Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Polycistronic Animal Virus mRNAs

CHARLES E. SAMUEL

Section of Biochemistry and Molecular Biology
Department of Biological Sciences
University of California
Santa Barbara, California 93106

Of central importance to the understanding of biological processes is the elucidation of the mechanisms by which the products of genes are produced in a regulated fashion. The production of a particular protein in an animal cell may be regulated, in principle, at any of the many steps of mRNA formation (*1*, *2*) and translation (*3*). At the level of translation, most animal virus and animal cell mRNAs analyzed in detail appear to be functionally monocistronic; that is, a single primary translational product is produced from a single mature mRNA transcript (*4*, *5*). In striking contrast, many bacteriophage and bacterial cell mRNAs are polycistronic; that is, multiple primary translational products are specified by a single mature mRNA transcript using independent initiation and/or termination codons for protein synthesis (*4*).

Although the organization of animal virus genomes and the mechanisms of viral genome replication and transcription can differ substantially among the major families of animal viruses, the translation of viral mRNA by the host's protein-synthesizing machinery is a central step common to all viral multiplication strategies (*6*). Historically, molecular genetic studies of animal viruses have yielded important new insights into our understanding of the mechanisms by which

Progress in Nucleic Acid Research and Molecular Biology, Vol. 37

proteins are produced in animal cells (*1–3, 7–9*). Now, detailed analyses have revealed that, in some cases, the synthesis of animal virus proteins involves the translation of functionally polycistronic mRNAs to yield multiple primary translation products. Such fundamentally important conclusions have been made possible, in part, by the availability of nucleotide and peptide sequences of a number of animal virus genomes and their encoded products, together with complete cDNA clones of animal virus genomes and mRNAs that can be manipulated by recombinant DNA techniques to generate synthetic mRNAs or mutated versions of authentic mRNAs. Careful analyses of various viral mRNAs have now made it clear that some viral mRNAs are indeed recognized and translated as functionally polycistronic mRNAs by the protein-synthesizing machinery of animal cells. Furthermore, many different biochemical mechanisms may exist in animal cells to permit the expression of functionally polycistronic viral mRNAs.

In this essay, I review some observations concerning the natural occurrence and structural organization of polycistronic animal virus mRNAs, and the mechanisms by which they may be translated to yield two or more unique polypeptide products.

I. Natural Occurrence of Polycistronic Animal Virus mRNAs

Most animal virus mRNAs characterized so far are functionally monocistronic. However, functionally polycistronic animal virus mRNAs also exist. Initiation of translation in monocistronic viral mRNAs most commonly occurs at AUG and, furthermore, at the 5′-proximal AUG (*5–7, 10, 11*). In most polycistronic viral mRNAs, initiation of translation of both the 5′-proximal, upstream cistron and the internal, downstream cistron(s) likewise occurs at an AUG codon (*12–79* inclusive). Animal viruses encoding polycistronic mRNAs in which translation-initiation occurs alternatively at one or more AUG initiation sites include members of several virus families that utilize a variety of different replication strategies as parts of their life cycles (Table I). They include: viruses with DNA genomes and viruses with RNA genomes; viruses with circular genomes and viruses with linear genomes; viruses whose genomes are constituted by a single piece of nucleic acid as well as viruses with segmented genomes; and viruses that utilize the cell nucleus as the site for mRNA biogenesis as well as viruses whose mRNA is synthesized in the cytoplasm.

The identification and characterization of polycistronic mRNAs have been greatly facilitated by the availability of nucleotide se-

