Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7698–7702, December 1984
Biochemistry

Affecting gene expression by altering the length and sequence of the 5′ leader

(galactokinase fusions/translation initiation/AUG inserts/frameshift mutants)

HANNE JOHANSEN*†, DANIEL SCHÜMPERLI†‡, AND MARTIN ROSENBERG*†

*Department of Molecular Genetics, Smith Kline & French Laboratories, 709 Swedeland Road, Swedeland, PA 19479; and †Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205

Communicated by Maxine Singer, August 13, 1984

ABSTRACT **We have utilized a recombinant vector system designed to study gene control elements by fusing them to the *Escherichia coli* galactokinase gene (*galK*) to examine the effects of gene expression that result from introducing changes into the 5′ noncoding leader region of the *galK* transcription unit. We varied the length of the 5′ leader region and found no effect on *galK* expression, provided that the *galK* initiation codon remained the first AUG in the transcription unit. Using synthetic linkers, we then inserted specific sequences, each containing a single AUG codon at a defined position within the leader region. We found that the AUG inserts had widely different effects on *galK* expression and that the sequences surrounding the inserted AUG codons determined the magnitude of these effects. In addition, we placed these upstream inserted AUG codons into each of the three possible translation reading frames so that translation occurring in these frames terminates prior to, within, or downstream of the *galK* initiation codon, respectively. Single-nucleotide frameshift mutations were also introduced into one of these constructs to shift upstream translation into the other two possible reading frames. Depending upon where upstream translation stopped relative to the *galK* initiation codon, we observed consistently different effects on *galK* expression. Our results show that an upstream AUG that interferes with downstream translation initiation exerts its greatest effect when it translates out-of-frame through the downstream initiator into the gene. If translation is stopped upstream of or within the initiator, an unexpectedly high level of expression from the downstream AUG is maintained.**

In most eukaryotic mRNAs, translation initiates at the 5′-proximal AUG codon. Based on this observation, a scanning model was proposed to explain eukaryotic translation initiation (1). The model suggested that ribosomes bound initially to the 5′ end of a mRNA and then moved along the message until they encountered the first AUG triplet. The model, in its original form, implied that the only important feature was the position of the appropriate AUG codon 5′ to all other AUG triplets on the mRNA. Evidence has accrued which, in general, supports this scanning mechanism (2–5). Most notable are recent experiments showing that ribosomes initiate exclusively at the 5′-proximal AUG of a preproinsulin mRNA that contains tandemly repeated copies of the translation initiation site for the preproinsulin gene (6).

As additional eukaryotic mRNAs have been characterized, a number of exceptions to the "first AUG" rule have been noted (7–9). In these messages, one or more AUG triplets occur upstream of the initiator AUG. To account for these exceptions, the scanning model was modified (3). It was suggested that those AUGs that served as functional initiators were within certain sequences that were preferred. After comparison of more than 250 cellular and viral mRNAs, the sequence (A/G)-C-C-A-U-G-G emerged as the favored sequence for eukaryotic translation initiation (10). Strong support for the importance of favored sequences comes from recent experiments in which single nucleotide changes were introduced near the translation start site of a cloned preproinsulin gene (11). These mutations had dramatic effects on the expression of that gene. Thus, the more recent versions of the scanning model suggest that ribosomes initiate efficiently at the first AUG only if it is flanked by favored sequences. An AUG flanked by unfavored sequences will be ignored, thereby allowing initiation at a downstream AUG.

Previously, we described a recombinant vector system, pSVK, that uses gene fusion to study eukaryotic gene regulatory elements (12–14). The vector system is analogous to other gene-fusion systems (15, 16) in that gene control elements are fused in a common transcription unit with a gene whose expression can be readily monitored. In contrast to other gene-fusion systems, our vectors carry two separate gene transcription units, each of which controls the expression of a different assayable, genetically selectable gene. One is the *Escherichia coli* galactokinase gene (*galK*) and the other is the *E. coli* xanthine (guanine) phosphoribosyltransferase gene (*gpt*) (see Fig. 1). The modular design of the vector system allows each of the regulatory elements controlling *galK* expression to be removed selectively and replaced by alternative DNA segments. Any vector changes that result in corresponding changes in levels of *galK* expression are measured accurately by using *gpt* expression as the internal standard. Transient expression can be measured 24–48 hr after DNA transfection; alternatively, expression can be assayed after stable introduction of the vector DNA into the host genome under the appropriate selective conditions (13, 18). Both *galK* and *gpt* activity from the same cell lysate are assayed by use of either a starch gel separation procedure or a rapid filter assay technique (refs. 13 and 19; *Materials and Methods*).

