

Termination-Reinitiation Occurs in the Translation of Mammalian Cell mRNAs

DAVID S. PEABODY*† AND PAUL BERG

Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305

Received 3 December 1985/Accepted 7 April 1986

Many examples of internal translation initiation in eucaryotes have accumulated in recent years. In many cases terminators of upstream reading frames precede the internal initiation site, suggesting that translational reinitiation may be a mechanism for initiation at internal AUGs. To test this idea, a series of recombinants was constructed in the mammalian expression vector pSV2. Each contained a dicistronic transcription unit comprising the coding sequence for mouse dihydrofolate reductase (DHFR) followed by the gene for xanthine-guanine phosphoribosyl transferase (XGPRT) from *Escherichia coli*. Various versions of this pSV2dhfr-gpt recombinant plasmid altered the location at which the DHFR reading frame was terminated relative to the XGPRT initiation codon and demonstrated that this is a critical factor for the expression of XGPRT activity in transfected Cos-1 cells. Thus, when the DHFR frame terminated upstream or a very short distance downstream of the XGPRT initiator AUG, substantial levels of XGPRT activity were observed. When the DHFR frame terminated 50 nucleotides beyond the XGPRT initiator, activity was reduced about twofold. However, when the DHFR and XGPRT sequences were fused in-frame so that ribosomes which initiated at the DHFR AUG did not terminate until they encountered the XGPRT terminator, production of XGPRT activity was abolished. This dependence of internal translation initiation on the position of terminators of the upstream reading frame is consistent with the hypothesis that mammalian ribosomes are capable of translational reinitiation.

The mechanisms that eucaryotic ribosomes use to select the correct AUG for translation initiation are still unknown. Although the basic machinery of translation is similar in procaryotes and eucaryotes, analysis of the structure of a large number of eucaryotic mRNAs indicates that the initiation process must differ from its procaryotic counterpart in significant details (13). In bacteria, recognition of the initiation codon is facilitated by the so-called Shine-Dalgarno sequence located a narrowly variable number of nucleotides upstream of the initiator AUG. This sequence is complementary to a region near the 3' end of the 16S rRNA and ostensibly directs ribosome binding to the correct initiator codon. Bacteria are capable of initiation at several sites on a single polycistronic mRNA, presumably because initiation can occur by the ribosome's binding directly to an internal site on the mRNA. In contrast, eucaryotic mRNAs are nearly always monocistronic, and translation is initiated most frequently at the AUG nearest the 5' end. There is no indication that a eucaryotic equivalent of the procaryotic Shine-Dalgarno sequence is required.

To account for the differences in translation initiation with procaryotic and eucaryotic mRNAs, Kozak proposed that eucaryotes use a "scanning" mechanism for translation initiation (20). Accordingly, the 40S ribosomal subunit binds the mRNA molecule at or near the 5' cap structure and migrates along the mRNA until it encounters an AUG triplet; then the 60S subunit associates with the small subunit to form the complete polymerizing complex, and synthesis of the polypeptide chain ensues. As originally formulated, the scanning hypothesis maintained that the first AUG would always serve as the initiator and that no other sequence or

structural features were required. Biochemical evidence indicating that a free 5' capped end is required (21) and that 40S subunits can migrate along the mRNA (22-24, 27) support the model (26). The Kozak scanning model is also consistent with the results of an extensive genetic analysis of the *Saccharomyces cerevisiae* gene for iso-1-cytochrome *c*, which suggests that initiation occurs only at the first AUG and that the position of the first AUG can be altered with no apparent effect on translation efficiency (49).

Recently, however, an increasing number of exceptions to the first AUG rule have been discovered, prompting a reevaluation of the scanning hypothesis. Many naturally occurring cellular (14, 44, 55, 58) and viral mRNAs (5, 7, 10, 11, 17, 19, 30, 31, 53, 57) are now known to initiate translation at internal AUGs. Moreover, internal initiation has been demonstrated in molecules constructed artificially by recombinant DNA methods (29, 38, 50, 52).

Three different models may be proposed to account for the apparent exceptions to the first AUG rule. (i) Direct binding of ribosomes to an internal initiation site: for reasons mentioned above, this is unlikely but cannot be ruled out categorically as occurring in some cases. (ii) Termination-reinitiation: after termination of translation of an upstream reading frame, ribosomes might remain associated with the mRNA and reinitiate at a nearby internal AUG. (iii) Relaxed scanning: some AUGs may be recognized poorly as initiation codons, permitting ribosomes to scan past them and initiate at internal sites. Relaxed scanning implies that structural features other than the AUG itself contribute to initiation codon recognition, an idea that deserves closer examination because the sequence of nucleotides around initiator AUGs is nonrandom (25, 26).

In this work we have tested some predictions of these models by constructing multicistronic transcription units in the pSV2 plasmid vector (2) and examining the structural requirements for translation of the downstream coding re-

* Corresponding author.

† Present address: Department of Cell Biology, Cancer Research and Treatment Center, University of New Mexico, Albuquerque, NM 87131.

gion in cultured monkey cells. The results are consistent with the view that eucaryotic ribosomes are capable of termination-reinitiation and that such a mechanism may account for many examples of internal initiation of translation.

MATERIALS AND METHODS

Cell culture and DNA transfection. The care and maintenance of simian virus 40 (SV40)-transformed CV1 cells (COS cells) have been described previously (12). Twenty-four hours after about 2×10^6 cells were seeded in a 100-mm plate, the cultures were transfected with 10 μ g of a test plasmid and 5 μ g of pCH110 by the method of Graham and Van der Eb (14) as modified by Parker and Stark (42). The plasmid pCH110 (provided by F. Lee, DNAX Institute of Molecular and Cellular Biology) is a derivative of pSV2 which contains the coding sequence of *E. coli* β -galactosidase transcribed from the SV40 early promoter (16). Assays of β -galactosidase activity in the transfected cell extracts provide a convenient means for normalizing the efficiency of different transfections.

Enzymes and chemicals. *SalI* endonuclease was obtained from K. Burtis (Stanford University), and *E. coli* DNA polymerase I and T4 DNA ligase were provided by S. Scherer (Stanford University). S1 nuclease and calf intestine alkaline phosphatase were purchased from Boehringer Mannheim. Bal31 nuclease was from Bethesda Research Laboratories, and other enzymes were from New England Biolabs. Oligonucleotides containing the *SalI* restriction site were from Collaborative Research, Inc.

Construction of recombinant plasmids. The general methods for ligation, recovery of DNA fragments from agarose gels, transformation of *E. coli* HB101, and rapid screening of clones by restriction enzyme analysis have been described elsewhere (9). Plasmids were prepared on a large scale by an alkaline lysis method (4) followed by two successive equilibrium sedimentations in CsCl-ethidium bromide gradients.