quences derived from cDNA clones of viral mRNAs. From these cDNA sequences, it has been possible to identify potential initiation and termination codons and open reading frames capable of encoding polypeptide products of a predicted size. The availability of these cDNA clones has also permitted the construction of various types of vectors for translational studies, including *in vivo* expression vectors and *in vitro* transcription vectors. Eukaryotic expression vectors using various high-efficiency promoters have made possible the analysis of protein synthesis *in vivo* using constructions that encode chimeric mRNAs possessing defined mutations in regions of potential regulatory importance, and/or chimeric mRNAs specifying "reporter" genes that facilitate assays for translational expression. *In vitro* transcription vectors with SP6 and T7 promoters have made possible the efficient synthesis of wild-type and mutant mRNAs that can be examined for their translational activity with a variety of different cell-free protein-synthesizing systems. In addition, the ability to identify and quantitate the expression within infected or transfected cells of proteins predicted from cDNA sequences has been facilitated by the production of antibodies against synthetic peptides having the amino-acid sequences deduced from open reading frames revealed by cDNA sequence data, and antibodies produced against recombinant fusion proteins produced in bacteria. Primer extension analysis, S1-protection analysis, and RNA "blotting" and hybridization have been used to demonstrate the presence of mRNAs *in vivo* that possess a particular 5′ region in which potential translation initiation codons are localized. Thus, much experimental evidence exists for the occurrence of functionally polycistronic mRNAs, specific examples of which are considered in the following section, which emphasizes their possible mechanisms of expression.

II. Mechanisms of Expression of Polycistronic Viral mRNAs

A. Initiation by 5′ Ribosome-Binding and Leaky Scanning

The scanning model provides a conceptual framework for understanding the process of initiation of mRNA translation in animal cells. According to the scanning model, a 40-S ribosomal subunit binds at or near the 5′ end of an mRNA and advances linearly until an initiation codon in a favorable context is reached, at which point assembly of a complete 80-S ribosomes takes place and initiation of polypeptide synthesis occurs (*5, 10, 12, 80*). The efficiency of initiation of transla-

TABLE I
POLYCISTRONIC ANIMAL VIRUS mRNAs

Virus	mRNA	Products (5′; 3′)	Proposed expression mechanism[a]	Reference
I. DNA Viruses				
A. Adenoviridae				
Ad12, Ad5	E1b	p19 (p21); p54 (p55)	1c	*27*
Ad2	E3a	p6.7; gp19	1c	*28*
Ad2	DNA pol	p120; p62	2	*56*
B. Herpesviridae				
Herpes simplex virus	Thymidine kinase	p43; p39; p38	1a	*16, 17*
C. Papovaviridae				
Simian virus 40	19-S late	VP2; VP3	1a	*13–15*
	16-S late	LP1; VP1	1, 5?	*78, 135*
	19-S early	SELP; T, t	1, 5?	*69*
D. Parvoviridae				
AAV2	B/C	p70 (B); p60 (C)	1b	*24, 25*
B19	VP1	VP1; VP2	5?	*70*
II. RNA Viruses				
A. Bunyaviridae				
Bunyavirus	Segment S	N; NS_s	1c	*37–39*
B. Coronaviridae				
Mouse hepatitis virus	Gene 5	p13; p9.6	1c	*49*
C. Orthomyxoviridae				
Influenza B	Segment 6	NA; NB	1c?	*47, 48*

D. Hepadnaviridae				
Hepatitis B virus	Pregenomic	core; *pol*	2	*141*
E. Paramyxoviridae				
Sendai virus	P/C	P; C, C'; Y, Y'	1b, 1c	*21–23*
Parainfluenza	P/C	P; C	1c	*34–35*
Measles	P/C	P; C	1c	*36*
Newcastle disease virus	P/C	P; (p38,p29) C	1a	*20*
F. Picornaviridae				
Poliovirus	Genomic	Polyprotein	2	*101, 104*
Encephalomyocarditis virus	Genomic	Polyprotein	2	*102, 105*
G. Reoviridae				
Reovirus T1, T3	Segment 1	$\sigma 1$; $\sigma 1_{ns}$ (p12, p14)	1c	*40–46*
Rotavirus SA11	Segment 9	(VP7) p38; p35	1a	*19*
H. Retroviridae				
Rous sarcoma virus	Genomic	*gag; pol*	3	*61*
	Subgenomic	*src*	5?	*72*
Human T-cell leukemia virus	Genomic	*gag; pol*	3	*62*
	Subgenomic	*rex; tax*	1a	*18*
Human immunodeficiency virus	Genomic	*gag; pol*	3	*64*
Mouse mammary tumor virus	Genomic	*gag*-X; *pro* *x-pro; pol*	3	*65, 66*
Bovine leukemia virus	Genomic	*gag; pol*	3	*63*
Murine leukemia virus	Genomic	*gag; pol*	4	*67*
Feline leukemia virus	Genomic	*gag; pol*	4	*68*
I. Rhabdoviridae				
Vesicular stomatitis virus	NS	NS (P); p7	2	*55*