We report here the use of this vector system to study the effects of gene expression that result from systematically varying the length and sequence of the 5′ noncoding leader region of the *galK* transcription unit. In particular, we examined the effects of inserting AUG-containing sequences upstream of the *galK* initiation codon. Moreover, these inserts were positioned in all three possible translation reading frames occurring before the *galK* gene. Our results suggest that sequences immediately surrounding an upstream AUG play an important role in determining the effect of that AUG on downstream translation. In addition, different effects on expression are observed depending on whether potential

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *galK*, *Escherichia coli* galactokinase gene; *gpt*, *E. coli* xanthine (guanine) phosphoribosyltransferase gene; bp, base pair(s); SV40, simian virus 40.

‡Present address: Institut fur Molekularbiologie II der Universitat Zurich, Winterthurerstr 266A, CH 8057 Zurich, Switzerland.

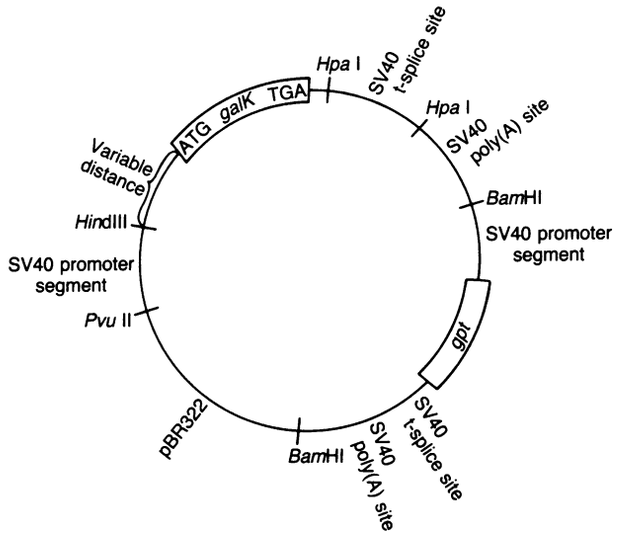


FIG. 1. Schematic representation of the basic pSVK vector (17) carrying DNA segments of various lengths in the 5′ leader of the *galK* transcription unit (see text for details). The *galK* transcription unit on pSVK has been described in detail (17) and consists of the simian virus 40 (SV40) early promoter region, the *galK* gene, the SV40 small tumor-antigen splice signal (t-splice site), and the SV40 early region polyadenylylation signal [poly(A) site]. A second transcription unit containing the *E. coli gpt* gene was obtained from pSV-*gpt* (15), adapted with *Bam*HI linkers, and inserted into the unique *Bam*HI site on pSVK (17). Restriction sites that separate the various regulatory elements of the *galK* transcription unit are also indicated. The *galK* and *gpt* coding regions are boxed.

translation from the upstream-inserted AUG terminates before, within, or after the downstream initiation codon. The implications of these results on the scanning model for eukaryotic translation initiation and the possible occurrence of multiple primary translation products from a single mRNA transcript are discussed.

MATERIALS AND METHODS

Vector Constructions. The construction of the basic pSVK vector has been described in detail (12, 13). We have also described the construction of a series of vectors, pDSΔK, which carry various segments of DNA sequence preceding the *galK* gene (13, 17). Each of these pDSΔK vectors was converted to a pSVK derivative using the procedures described in ref. 13. The vectors pSVK 100, 200, 300, 400, and 500 (Figs. 1 and 2) were derived from these pSVK derivatives by inserting a second transcription unit carrying the *gpt* gene, derived from the vector pSV2-*gpt* (15). The *gpt* transcription unit, contained on a *Bam*HI fragment, was inserted into the unique *Bam*HI site on each pSVK derivative (see Fig. 1). The 16-base-pair (bp) synthetic linker (obtained from G. Brown, Genetics Institute) (Fig. 3) was inserted at the unique *Hin*dIII restriction site on the vectors pSVK 100, 200, and 300. Prior to insertion of the linker, the *Hin*dIII site was filled in using DNA polymerase I Klenow fragment (20). The linker was inserted in both possible orientations by standard blunt-end ligation with T4 DNA ligase (20). The resulting constructs pSVK 150, 250, 350, 160, 260, and 360 were characterized by fine restriction and DNA sequencing (21) (Fig. 4).