The construction of pSV2dhfr26 (52) and pSV2gpt (37) have been described previously. The plasmid pSV2dhfr-gpt1 was constructed by excising the *gpt* sequence from pBRgpt (37) as a *Bam*HI-*Bgl*III fragment and inserting it into the *Bgl*III site of pSV2dhfr6 (Fig. 1). Clones which contained the *gpt* gene in the correct orientation were easily identified since they reconstructed the *Bgl*III restriction site at the junction of the *dhfr* and *gpt* sequences. The conjunction of sequences in this region created two short open reading frames between the *dhfr* and *gpt* coding regions. To delete sequences in this region, pSV2dhfr-gpt1 was cleaved at the single *Bgl*III restriction site and treated with Bal31 nuclease. *SalI* linkers were ligated to the resected blunt ends, and after digestion with an excess of *SalI* endonuclease the plasmid DNA was separated from the linkers by electrophoresis in an agarose gel. The recovered DNA was recircularized by ligation and used to transform *E. coli*. The deletions were mapped by restriction analysis, and the exact end points of suitable candidates were determined by nucleotide sequence analysis with the method of Maxam and Gilbert (32). Since no single molecule had the desired deletions at both ends, it was necessary to recombine different mutants at the *SalI* restriction site in vitro to obtain the desired arrangement of reading frames. Thus, to construct pSV2dhfr-gpt6 and -8, plasmids containing the appropriate right- or left-hand deletion end points were digested with *SalI* and *Eco*RI endonucleases, and the fragments were isolated by agarose gel electrophoresis. Different halves of the molecules were ligated together

and introduced into *E. coli* to obtain the appropriate recombinants.

The plasmid pSV2dhfr-gpt13 was derived from pSV2dhfr-gpt8 by insertion of 4 base pairs (bp) at the *SalI* restriction site. This was accomplished by treating the *SalI*-cut plasmid with DNA polymerase I in the presence of the four deoxynucleotide triphosphates to create blunt ends, which were then joined together by ligation with T4 DNA ligase. After a second digestion with *SalI* endonuclease to eliminate molecules which had retained the site, the DNA was introduced into *E. coli*. The exact duplication of the 4-base overhang of the *SalI* restriction site created a *Pvu*II cleavage site, providing a convenient means of verifying that the clones had acquired the correct sequence.

The plasmid pSV2dhfr-gpt14 was constructed by deletion of 4 bp at the *SalI* site of pSV2dhfr-gpt8 by the following scheme. Five micrograms of pSV2dhfr-gpt8 was digested to completion with *SalI* endonuclease, ethanol precipitated, and redissolved in S1 nuclease buffer (0.25 M NaCl, 0.03 M sodium acetate, pH 4.6, 0.001 M ZnCl₂). The DNA was then divided into five equal portions and treated with different amounts of S1 nuclease (ranging from 50 to 5,000 U) for 30 min at 12°C. The reactions were terminated by phenol-chloroform extraction followed by ether extraction and ethanol precipitation. After recircularization by overnight ligation at 12°C, the plasmids were digested again with *SalI* to eliminate sites that had not been destroyed by the S1 nuclease treatment. Each of the five portions was used separately to transform *E. coli*. When clones from the plates

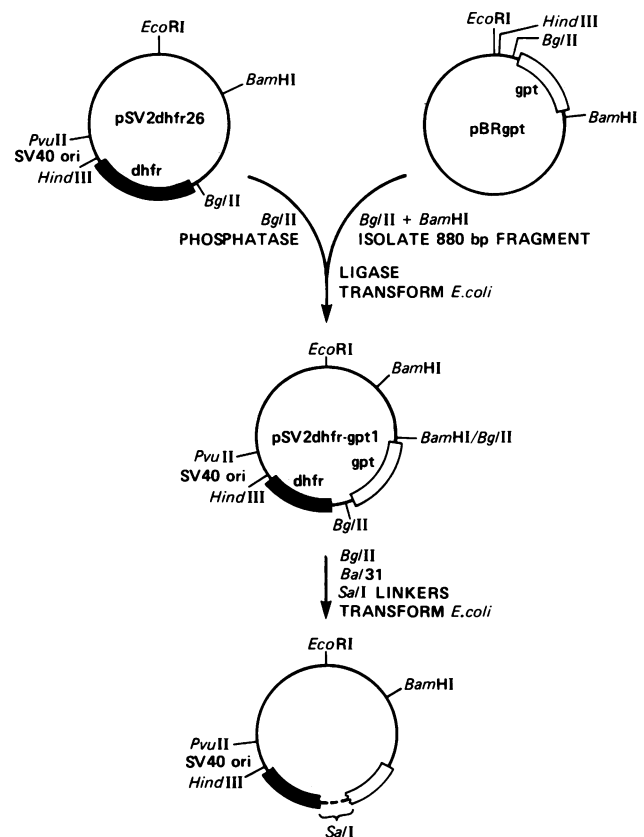


FIG. 1. Construction of the dicentric transcription unit in pSV2dhfr-gpt1 and the derivation of the deletion mutants as described in the text.

with the highest numbers of transformants were analyzed, about 90% of the molecules had undergone exact deletion of the 4-base overhang of the *Sa*II site. Under the conditions described here, maximization for transformation efficiency also optimizes blunt-end formation. This was generally achieved when the linear plasmids were treated with about 300 U of S1 nuclease per μ g of DNA.

Characterization of RNA. Total cytoplasmic RNA was extracted from cells 48 h after transfection by methods already described (39) and analyzed by the Weaver and Weissman modification (56) of the S1 nuclease protection technique of Berk and Sharp (3). In all the constructions described in this paper, the *Bam*HI site at the 3' end of the *gpt* gene was joined to a *Bgl*II site in the vector so that both restriction sites were destroyed. To generate probes for S1 nuclease analysis, special derivatives were constructed which retained the *Bam*HI site in this position (D. Peabody, unpublished data). These plasmids were digested with *Bam*HI, and their 5' ends were labeled with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and polynucleotide kinase (32). After digestion with *Pst*I endonuclease, the appropriate fragments were isolated by electrophoresis in agarose gels. Probes labeled in this position can be used to assay for intactness of the mRNA along the entire length of the coding regions. DNA fragments protected from S1 nuclease digestion were analyzed by electrophoresis in alkaline agarose gels (33).

Transcripts were also analyzed by a blotting analysis of polyadenylated RNA (1). Samples were denatured by glyoxalation, electrophoresed in agarose gels, and blotted to nitrocellulose essentially as described by Thomas (54). Probes were prepared by nick translation (46) of isolated *dhfr* or *gpt* fragments and hybridized to the immobilized RNA by established methods (54).