[a] Candidate expression mechanisms. 1, Leaky scanning: 1a, alternative in-frame AUG initiation codon; 1b, alternative in-frame non-AUG initiation codon; 1c, alternative AUG initiation codon in a different, overlapping reading frame. 2, Initiation by internal binding of ribosomes. 3, Ribosomal frameshifting during elongation. 4, Suppression of in-frame UAG termination codon. 5, Reinitiation following termination.

tion is thought to be modulated, at least in part, by the position and context of the initiation codon (*5, 10, 12, 80–83*). If the 5′-proximal initiation codon is in an optimal context, initiation at that codon is efficient. If the context of the 5′-proximal initiation codon is suboptimal, initiation at that codon is inefficient. Although a fraction of the scanning ribosomes may indeed initiate translation at a suboptimal 5′-proximal initiation codon, a portion of the 40-S ribosomal subunits would presumably bypass the suboptimal codon and initiate downstream at an internal AUG codon positioned in a more favorable context. This "leaky scanning" model provides a mechanism for the initiation of translation at internal initiation codons that would be present in polycistronic mRNAs (*12*).

Considerable evidence consistent with the leaky scanning model for initiation of translation has accumulated. This evidence includes the unique 5′-terminal $m^7G(5')ppp(5')N$ . . . cap structure characteristic of most animal virus and cell mRNAs, a structure that often significantly stimulates translation in many types of cell-free protein-synthesizing systems (*84–86*); the inability of ribosomes to bind and translate synthetic circular mRNAs (*87, 88*); the ability of 40-S ribosomal subunits to migrate along mRNA (*89, 90*); the inhibitory effect on translation-initiation of synthetic secondary structures inserted upstream from an "authentic" initiation codon (*91–95*); and the inhibitory effect on translation of AUG initiation codons inserted upstream from an "authentic" initiation codon (*83, 96–99*).

By inspection of sequences near the 5′ ends of animal cell and viral mRNAs (*5, 11, 80, 81*) and by site-directed mutagenesis of a cDNA copy of a cellular gene (*82, 83*), two nucleotide postions relative to the initiation codon were found to have a dominant effect on the efficiency of initiation: position −3 (three nucleotides upstream from the initiation codon), and position +4 (immediately following the initiation codon). A purine in position −3 and a guanine in position +4 were optimal for efficient initiation at 5′-proximal AUG codons. Furthermore, the efficiency of initiation was, to a large extent, modulated by the context—that is, the −3/+4 nucleotide positions flanking the 5′-proximal AUG—rather than by the relative distance of the AUG from the 5′ end of the mRNA (*5, 10–12, 80–83*).

For several of the polycistronic animal virus mRNAs, the initiation codon of the 5′-proximal upstream cistron is present in a suboptimal context with regard to the −3/+4 flanking nucleotides, whereas the initiation codon of the internal downstream cistron is often, but not always, present in an optimal −3/+4 context. Thus, some of the scanning 40-S ribosomal subunits would presumably be able to "leak" past the suboptimal initiation site of the upstream cistron and con-

tinue scanning until reaching the optimal initiation site of the downstream cistron, present in either the same or a different reading frame. The relative efficiency of utilization of the upstream and downstream cistron initiation codons by ribosomes scanning the same mRNA is of possible regulatory importance, as certain polycistronic mRNAs appear to be organized so that the relative abundance of the proteins encoded by different cistrons correlates with the relative strength of the nucleotides flanking the initiation sites of the cistrons.