The vectors pSVK 151 and 152 were derived from vector pSVK 150 by restriction with *Eco*RV at the unique site positioned 30 bp upstream of the translation initiation site of the *galK* gene, followed by treatment with T4 DNA polymerase (using conditions that promote 3′ exonucleolytic activity) to remove the 3′ overhanging end. The DNAs were then blunt-end ligated (20). The resulting constructs were characterized by fine restriction and DNA sequence analysis. All other methods used in vector constructions and characterization were standard procedures (20).

Cell Culture, DNA Transfection, and Enzyme Assay. African green monkey kidney cells (CV-1), and Chinese hamster cells (R1610, ref. 22) were propagated as described (13). Cells at passage 10 were transfected with various vector DNAs as described (13). The cells were harvested 24 hr (R1610) or 48 hr (CV-1) posttransfection, and cell extracts were prepared (23) and assayed for *galK* and *gpt* activity. Starch gel electrophoretic assay of *galK* and *gpt* were described previously (13, 14). *galK* filter assays were described by McKenney *et al.* (19). *gpt* filter assays were performed as described by Chasin and Urlaub (24), except that the standard reaction mixture contained 0.1 mM [8-C^{14}]guanine·HCl (48 Ci/mol; 1 Ci = 37 GBq), 1 mM 5-phosphoribosyl diphosphate, 5 mM $MgCl_2$, 3.2 mM NaF, bovine serum albumin at 2.4 mg/ml, and 0.05 M Tris HCl, pH 7.4, in a total volume of 50 μl. The reaction was initiated by the addition of 3 μl of cell extract and incubation was at 37°C for 40 min.

RESULTS AND DISCUSSION

Varying the Length of the 5′ Leader. In initial experiments, we used the pSVK vector system to examine the effects of varying the length of the 5′ leader on *galK* expression. We have described previously the construction of a set of plasmids that carried various lengths of DNA sequence preceding the translation initiation codon of the *E. coli galK* gene (17). The additional upstream sequences on these vectors were derived from the region of the bacterial genome that normally precedes *galK*. Constructs, characterized by direct DNA sequencing, carried precisely 37, 59, 74, 99, and 120 bp, respectively, of contiguous *galK* upstream sequence. These vectors were used to construct a set of pSVK derivatives (see the Methods and Fig. 1) that carry the various lengths of *galK* upstream sequences as part of the 5′ noncoding leader region of the *galK* transcription unit. In each case, the upstream segment was fused precisely at the *Hin*dIII restriction site positioned 64 bp downstream of the early simian virus 40 (SV40) promoter start site (i.e., cap site) carried on the pSVK vector. The resulting derivatives (pSVK 100, 200, 300, 400, and 500; Fig. 2*A*) contain nearly identical *galK* transcription units. Blot analysis of RNA produced by cells transfected with various pSVK derivatives (including pSVK 100) show that the *galK* and *gpt* transcription units utilize the predicted SV40 initiation and polyadenylylation signals (unpublished results; see ref. 15). The vectors pSVK 100–500 all contain the same SV40 promoter segment, the same predicted transcription start site, the same initial 64-nucleotide SV40-derived 5′ leader sequence, the same 37-bp segment immediately preceding the *galK* gene, and the same splice and polyadenylylation signals beyond *galK*. They differ, however, in that the overall lengths of their 5′ leader sequences vary from 101 to 184 nucleotides due to the different lengths of bacterial sequences that have been added upstream of *galK* (Fig. 2). Most important, on four of these vectors, pSVK 100, 200, 300, and 400, the *galK* initiation codon remains the first AUG triplet occurring downstream from the mRNA start site. Only one of the vectors, pSVK 500, contains additional AUG triplets in its leader region (Fig. 2).

We monitored *galK* and *gpt* expression from the five vector constructs introduced by DNA transfection into both monkey (CV-1) and hamster (R1610) cells using a transient expression assay (see the Methods). The results (Fig. 2*B*) indicate that the levels of *galK* expression in CV-1 cells from the pSVK 100, 200, 300, and 400 derivatives are identical. Similar results were obtained in R1610 cells (data not shown). The changes made in the length of the *galK* 5′ leader