Analysis of protein products. Protein was extracted from approximately 10^7 cells by methods previously described (52). The production of dihydrofolate reductase (DHFR) was measured by the blotting technique (45) described by Burnette (5). Proteins separated on 15% polyacrylamide gels in the presence of sodium dodecyl sulfate (28) were transferred to nitrocellulose, and DHFR was detected by incuba-

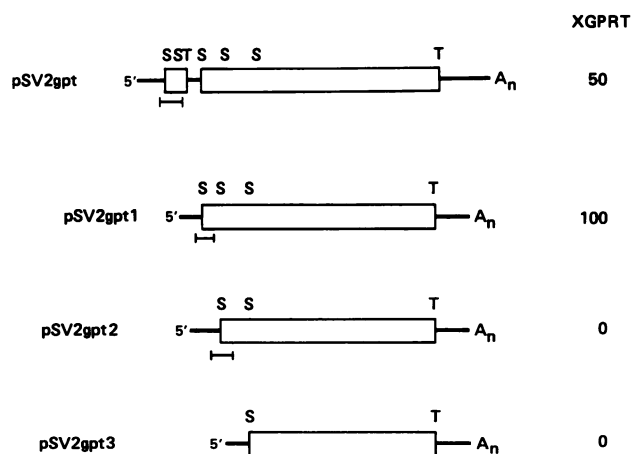


FIG. 2. Predicted structures of the mRNAs produced by recombinants used to locate the XGPRT initiator AUG. Boxes indicate open reading frames, and the letters S and T denote, respectively, the translation initiation and termination codons at their boundaries. The large box includes the XGPRT open reading frame. The relative amounts of XGPRT produced after transfection of COS cells with these recombinants is indicated in the column at the right; the activity obtained with pSV2 gpt1 is arbitrarily defined as 100.

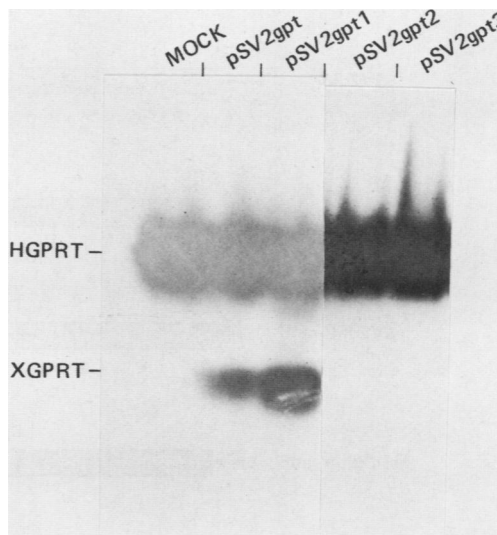


FIG. 3. Electrophoretic assays of XGPRT in extracts of cells transfected with the recombinants illustrated in Fig. 2. Each track contains extract from about 10^6 cells, the exact amount having been determined by reference to β -galactosidase activity as described in the text.

tion with antiserum to mouse DHFR (a gift of R. Schimke, Stanford University) and subsequent treatment with ^{125}I -labeled protein A (provided by S. Schmid, Stanford University). The filter was exposed to Kodak XAR-5 film with enhancement by a Dupont Cronex Lightning-Plus screen.

β -Galactosidase activity was determined as previously described (36) and was used to determine how much of each extract was assayed for xanthine-guanine phosphoribosyl transferase (XGPRT). Thus, each sample was normalized for variations in transfection efficiency. XGPRT activity was measured by a modification of the in situ gel assay described previously (37). Extracts were fractionated on 7.5% polyacrylamide gel slabs in the Davis Tris-glycine system (7). After the electrophoresis, the gel was removed from the glass plates, soaked for 10 min in 50 mM Tris hydrochloride, pH 7.5, and then placed on a prewetted sheet of polyethyleneimine (Polygram CEL 300; Brinkmann Instruments, Inc.). A 1.5-ml assay mixture containing 10 mM MgCl_2 , 10 mM phosphoribosyl pyrophosphate, 100 mM Tris hydrochloride, pH 7.5, and 150 μCi of $[\text{H}^3]\text{guanine}$ (20 Ci/mmol) was spread on the surface of the gel and incubated at 37°C for 60 min on a level surface. During this time the reaction products diffused through the gel to the polyethyleneimine sheet. After removal of the gel, the sheet was washed in water, dried, and fluorographed. XGPRT activities were quantitated by scanning the fluorogram with a densitometer.

RESULTS

Identification of the XGPRT initiator AUG. As originally cloned in pSV2gpt (37, 38), the *gpt* sequence contains two AUGs in a single reading frame, 22 codons long, upstream of the XGPRT coding region (Fig. 2). Deletion of 120 nucleotides of upstream sequence in pSV2gpt1 (referred to as pSV2gpt BglII del in reference 37) resulted in a twofold increase in XGPRT production (Fig. 3). This increase in activity could be due either to removal of the two upstream AUGs or to the decreased length of the mRNA leader. The open reading frame that encodes XGPRT contained three

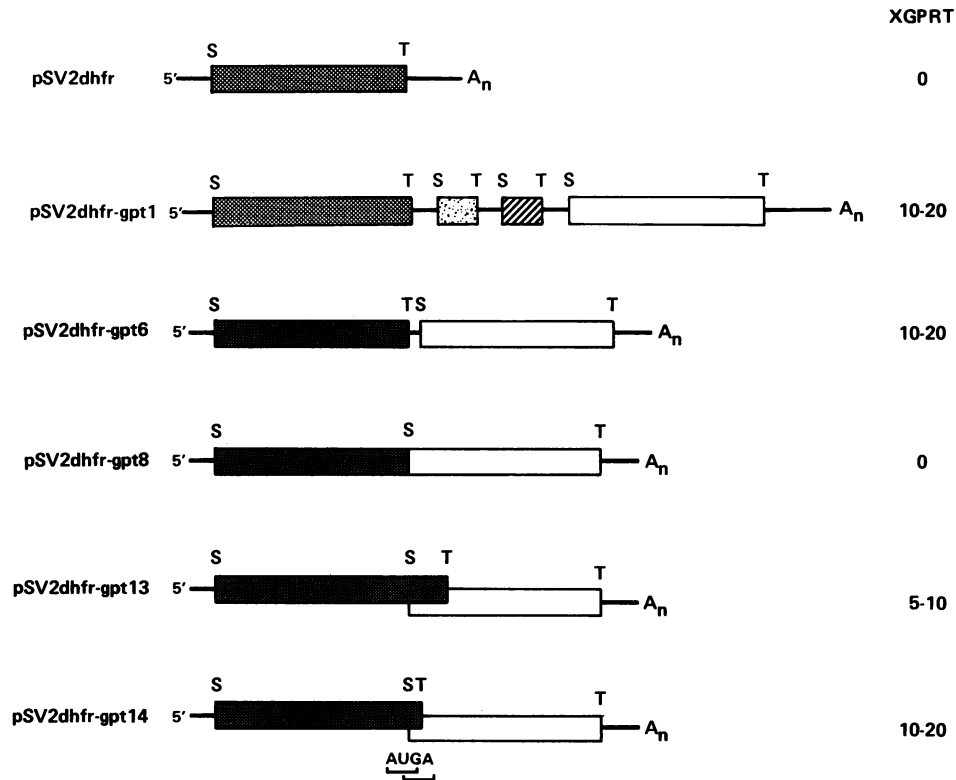


FIG. 4. Predicted structures of the mRNAs produced by pSV2dhfr and the dicistronic pSV2dhfr-gpt derivatives. Boxes indicate coding regions, and the letters S and T refer to translation initiation and termination codons as for Fig. 2. The solid black box denotes the DHFR coding region, whereas the open box illustrates the XGPR T reading frame. The stippled and hatched regions are the short reading frames which result from the conjunction of sequence in the 3' untranslated portion of *dhfr* and the 5' flanking sequences of the *gpt* gene. The overlapping coding regions of pSV2dhfr-gpt 13 and 14 are indicated by the overlapping boxes. Relative amounts of XGPR T are listed on the right, with the activity produced by pSV2gpt 1 defined as 100.