One mechanism by which one gene can encode two forms of a protein that differ only in the amino-terminal region is the production of one mRNA transcript from which translation initiation can occur alternatively at one of two in-frame initiation codons. A number of animal virus genes encode mRNA species in which different in-frame AUG initiation sites are utilized in a manner such as to generate related protein products that differ in their amino-terminal sequences. For example, the following viral polycistronic mRNAs each encode two or more proteins in the same reading frame: simian virus 40 (SV40) late 19-S mRNAs encoding capsid proteins VP2 and VP3 (*13–15*); the herpes simplex virus thymidine kinase gene mRNA encoding 43-kDa, 39-kDa, and 38-kDa proteins (*16, 17*); the human T-cell leukemia virus subgenomic mRNA encoding the *rex* and *tax* nonstructural regulatory proteins (*18*); the rotavirus segment-9 mRNA encoding 38-kDa and 35-kDa VP7 capsid glycoproteins (*19*); and the Newcastle disease virus mRNA encoding capsid phosphoprotein P and nonstructural 38- and 29-kDa C-like nonstructural proteins (*20*). In each of these cases (*13–20*), the upstream or 5′-proximal AUG initiation site is in a suboptimal −3/+4 context relative to the downstream AUG initiation site.

The relative translational efficiency of the upstream and downstream cistrons and the abundance of their encoded products may largely be determined by the −3/+4 sequence context of these viral mRNAs. For example, the less abundant VP2 capsid protein of SV40 is synthesized from the 5′-proximal AUG, which is in a suboptimal context, and the 3-fold more abundant VP3 capsid protein is synthesized from a downstream AUG positioned in a more favorable −3/+4 flanking nucleotide context. The results of genetic and biochemical studies of deletion and insertion mutants of SV40, and studies of SV40 transient expression vectors utilizing the bacterial chloramphenicol acetyltransferase (CAT; EC 2.3.1.28) as a reporter activity, provide strong evidence consistent with the conclusion that both VP2 and VP3 are synthesized from each of the alternatively spliced late 19-S mRNAs in a regulated manner by independent initiation of translation via a leaky scanning mechanism (*13–15*).

The Sendai P/C mRNA (*21–23*), the adeno-associated virus B/C mRNA (*24, 25*), and the human c-*myc* mRNA (*26*) also each encode two or more proteins in the same reading frame. Different in-frame initiation sites are utilized by these mRNAs to generate two protein products that differ in their N-terminal sequence. However, in these cases, the most upstream 5′-proximal initiation site is not AUG, but rather is either ACG (*21–25*) in the Sendai P/C and AAV B/C mRNAs, or CUG (*26*) in the c-*myc* mRNA. These 5′-proximal non-AUG initiation codons are in a favorable −3/+4 flanking nucleotide context. However, as discussed under Section II, C, the non-AUG codons are not as efficient as is AUG for initiation, and thus may possibly facilitate a leaky scanning of the 40-S subunit and subsequent initiation at the downstream AUG initiation codon at a frequency sufficient to permit the required levels of synthesis of the encoded proteins.

Several polycistronic viral mRNAs encode two proteins in different, overlapping reading frames. Examples include: the adenovirus (Ad12 and Ad5) 2.2-kb E1b mRNA encoding the 19/21-kDa and 54/55-kDa tumor antigens (*27*); the adenovirus E3a mRNA encoding a 6.7-kDa protein and a 19-kDa glycoprotein (*28*); the paramyxovirus P/C mRNAs of Sendai virus (*21–23, 29–33, 140*), parainfluenza virus 3 (*34, 35*) and measles virus (*36*), each of which encodes the phosphoprotein P and one or more nonstructural proteins designated C, C′, Y, and Y′; the bunyavirus-S segment mRNA encoding the nucleoprotein N and the nonstructural protein NS_s (*37–39*); and the reovirus segment-S1 mRNA encoding the minor capsid protein sigma-1 and the nonstructural protein sigma-1_{NS} (*40–46*). The upstream AUG initiation site in each of these polycistronic viral mRNAs possessing overlapping reading frames is in a suboptimal −3/+4 context, and the relative abundance of the upstream and downstream cistron products may be controlled by the −3/+4 context of their initiation codons, both *in vivo* and *in vitro*. For example, the adenovirus E3 downstream AUG codon is strong and the downstream cistron gp19K protein is the most abundantly synthesized E3 protein *in vivo* (*28*); the synthesis of the Sendai virus C-protein is about five times more efficient than that of P protein and the P/C mRNA initiation codon for C protein synthesis has an adenine at −3 whereas the P protein AUG has a cytosine at −3 (*21, 32*); and the reovirus S1 mRNA 5′-proximal AUG encoding the less abundant sigma-1 protein is in a weak −3/+4 context as compared to the internal AUG encoding sigma-1_{NS}, which is in a strong context and is more efficiently synthesized *in vitro* than is sigma-1 (*40–44, 103*).