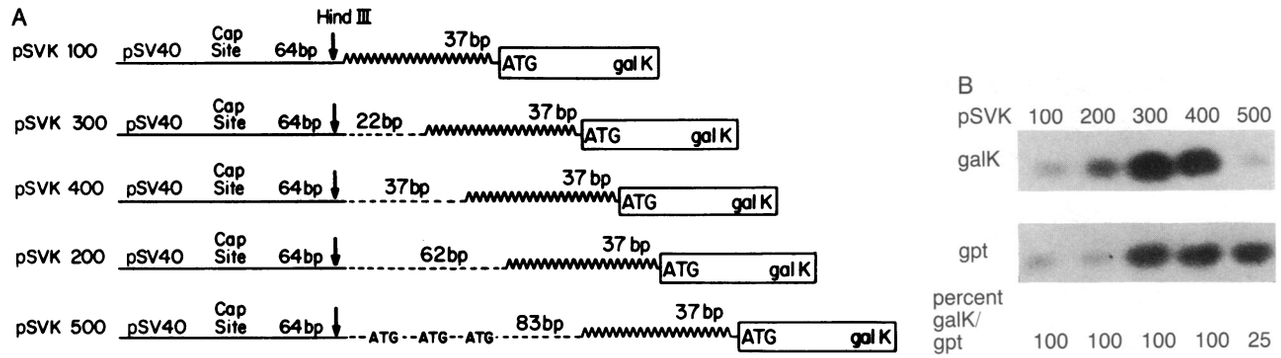


FIG. 2. (*A*) Schematic representation of the *galK* transcription unit contained in a set of similarly constructed vectors (pSVK 100–500) carrying 5′ noncoding leader regions of different lengths. The position of the *galK* gene is indicated, as is the length of various portions of the leader region and other ATG codons occurring in the leader. Each of these vectors carries the same SV40 early promoter segment (pSV40), the same transcription start site (cap site), the same 5′ leader segment derived from the *gal* operon of *E. coli* (〰). These vectors differ only in the length of the *E. coli gal* operon DNA segment (------) introduced between the 64-bp and 37-bp 5′ leader segments. See text for other details. (*B*) *E. coli galK* and *gpt* activities measured in CV-1 African green monkey kidney cells after transfection with vectors pSVK 100–500. Fluorograph shows the activity in cell extracts that have been subjected to starch gel electrophoresis (17). *galK* and *gpt* gene product activities were quantified by scanning the fluorograph using a Zenith laser densitometer. Ratios of *galK* and *gpt* gene product activities were determined and expressed as percents relative to the galK/gpt ratio obtained for pSVK 100. Values given are the average of three independent transfection experiments; values never deviated from the average by more than 20%.

had no effect on expression in either cell type. Apparently, the *galK* initiation signal is recognized with the same efficiency when placed at any of the different distances relative to the mRNA cap site. In contrast to the results obtained with the vectors pSVK 100–400, the pSVK 500 derivative showed a decrease of ≈75% in *galK* expression. The *galK* 5′ leader region in this vector is only 21 nucleotides longer than that of pSVK 200, and thus, the added leader length is not likely to be responsible for the reduction of *galK* expression. Instead, the decrease is more likely due to the occurrence of the three additional AUG codons contained within this 21-nucleotide segment. Presumably, one or more of these upstream AUGs competes for translation with the downstream *galK* initiation codon. Surprisingly, however, the *galK* initiation codon is still recognized with ≈25% efficiency in pSVK 500 even though three different AUG codons are positioned upstream. The results presented below provide an explanation for this rather high residual initiation from the downstream AUG codon.

Inserting AUG-Containing Sequences into the Leader. Our rationale was to use the pSVK 100, 200, and 300 vectors to examine the effects of systematically introducing specific AUG-containing sequences into the 5′ noncoding leader on the *galK* transcription unit. The approach was to insert small DNA segments, each containing a single AUG codon flanked by different sequences, at defined locations in the 5′ leader region of the *galK* transcription unit and then monitor the effect on *galK* expression. For this work, we obtained a 16-bp synthetic DNA linker (generously provided by G. Brown, Genetics Institute) that encoded one AUG triplet on each of its DNA strands (Fig. 3). This linker was inserted in both orientations at the unique *Hin*dIII restriction site positioned 64 bp into the 5′ leader region of the *galK* transcription unit on the vectors pSVK 100, 200, and 300 (Fig. 3). Irrespective of its orientation, this linker adds a single AUG codon into the 5′ leader sequence upstream of *galK*. In one orientation the inserted AUG is flanked by the favored sequence (A-C-C-A-U-G-G) for eukaryotic translation initiation, whereas in the opposite orientation the AUG is surrounded by an unfavorable sequence (U-C-C-A-U-G-G). In particular, the favored sequence contains the requisite purine three nucleotides 5′ to the AUG, whereas the unfavored sequence encodes a uridine residue at this position. The downstream authentic *galK* initiator is flanked by the rather favorable sequence G-A-A-A-U-G-G.