potential translation initiation sites (Fig. 2) that could encode polypeptides of 152, 142, or 129 amino acids. For the present work, it was important to determine which of these three AUGs serves as the initiation codon for XGPR T synthesis in animal cells. The plasmids pSV2gpt2 and -3 were constructed to delete the first two AUGs in the XGPR T reading frame (Fig. 2). Each of these deletions completely abolished the capacity to synthesize active XGPR T (Fig. 3), leading us to conclude that the first AUG in pSV2gpt1 is the bona fide initiation codon. If truncated forms of XGPR T were produced by pSV2gpt2 and -3, they lacked enzymatic activity and were therefore undetected.

Structures of the mRNAs. The principle objective of this work was to test the hypothesis that initiation of translation at internal AUGs can occur by a termination-reinitiation mechanism. Accordingly, a series of plasmids was constructed to test the effects of deleting termination codons of reading frames upstream of the coding region for XGPR T. The structure of the pSV2 vectors and their use in the expression of foreign genes in COS cells has been described in detail elsewhere (2). Briefly, the plasmids used in this study use the SV40 early promoter and RNA processing signals to produce mRNAs whose predicted structures are shown in Fig. 2 and 4. To verify that these mRNAs were indeed produced, transcripts were analyzed by both the blotting procedure for mRNAs and the S1 nuclease protection method. Polyadenylated RNA from transfected cells was denatured by glyoxalation, electrophoresed in a 1.5% agarose gel, transferred to nitrocellulose, and annealed with

³²P-labeled *dhfr*- and *gpt*-specific probes. pSVdhfr26, pSV2gpt1, and pSV2dhfr-gpt1 produced mRNA species migrating in positions consistent with their expected sizes of approximately 1,700, 1,850, and 2,600 nucleotides, respectively (Fig. 5). It is important that no significant quantities of species other than the 2,600-nucleotide RNA were produced by the pSV2dhfr-gpt1 recombinant.

S1 nuclease analyses were performed with RNAs from cells transfected with pSV2gpt, pSV2gpt1, pSV2dhfr-gpt1, pSV2dhfr-gpt6, and pSV2dhfr-gpt8. Probes were labeled at their 5' ends with ³²P as illustrated in Fig. 6A. After hybridization and S1 nuclease digestion, the protected fragments were separated by electrophoresis in an alkaline agarose gel (Fig. 6B). The results confirm that each recombinant produced an mRNA species that was contiguous with the probe from the labeled site at the 3' end of the *gpt* sequence all the way to the start point of transcription near the *Bgl*I site in the SV40 early promoter. With pSV2dhfr-gpt1, -6, and -8 this was the only abundant species produced. The pSV2gpt and pSV2gpt1 species, however, produced an additional component of about 0.7 kilobases. The discontinuity of this RNA with the probe mapped to a site well within the XGPR T coding sequence and presumably cannot direct the synthesis of functional enzyme. This species indicates the existence of an S1 nuclease-sensitive site within the XGPR T coding sequence, as was noted by Mulligan and Berg (38) for plasmids containing the *gpt* sequence in the SV40 late region. Since no known splice donors were present upstream from this site in the pSV2

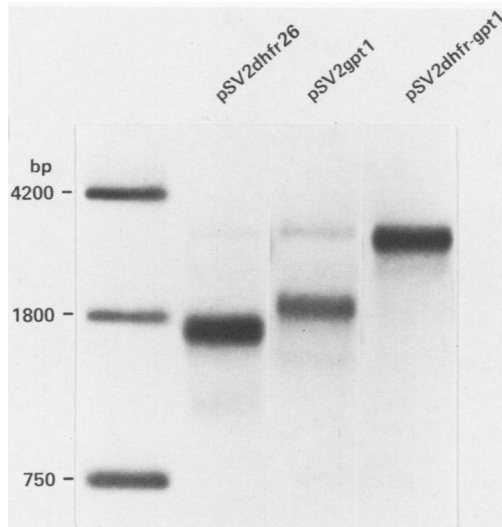


FIG. 5. Blot analysis of *gpt*- and *dhfr*-specific RNAs produced in transfected cells. The pSV2dhfr26 lane was probed with an isolated DNA fragment containing only *dhfr* sequences labeled by nick translation. Similarly, the pSV2gpt1 and pSV2dhfr-gpt1 samples were probed with sequences specific for *gpt*. The leftmost lane contained size markers.

derivatives and since the blotting analysis of mRNAs provided no evidence for a shorter species of this abundance, it is likely that the 0.7-kilobase species represents an S1 nuclease-sensitive site in the RNA-DNA hybrid.

DHFR production. The production of mouse DHFR in transfected cells was measured by the electrophoretic blotting technique described in Materials and Methods. Cells transfected with pSV2dhfr, pSV2dhfr-gpt1, and pSV2dhfr-gpt6 produced mouse DHFR (Fig. 7). Since these recombinants contained nondisrupted DHFR coding sequences, each was expected to produce the authentic 22,000-dalton mouse enzyme. However, in pSV2dhfr-gpt8 the *dhfr* termination codon had been deleted and its reading frame fused to that of *gpt* (Fig. 4). This recombinant did not produce the normal form of mouse DHFR but a larger protein; its electrophoretic mobility indicated a polypeptide of approximately 39,000 daltons. Although its identity has not been firmly established, its size is consistent with its being the predicted DHFR-XGPRT fusion protein. Likewise, pSV2dhfr-gpt13 (Fig. 4) produced a form of DHFR consistent with the addition of 22 amino acids at its carboxyl terminus, as predicted from its DNA sequence. No product related to mouse DHFR could be detected in cells transfected with pSV2dhfr-gpt14, but its nucleotide sequence predicted that it should direct the synthesis of a product three amino acids longer than normal DHFR that probably comigrated with the endogenous monkey enzyme.

Production of XGPRT. A prediction of the termination-

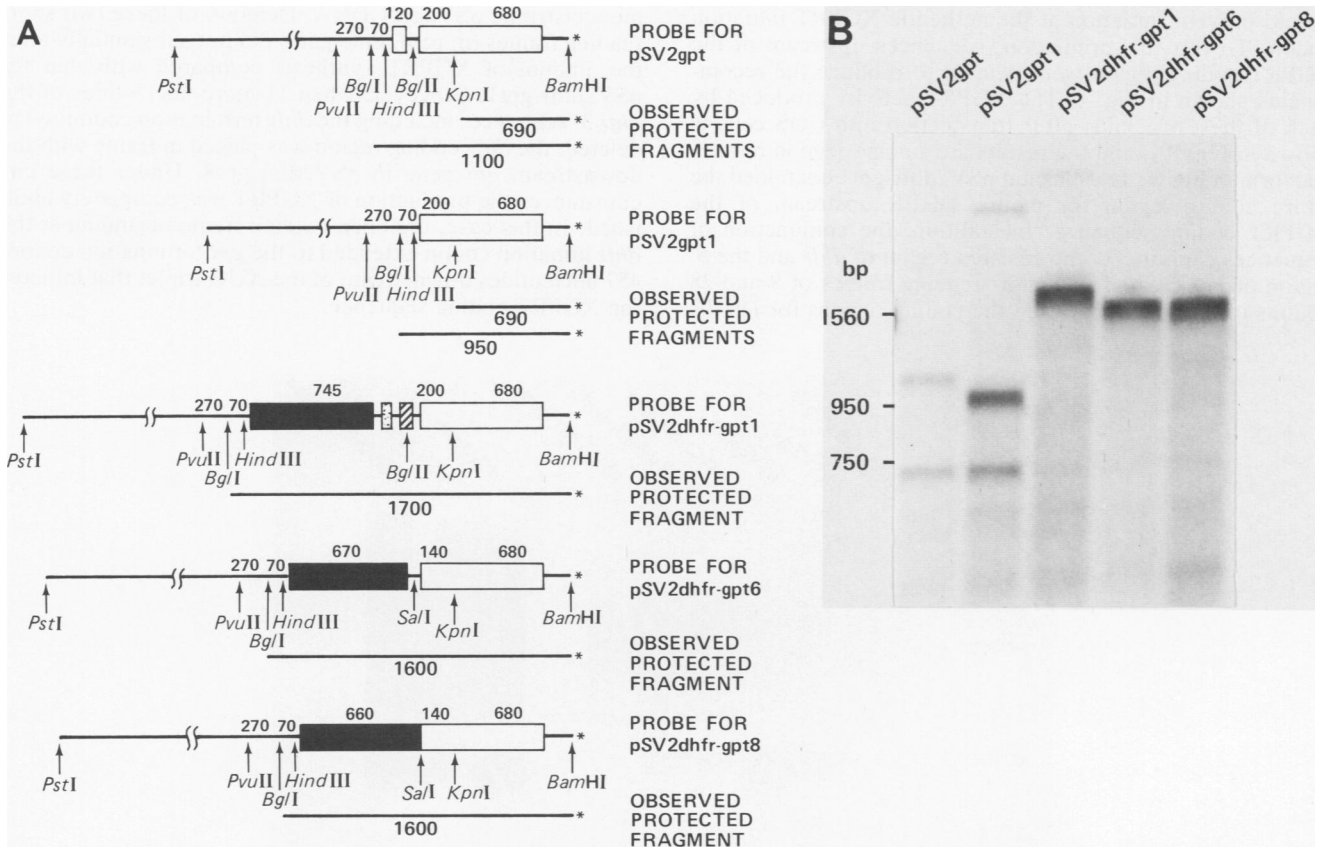


FIG. 6. (A) S1 nuclease analysis of RNAs from transfected cells. The specific end-labeled probes hybridized to the respective RNA samples are indicated with solid lines for vector sequences and boxes (coded as in Fig. 2) to indicate the location of the coding regions. The distances (in nucleotides) between some restriction endonuclease cleavage sites are indicated, as are the lengths of protected fragments. The asterisks indicate the label at the 5' end. (B) After hybridization of the RNA to the DNA probe, samples were treated with S1 nuclease and electrophoresed in alkaline agarose gels. The positions of size markers are shown on the left.

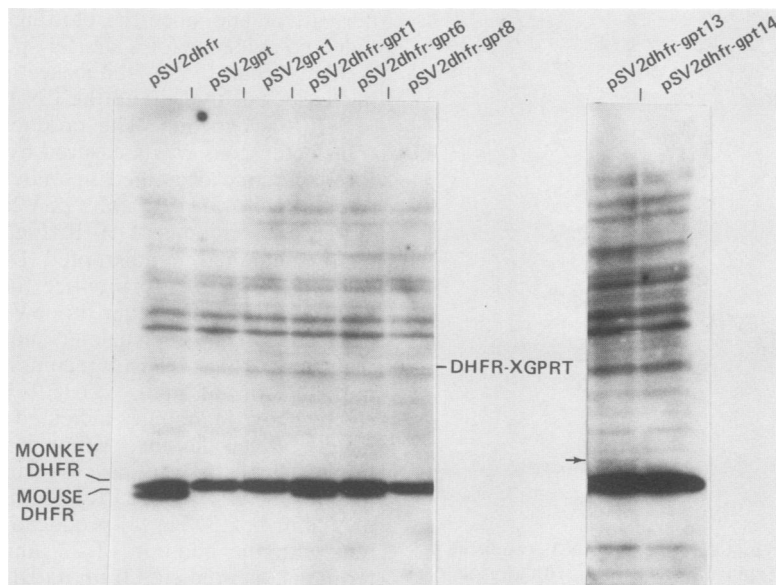


FIG. 7. Blot analysis for DHFR-related proteins in transfected cells. The positions of the endogenous monkey enzyme, present in all samples, and the mouse DHFR are indicated at the left. The position of the putative DHFR-XGPRT fusion protein is indicated. The arrow on the right indicates the elongated form of DHFR produced by pSV2dhfr-gpt13.

reinitiation hypothesis is that deletion of the translation termination codon upstream of the XGPRT coding region should prevent initiation at the authentic XGPRT initiation codon. To test this prediction, sequences upstream of the XGPRT coding region were deleted to produce the recombinants shown in Fig. 4. The XGPRT activity produced by each of these plasmids after transfection into COS cells is shown in Fig. 8A, and the results are summarized in numerical form in Fig. 4. The plasmid pSV2dhfr-gpt1 contained the entire coding region for mouse DHFR upstream of the XGPRT coding sequence. In addition, the conjunction of sequences from the 3' untranslated region of *dhfr* and the 5' region of *gpt* created two short reading frames of 9 and 28 codons in the region between the coding regions for DHFR

and XGPRT. Cells transfected with this plasmid produced only 10 to 20% as much XGPRT as cells transfected with monocistronic pSV2gpt1 DNA. Deletion of these two short reading frames (in pSV2dhfr-gpt6) did not substantially alter the amount of XGPRT synthesis compared with that for pSV2dhfr-gpt1. However, when 11 more nucleotides of the *dhfr* 3' sequence, including the *dhfr* termination codon, were deleted, the *dhfr* coding region was placed in frame with the downstream *gpt* gene in pSV2dhfr-gpt8. Under these circumstances the production of XGPRT was completely abolished. In this case, the open reading frame beginning at the *dhfr* initiation codon extended to the *gpt* termination codon, 457 nucleotides downstream of the AUG triplet that initiates the XGPRT coding sequence.