Six vector derivatives, three containing the favored AUG sequence (pSVK 150, 250, and 350; Fig. 4) and three containing the unfavored AUG sequence (pSVK 160, 260, and 360; Fig. 4) were obtained and characterized by fine restriction and direct DNA sequencing. All six constructions were assayed for transient expression in both CV-1 monkey cells and R1610 hamster cells. The results (Fig. 5) indicate that the two different AUG sequences had widely different effects on *galK* expression.

In monkey cells, the three constructs carrying the favored AUG expressed *galK* ≈90% less efficiently than did the parent vectors. In contrast, the three vectors carrying the unfavored AUG showed only a 50% reduction in *galK* expression. We obtained similar results using the hamster cell line (Fig. 5). Assuming that the observed interference with *galK* expression correlates directly with the relative ability of the two different AUG sequences to be recognized by the translational machinery of the cell, then the favored sequence is recognized about five times as well as the unfavored se-

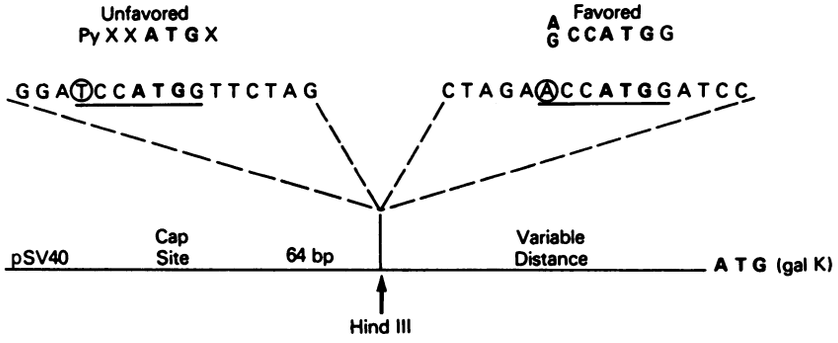


FIG. 3. Schematic representation of the insertion of a 16-bp linker into the unique *Hin*dIII site in the leader region of pSVK 100, 200, and 300 (see Fig. 2 and text for details). The linker is shown inserted in both possible orientations and ATG codons (bold) and flanking sequences are underlined. Also shown is the consensus "favored" and "unfavored" sequence for eukaryotic translation initiation sites (10). Py, pyrimidine nucleotide; X, any nucleotide. The circled residue indicates the residue in the linker at the important position −3 nucleotides upstream of the ATG. All other designations are as in Figs. 1 and 2.

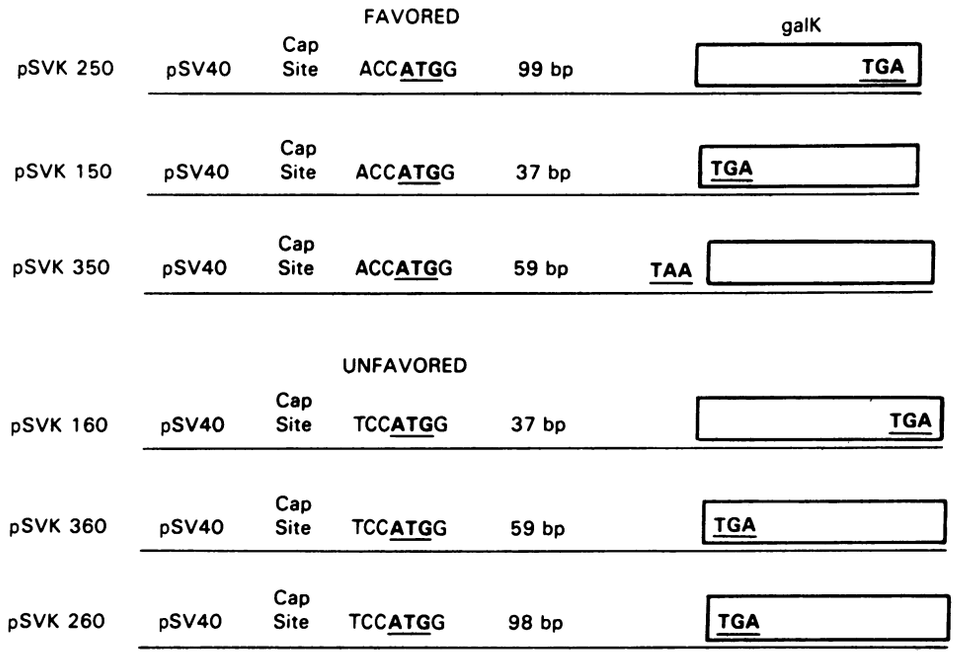
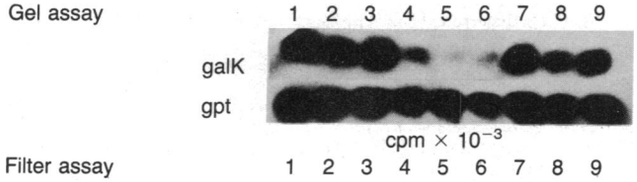