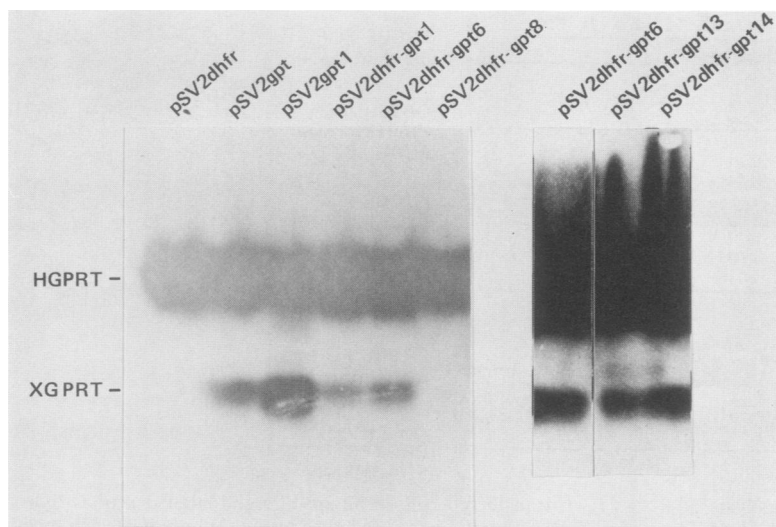


FIG. 8. Electrophoretic assays of XGPRT in extracts of cells transfected with the recombinants whose structures are illustrated in Fig. 4, along with pSV2gpt and pSV2gpt1 for comparison (see Fig. 2). Each track contains extract from about 10^6 cells, the exact amount being determined by reference to β -galactosidase activity as described in Materials and Methods.

To evaluate the effect of termination of DHFR translation at varying distances downstream from the XGPRT initiator, pSV2dhfr-gpt13 and -14 were constructed by the insertion and deletion, respectively, of 4 bp at the *SalI* site that marks the boundary between the *dhfr* and *gpt* sequences in pSV2dhfr-gpt8. The frameshift mutations created by these manipulations caused translation that initiated at the *dhfr* AUG to terminate at a UAA triplet 50 nucleotides downstream of the *gpt* AUG in pSV2dhfr-gpt13 or at a UGA triplet that overlapped the *gpt* AUG in pSV2dhfr-gpt14 (Fig. 4). The results of assays for the ability of these recombinants to produce XGPRT are shown in Fig. 8, and the data are summarized numerically in Fig. 4. Although pSV2dhfr-gpt8 produced no detectable XGPRT activity, pSV2dhfr-gpt13, which contained a termination codon in the *dhfr* reading frame 50 nucleotide downstream of the *gpt* initiation codon, produced XGPRT at about 50% of the level produced by pSV2dhfr-gpt6. When the termination and initiation triplets actually overlapped, as in pSV2dhfr-gpt14, XGPRT was produced at about the same level as by pSV2dhfr-gpt6. Since functional XGPRT was not produced when initiation occurred at the AUG codon downstream of the true XGPRT initiator, it is unlikely that the activity resulted from reinitiation at these sites (see earlier data, Fig. 2 and 3). Thus, the synthesis of functional XGPRT in pSV2dhfr-gpt13 and -14 suggests that ribosomes can "reach back" after translation terminates to reinitiate translation at a site at least 50 nucleotides upstream.

DISCUSSION

The results presented here add to the growing list of examples of translation initiation at internal AUGs. In cells transfected with pSV2dhfr-gpt1, for example, XGPRT synthesis was initiated at an AUG that was about 900 nucleotides from the 5' end of the mRNA molecule and was preceded by 11 AUG triplets. Our analysis of the species of mRNA made from the recombinant plasmids described here indicates that XGPRT is not made from a smaller mRNA in which the XGPRT AUG initiator codon is the most 5' proximal. The very small structural differences between plasmids which produced XGPRT and those which did not also make this an extremely unlikely possibility. Plasmid pSV2dhfr-gpt8 did not promote the synthesis of XGPRT, although it differed from pSV2dhfr-gpt6 by a deletion of only 11 bp from the 3' end of the *dhfr* sequence. Moreover, XGPRT synthesis was restored with both pSV2dhfr-gpt13 and -14, which were derived from pSV2dhfr-gpt8 by the insertion or deletion of 4 bp at the *SalI* site that joins the *dhfr* and *gpt* sequences. These results are difficult to reconcile with an explanation that relies on the existence of an unusual mRNA species. One would have to maintain that a cryptic promoter or RNA processing signal is disrupted by the 11-bp deletion in pSV2dhfr-gpt8 and then reconstructed by a 4-bp insertion in pSV2dhfr-gpt13 and by a 4-bp deletion in pSV2dhfr-gpt14. This argument can also be invoked against the possibility that the failure of pSV2dhfr-gpt8 to produce XGPRT is due to the disruption of a sequence that mediates the direct binding of ribosomes to an internal initiation site.

Another possible explanation of our data is that internal initiation involves direct binding of ribosomes at an internal site but that such binding is inhibited by active translation from an upstream sequence through the internal initiation site. In this view, the role of upstream terminators is to relieve this inhibition by causing the ribosomes to dissociate from the mRNA upstream of the internal initiation site. This

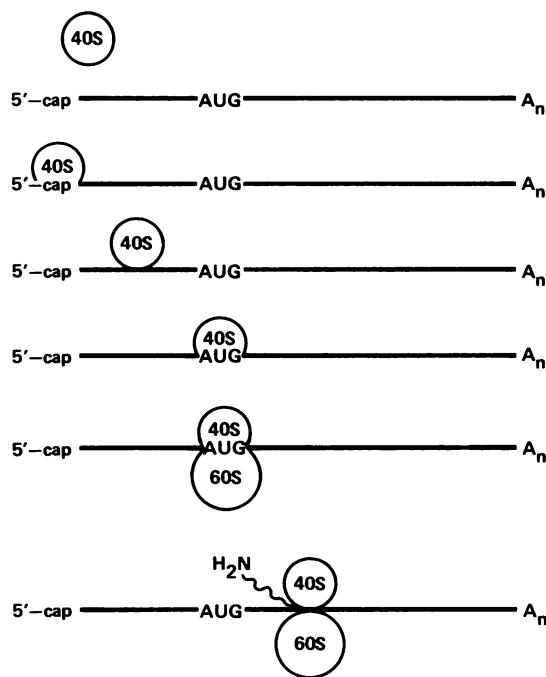


FIG. 9. Schematic illustration of Kozak's scanning hypothesis (20). According to this scheme, the 40S ribosomal subunit initially binds the 5' end of the mRNA molecule in a reaction that is facilitated by the presence of a 5' cap structure. The 40S subunit then migrates along the mRNA until it encounters an AUG triplet, at which point the 60S ribosomal subunit becomes associated and protein synthesis begins.

explanation seems unlikely since quite efficient initiation of XGPRT synthesis occurred even when the two coding regions of the dicistronic mRNA overlapped by 50 nucleotides.

The termination-reinitiation model we propose embodies the basic features of the scanning hypothesis but maintains that after termination the ribosome remains associated with the mRNA long enough to reinitiate at a nearby AUG. According to the scanning hypothesis (Fig. 9), the 40S ribosomal subunit binds the 5' end of the mRNA and migrates until it encounters an initiation codon, at which point the 60S subunit becomes associated and synthesis of the polypeptide chain begins. In one version of our model (Fig. 10), the 60S subunit dissociates after termination of translation and the 40S subunit returns to a scanning mode until it encounters another AUG, at which the initiation process is repeated. Alternatively, the entire 80S complex may remain attached to the mRNA, retaining the capacity to initiate at a nearby AUG. Since virtually nothing is known at present about the immediate fate of eucaryotic ribosomes upon termination, we have no basis by which to judge which is the most likely alternative. These are questions which are not yet fully resolved even for bacterial systems, in which termination-reinitiation is a well-documented phenomenon (35, 40, 41, 47, 51).

Termination-reinitiation is one mechanism by which initiation at internal AUGs might occur. The experiments described here confirm a major prediction of the model, that deletion of the terminator of an upstream coding region can abolish initiation at a downstream site. Thus, when the *dhfr* termination codon is deleted (pSV2dhfr-gpt8), the capacity for initiation at the *gpt* AUG is eliminated. Results with

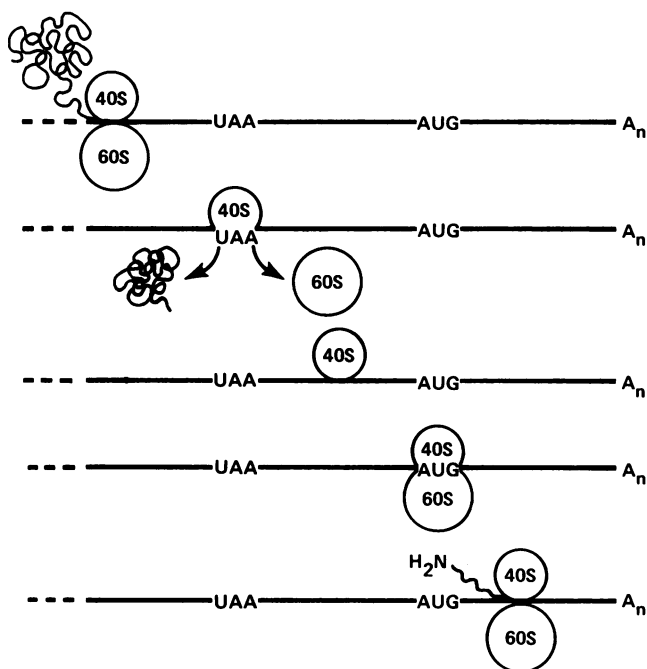


FIG. 10. Termination-reinitiation model. This hypothesis maintains that after termination of translation, the 40S ribosomal subunit may remain associated with the mRNA, regaining the capacity to scan along the mRNA and initiate translation again at the next available AUG.

pSV2dhfr-gpt13 and -14, however, indicate that the reinitiation process does not necessarily require that the termination event occur upstream of the reinitiation site, only that it not occur too far downstream. Our data do not permit a determination of the maximum distance a ribosome may reach back to reinitiate, only that the process is about 50% efficient over a distance of 50 nucleotides and undetectable at 457 nucleotides. This is consistent with other results (43) which show efficient reach-back reinitiation over a distance of 13 but not 137 nucleotides. These results suggest that ribosomes can migrate bidirectionally on the mRNA molecule.

Although mammalian cells apparently possess the ability to initiate translation at internal AUG codons, they rarely use that ability. A recent comparison of the sequences of 180 different naturally occurring mRNAs revealed that only 18 of them had one or more AUGs in their 5' leader sequences (26). An even smaller number have been shown to actually use an upstream AUG for initiation. A possible reason for the apparent selection against the appearance of upstream AUGs is that they decrease the efficiency of translation of the downstream sequence. The dicistronic plasmid pSV2dhfr-gpt6, for example, produced XGPRT at a level only 10 to 20% of that observed for the monocistronic pSV2gpt1. This tendency to reduce translation efficiency has been observed in some other cases as well but is not universally true, since we have encountered at least one example of reinitiation that seems to be 100% efficient (43).

It should be noted that the model we have proposed does not exclude the possibility that other mechanisms may be involved in some cases. Indeed, termination-reinitiation cannot explain all the known examples of internal initiation, since some mRNAs contain upstream AUGs with no in-frame terminator before or soon after the bona fide initiation

codon. Although many of the known examples of mRNAs that initiate translation at internal AUGs conform to the requirements of the termination-reinitiation model, some do not, and in these cases the relaxed scanning or direct internal binding models must be invoked. The experiments described here and in the accompanying paper (43) suggest, however, that termination-reinitiation is an important mechanism of internal initiation of translation in many cases.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant R01-GM13235 from the National Institutes of Health. D.S.P. was a postdoctoral fellow of the Jane Coffin Childs Fund for Medical Research.

LITERATURE CITED

1. Alwine, J. C., D. J. Kemp, and G. R. Stark. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyl oxymethyl paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* **74**:5350-5354.
2. Berg, P. 1981. Dissections and reconstructions of genes and chromosomes. *Science* **213**:296-303.
3. Berk, A. J., and P. Sharp. 1978. Spliced early mRNAs of simian virus 40. *Proc. Natl. Acad. Sci. USA* **75**:1274-1278.
4. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**:1513-1523.
5. 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 Elb mRNA of human Ad12 and Ad5 codes for two tumor antigens starting at different AUG triplets. *Cell* **27**:121-131.
6. Burnette, W. N. 1981. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* **112**:195-203.
7. Clerx-van Haaster, C., H. Akashi, D. Auperin, and D. Bishop. 1982. Nucleotide sequence analyses and predicted coding of bunyavirus genome RNA species. *J. Virol.* **41**:119-128.
8. Davis, B. J. 1964. Disc electrophoresis II: method and application to human serum proteins. *Ann. N.Y. Acad. Sci.* **121**:404-427.
9. Fromm, M., and P. Berg. 1982. Deletion mapping of DNA regions required for SV40 early region promoter function *in vivo*. *J. Mol. Appl. Genet.* **1**:457-481.
10. Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. V. Choudary, P. Lebowitz, and S. Weissman. 1978. The 5' terminal leader sequence of late 16S mRNA from cells infected with simian virus 40. *J. Biol. Chem.* **253**:3643-3647.
11. Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman. 1978. Heterogeneity and 5' terminal structures of the late mRNAs of simian virus 40. *J. Mol. Biol.* **126**:813-846.
12. Gluzman, Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**:175-182.
13. Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo. 1981. Translation initiation in prokaryotes. *Annu. Rev. Microbiol.* **35**:365-403.
14. Graham, F. L., and A. J. Van der Eb. 1973. A new technique for assaying the infectivity of adenovirus 5 DNA. *Virology* **52**:456-67.
15. Hagenbuchle, O., M. Tosi, U. Schibler, R. Bovey, P. K. Wellaver, and R. A. Young. 1981. Mouse liver and salivary gland α -amylase mRNAs differ only in 5' non-translated sequences. *Nature (London)* **289**:643-646.
16. Hall, C. V., P. E. Jacob, G. M. Ringold, and F. Lee. 1983. Expression and regulation of *Escherichia coli* lacZ gene fusions in mammalian cells. *J. Mol. App. Genet.* **2**:101-109.
17. Hendy, G., H. Kronenberg, J. Potts, Jr., and A. Rich. 1981. Nucleotide sequence of cloned cDNAs encoding human parathyroid hormone. *Proc. Natl. Acad. Sci. USA* **78**:7365-7369.