FIG. 4. Schematic representation of the six vector derivatives that result from the insertion of the synthetic linker (Fig. 3). The name of each vector and the position and orientation of the synthetic linker are indicated. Also shown is the position at which translation initiating from the inserted ATG will terminate (stop codon) relative to the *galK* coding sequence. In vectors pSVK 150, 260, 360 translation terminates within the *galK* initiation codon (ATGA). Two codons upstream and in-frame with the *galK* sequence is a chain-terminating codon (TAA). This stop signal prevents any translation initiating upstream of *galK* from reading through in-frame with the *galK* gene (e.g., pSVK 350). DNA sequencing indicated that pSVK 260 had undergone a single-base-pair deletion during insertion of the synthetic linker and thus has 98 bp (rather than 99 as in pSVK 250) between the linker and the *galK* coding sequence. All other designations are as in Figs. 1 and 2.

quence. Apparently, the sequence flanking the AUG codon, in particular, the residue in the −3 position, plays a major role in determining the efficiency with which a particular AUG is recognized as an initiation codon. These results are in complete accord with those recently obtained by Kozak (10).

Vector	Position of stop codon in relation to ATG of *galK*	Percent galK/gpt	
		CV-1	R1610
1. pSVK100	—	100	100
2. pSVK200	—	100	100
3. pSVK300	—	100	100
4. pSVK150	Within	≈10	≈20
5. pSVK250	Beyond	≈10	≈5
6. pSVK350	Before	≈10	≈20
7. pSVK160	Beyond	≈50	≈65
8. pSVK260	Within	≈50	≈65
9. pSVK360	Within	≈50	≈65



Filter assay	1	2	3	4	5	6	7	8	9
[C¹⁴]galactose-1-phosphate	188	190	218	50	15	32	218	107	175
[C¹⁴]guanosine-5-monophosphate	26	35	57	33	31	20	62	15	43

FIG. 5. (*Top*) Expression of *galK* relative to expression of *gpt* for the vectors described in Figs. 2 and 4. DNA transfections with each vector were carried out both in CV-1 monkey cells and R1610 hamster cells. Assays were performed and quantified as described in the Methods and the legend to Fig. 2 (13, 14, 19, 24). Values are expressed as percentages relative to the galK/gpt ratio determined for pSVK 100 (as in Fig. 2*B*) and are averages of four independent experiments. (*Bottom*) A representative starch-gel assay and filter assay of galK and gpt activities measured for the nine vectors shown above (corresponding numbers) transfected into *galK*⁻ *gpt*⁻ R1610 hamster cells. The gel assay is a fluorograph, as in Fig. 2*B*. The filter assay gives radioactivity from [C¹⁴]galactose and [C¹⁴]guanine converted to the appropriate phosphorylated derivatives by the bacterial galK and gpt activities.

Translation Reading-Frame Effects. In the experiments reported above, similar results were obtained using either the monkey or hamster cell lines. However, using the hamster cell line, we detected reproducible differences in expression among the three vectors carrying the favored AUG sequence (pSVK 150, 250, and 350; Fig. 5). These differences were discernible in the hamster cell line because of the higher levels of expression that could be obtained (apparently due to more efficient DNA transfection) and also because of the high accuracy of the filter method used to assay extracts from these cells (see the Methods). Our results showed that the pSVK 250 derivative exhibited ≈75% lower *galK* expression than did the other two vectors, pSVK 150 and 350. The three vectors differ only in that the inserted linker occurs in different positions relative to the *galK* initiation codon. It is most interesting that the position of the insert places the added AUG codon in these vectors into each of the three possible reading frames occurring upstream of the *galK* gene. As shown in Fig. 4, translation occurring from the various inserted AUG codons terminates before (pSVK 350), within (pSVK 150, ATGA), or downstream of (pSVK 250) the *galK* initiation codon. Our results suggest that the *galK* AUG is used least efficiently when translation initiating upstream continues out-of-frame well beyond this codon (i.e., terminates 98 codons into the *galK* gene, as in pSVK 250). In contrast, stopping translation before (as in pSVK 350) or even within (as in pSVK 150) the *galK* initiation codon results in increased utilization of this AUG. We did not observe any such reading-frame effects among those vectors that carried the unfavored AUG sequence (pSVK 160, 260, and 360; Figs. 4 and 5). Apparently, the inserted AUG codon in these vectors is poorly recognized and, thus, its reading frame is unimportant.