18. Kastelein, R. A., E. Remaut, W. Fiers, and J. van Duin. 1982. Lysis gene expression of RNA phage MS2 depends on a frameshift during translation of the overlapping coat protein gene. *Nature (London)* **295**:35-41.
19. Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. van der Werf, C. W. Anderson, and E. Wimmer. 1981. Primary structure gene organization and polypeptide expression of polio virus RNA. *Nature (London)* **291**:547-553.
20. Kozak, M. 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**:1109-1123.
21. Kozak, M. 1979. Inability of circular mRNA to attach to eucaryotic ribosomes. *Nature (London)* **280**:82-85.
22. 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.
23. Kozak, M. 1980. Role of ATP in binding and migration of 40S ribosomal subunits. *Cell* **22**:459-467.
24. Kozak, M. 1980. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell* **19**:79-90.
25. Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiation codon by eucaryotic ribosomes. *Nucleic Acids Res.* **9**:5233-5252.
26. Kozak, M. 1983. Comparison of initiation of protein synthesis in prokaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**:1-45.
27. Kozak, M., and A. J. Shatkin. 1978. Migration of 40S ribosomal subunits on messenger RNA in the presence of Edeine. *J. Biol. Chem.* **253**:6568-6577.
28. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
29. Lomedico, P. T., and S. J. McAndrew. 1982. Eucaryotic ribosomes can recognize preproinsulin initiation codons irrespective of their position relative to the 5' end of mRNA. *Nature (London)* **299**:221-226.
30. Mardon, G., and H. E. Varmus. 1983. Frameshift and intragenic suppressor mutations in a Rous sarcoma provirus suggests *src* encodes two proteins. *Cell* **32**:871-879.
31. Marsden, H. S., L. Haarr, and C. M. Preston. 1983. Processing of herpes simplex virus protein and evidence that translation of thymidine kinase mRNA is initiated at three separate AUG codons. *J. Virol.* **46**:434-445.
32. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
33. McDonnell, M. W., M. N. Simon, and F. W. Studier. 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J. Mol. Biol.* **110**:119-146.
34. McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74**:4835-4838.
35. Michels, C. A., and D. Zipser. 1969. Mapping of polypeptide reinitiation sites within the β -galactosidase structural gene. *J. Mol. Biol.* **41**:341-347.
36. Miller, J. H. 1972. Experiments in molecular genetics, p. 352-355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. Mulligan, R. C., and P. Berg. 1980. Expression of a bacterial gene in mammalian cells. *Science* **209**:1422-1427.
38. Mulligan, R. C., and P. Berg. 1981. Factors governing the expression of a bacterial gene in mammalian cells. *Mol. Cell. Biol.* **1**:449-459.
39. Mulligan, R. C., B. H. Howard, and P. Berg. 1979. Synthesis of rabbit β -globin in cultured monkey cells following transfection with SV40 β -globin recombinant. *Nature (London)* **227**:108-114.
40. Napoli, C., L. Gold, and B. S. Singer. 1981. Translational reinitiation in the rIB1B cistron of bacteriophage T4. *J. Mol. Biol.* **149**:433-449.
41. Oppenheim, D. S., and C. Yanofsky. 1980. Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics* **95**:785-795.
42. Parker, B., and G. Stark. 1979. Regulation of simian virus 40 transcription: sensitive analysis of the RNA species present early in infections of virus or viral DNA. *J. Virol.* **31**:360-369.
43. Peabody, D. S., S. Subramani, and P. Berg. 1986. Effect of upstream reading frames on translation efficiency in simian virus 40 recombinants. *Mol. Cell. Biol.* **6**:2704-2711.
44. Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson. 1980. The evolution of genes: the chicken preproinsulin gene. *Cell* **20**:555-566.
45. Renart, J., J. Reiser, and G. R. Stark. 1979. Transfer of proteins from gels to diazobenzyloxymethyl paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proc. Natl. Acad. Sci. USA* **76**:3116-3120.
46. Rigby, P. W., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**:237-251.
47. Sarabhai, A., and S. Brenner. 1967. A mutant which reinitiates the polypeptide chain after chain termination. *J. Mol. Biol.* **27**:145-162.
48. Schumperli, D., K. McKenney, D. A. Sobieski, and M. Rosenberg. 1982. Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell* **30**:865-871.
49. Sherman, F., J. W. Stewart, and A. M. Schweingruber. 1980. Mutants of yeast initiating translation of iso-1-cytochromes within a region spanning 37 nucleotides. *Cell* **20**:215-222.
50. Southern, P. J., and P. Berg. 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* **1**:327-341.
51. Steege, D. A. 1977. 5'-Terminal nucleotide sequence of *Escherichia coli* lactose repressor mRNA: features of translational initiation and reinitiation sites. *Proc. Natl. Acad. Sci. USA* **74**:4163-4167.
52. Subramani, S., R. C. Mulligan, and P. Berg. 1981. Expression of the mouse dihydrofolate reductase complementary deoxyribonucleic acid in simian virus 40 vectors. *Mol. Cell. Biol.* **1**:854-864.
53. Swanstrom, R., H. E. Varmus, and J. M. Bishop. 1982. Nucleotide sequence of the 5' noncoding region and part of the *gag* gene of Rous sarcoma virus. *J. Virol.* **41**:535-541.
54. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77**:5201-5205.
55. Vogeli, G., H. Ohkubo, M. E. Sobel, Y. Yamada, I. Pastam, and B. de Crombrughe. 1981. Structure of the promoter for chicken $\alpha 2$ type I collagen gene. *Proc. Natl. Acad. Sci. USA* **78**:5334-5338.
56. Weaver, R. F., and C. Weissmann. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure. *Nucleic Acids Res.* **75**:1175-1193.
57. Wengler, G., G. Wengler, and H. J. Gross. 1979. Replicative form of Semliki Forest virus RNA contains an unpaired guanosine. *Nature (London)* **282**:754-756.
58. Yoo, O. J., C. T. Powell, and K. L. Agarwal. 1982. Molecular cloning and nucleotide sequence of full length cDNA coding for porcine gastrin. *Proc. Natl. Acad. Sci. USA* **79**:1049-1053.