Frameshift Mutations. The results obtained above suggested that translation initiating upstream of a gene can differentially affect its expression, depending on where this translation terminates. However, since the vectors used in our study have different length 5′ leader regions and differently positioned AUG inserts, we could not rule out the possibility that these factors were affecting our results. In an effort to examine more directly the effect on *galK* expression resulting from alteration of the upstream translation reading

frame, we introduced frameshift mutations into the pSVK 150 vector between the inserted AUG sequence and the *galK* initiation codon (see the Methods). In particular, two pSVK 150 derivatives were characterized (pSVK 151 and 152, Fig. 6) that contained one and two nucleotide deletions, respectively; these deleted residues frameshift the upstream translation into the two other possible reading frames. In pSVK 151, translation traverses well beyond the *galK* initiator, stopping within the *galK* gene (as in pSVK 250), whereas in pSVK 152 translation stops upstream of the *galK* initiation codon (as in pSVK 350). However, unlike the situation with pSVK 250 and 350 (Fig. 4), the pSVK 150, 151, and 152 vectors are identical except for the single and double nucleotide deletions.

We monitored transient *galK* and *gpt* expression from the three constructs in the hamster cell line. The results (Fig. 6) are consistent with those obtained above for pSVK 150, 250, and 350. Stopping upstream translation at or before the *galK* AUG (pSVK 150 and 152, respectively) resulted in ≈3.5-fold higher levels of *galK* expression than when translation was allowed to read through the *galK* AUG into the gene. One possible explanation for this effect is that ribosomes terminating upstream or within the *galK* AUG can reinitiate at this AUG, thereby increasing *galK* expression. Alternatively, but perhaps less likely and more contradictory to the scanning model, ribosomes may initiate independently at the authentic *galK* AUG, and this interaction may be inhibited by the read-through translation from upstream. These results provide an explanation for the high basal level of *galK* expression observed with the pSVK 500 derivative (Fig. 2*B*). The 5′ leader of the *galK* transcript from this vector contained three AUG triplets upstream of the *galK* initiator, yet continued to express *galK* at 25% efficiency (relative to *gpt* expression). Two of these three upstream AUGs are flanked by unfavorable sequences (U-U-U-A-U-G-G and G-A-A-A-U-G-C), but the third AUG (G-U-U-A-U-G-A) should compete favorably with the *galK* initiator. However, translation initiating from any of these upstream AUG triplets would terminate before or within the *galK* initiator, thereby allowing *galK* expression.

Implications. We have varied the distance between the transcription start site and the translation initiation codon of the *galK* transcription unit from 101 to 163 nucleotides without affecting gene expression. This result implies that the translation efficiency of an initiation codon can be maintained at different distances from the 5′ cap site. In contrast, when sequences containing one or more AUG triplets were introduced into the leader, *galK* expression was affected and the magnitude of the effect varied considerably depending upon the sequences flanking the inserted AUG codons. Our results, as well as the results of others (4, 10), suggest the existence of sequence determinants close to the AUG that are important for mediating translation efficiency. This feature is reminiscent of the situation known to exist in prokaryotes.

In addition, we found that an upstream AUG that interferes with initiation downstream does so most effectively when its translation reading frame traverses the downstream initiator and continues into the gene. If the reading frame is terminated either before or within the downstream initiator, then substantial expression occurs from the downstream signal. We emphasize, however, that this effect is of secondary importance when compared to the effect of the AUG flanking sequences.

Most important, our results imply that the factors controlling the efficiency of translation initiation in eukaryotic systems may be far more complex than the initially proposed concept of "first AUG on a monocistronic mRNA." AUG codons positioned 5′-proximally on a mRNA may well be important determinants of downstream translation efficiency. Moreover, the apparent ability of multiple AUGs on a single RNA to be recognized differentially (due to their position, flanking sequences, and translation reading frame) implies that multiple translation products can be derived from the same transcript. In fact, overlapping in-frame translation starts have already been found (25, 26), and it is possible that entirely separate cistrons may exist on a single eukaryotic mRNA.

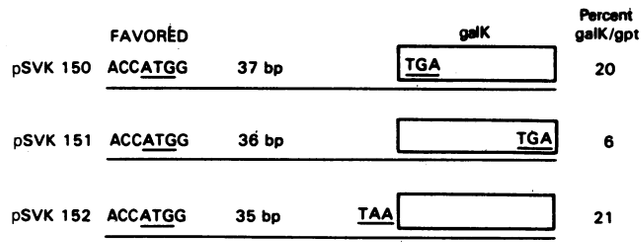


FIG. 6. Schematic representation of the *galK* transcription unit in pSVK 150 (as in Fig. 4) and in two different derivatives obtained by deleting 1 (pSVK 151) and 2 (pSVK 152) bp from the leader region between the synthetic linker insert and the *galK* gene (see text for details). Also shown are the galK/gpt activities determined for each vector. These activities were measured (as in Fig. 5) by the filter assay method and are expressed as percentages relative to the galK/gpt value determined for pSVK 100. Values given are averages of three independent transfection experiments with R1610 hamster cells. All other designations are as in Figs. 2–4.

1. Kozak, M. (1980) *Cell* **22**, 7–8.
2. Paterson, B. & Rosenberg, M. (1979) *Nature (London)* **279**, 692–696.
3. Kozak, M. (1981) *Nucleic Acids Res.* **9**, 5233–5252.
4. Lomedico, P. T. & McAndrew, F. (1982) *Nature (London)* **299**, 221–226.
5. Liu, C.-C., Simonsen, C. C. & Levinson, A. D. (1984) *Nature (London)* **309**, 82–85.
6. Kozak, M. (1983) *Cell* **34**, 971–978.
7. Ghosh, P. K., Reddy, V. B., Swinscoe, F., Choundary, P. V., Lebowitz, P. & Weissman, S. M. (1978) *J. Biol. Chem.* **253**, 3643–3647.
8. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodnes, R. & Dodgson, F. (1980) *Cell* **20**, 555–566.
9. Kitamura, N., Semiler, B. L., Rothberg, P. G., Larsen, G. R., Adler, C. F., Dorner, A. J., Emini, E. A., Hanecak, R., Lee, F. F., van der Werf, S., Anderson, C. W. & Wimmer, E. (1981) *Nature (London)* **291**, 547–553.
10. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
11. Kozak, M. (1984) *Nature (London)* **308**, 241–246.
12. Johansen, H., Reff, M., Schümperli, D. & Rosenberg, M. (1984) in *Gene Expression*, The Alfred Benzon Symposium No. 19, eds. Clark, B. F. C., Peterson, H. & Thaysen, J. H. (Munksgaard, Copenhagen), pp. 413–429.
13. Schümperli, D., Howard, B. H. & Rosenberg, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 257–261.
14. Berg, P., Yu, J. U., Popovic, Z., Schümperli, D., Johansen, H., Rosenberg, M. & Anderson, F. (1983) *Mol. Cell. Biol.* **3**, 1246–1254.
15. Mulligan, R. C. & Berg, P. (1981) *Mol. Cell. Biol.* **1**, 449–459.
16. Gorman, C., Merlina, G. T., Willingham, M. C., Pastan, I. & Howard, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6777–6781.
17. Schümperli, D., McKenney, K., Sobieski, D. A. & Rosenberg, M. (1982) *Cell* **30**, 865–871.
18. Mulligan, R. C. & Berg, P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2072–2076.
19. McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, L. & Rosenberg, M. (1981) in *Gene Amplification and Analysis*, eds. Chirikjian, J. S. & Papas, T. S. (Elsevier/North-Holland, New York).
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
21. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
22. Thirion, J. P., Banville, D. & Noel, H. (1976) *Genetics* **83**, 137–147.
23. Peterson, F. L. & McBride, O. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1583–1587.
24. Chasin, L. A. & Urlaub, G. (1976) *Somatic Cell Genet.* **2**, 453–467.
25. Rosenberg, M. & Paterson, B. M. (1979) *Nature (London)* **279**, 696–701.
26. Laughon, A., Driscoll, R., Wills, N. & Gesteland, F. (1984) *Mol. Cell. Biol.* **4**, 268–279.