Although the apparent efficiencies of utilization of the AUG codons for the first and second cistrons of polycistronic RNAs expressed *in vivo* or translated *in vitro* often agree with the predictions of the leaky scanning model when the context of the −3/+4 flanking nucleotides is considered, exceptions do exist. For example, the influenza virus B segment 6 NB/NA mRNA encodes two glycoproteins in different overlapping reading frames, the neuraminidase NA protein and a nonstructural glycoprotein NB (*47, 48*). The −3/+4 context of the 5′-proximal AUG initiation site used for NB synthesis is more favorable than the context of the downstream AUG initiation codon used for NA synthesis, yet NA and NB accumulate to approximately equal amounts in infected cells (*48, 138*). Furthermore, mutations in the sequence immediately around the 5′-proximal AUG codon do not make a large difference in the amounts of NB and NA that accumulate in transfected cells, but when the 5′ AUG is displaced from its normal position it becomes efficient at preventing downstream initiation events (*138*). The coronavirus gene 5 mRNA of mouse hepatitis virus (MHV) also possesses two open reading frames. Although MHV polypeptides corresponding in size to both gene-5 mRNA reading frames are synthesized *in vitro* from pGEM transcripts, so far only the second open reading frame product has been shown to be expressed within MHV-infected cells, and the −3/+4 context of the AUG initiation site for this reading frame is not optimal (*49–51*).

The analyses of certain mutant mRNAs—for example, reovirus s1 and s4 mRNAs (*52*), influenza virus NB/NA mRNA (*138*), avian retrovirus mRNAs (*53*), and hepatitis-B surface antigen mRNA (*54*)—suggest that nucleotides other than the previously identified consensus context nucleotides flanking the initiator AUG codon at the −3 (purine) and +4 (guanine) positions are also important and perhaps even play a dominant role in determining the efficiency of translation in animal cells. Conceivably, mRNA primary or higher ordered structures responsible for the differential ability of mRNAs to interact with mRNA binding initiation factors such as eIF-4A and eIF-4F may, in some cases, be a major determinant of the translational efficiency of 5′-capped mRNAs (*86, 100*). Thus, the selection of an "authentic" initiation codon with an appropriate efficiency may well be the consequence of collective contributions of many parameters, including both *cis*-acting sequences such as the −3/+4 context nucleotides, and *trans*-acting components such as protein synthesis-initiation factors that may discriminate among different mRNAs by virtue of different mRNA binding affinities.

B. Initiation of Translation by Internal Binding of Ribosomes

Binding of 40-S ribosomal subunits to internal sequences within the body of a polycistronic mRNA, rather than binding to the 5′ end of the mRNA as proposed by the modified scanning model, would provide a mechanism by which translation-initiation could occur from initiation codons positioned downstream from the 5′-proximal AUG codons. Among the best evidence for internal initiation of translation is that obtained with picornavirus mRNAs, both *in vitro* and *in vivo* (*101, 102*).

Picornaviruses possess nonsegmented, plus-strand RNA genomes of about 7500 nucleotides (*6*). Although picornavirus mRNAs appear functionally monocistronic in that a single large polypeptide precursor is synthesized that undergoes posttranslational proteolytic cleavages to generate the mature protein products (*9*), studies with poliovirus (*101, 104*) and encephalomyocarditis (EMC) virus (*102, 105*) suggest that initiation of translation occurs by a cap-independent mechanism that involves internal binding of ribosomes.

The 5′-noncoding region of poliovirus mRNA is unusually long, about 750 nucleotides, and contains several AUG codons upstream from the major initiator AUG located at nucleotide 743 (*106–109*). Unlike most animal virus mRNAs, poliovirus mRNA does not contain a 5′-terminal $m^7G(5')ppp(5')N$ . . . cap structure. Rather, poliovirus mRNA terminates in pUp . . . and is translated by a cap-independent mechanism (*110*). Furthermore, the initiation of translation of 5′-capped cellular and viral mRNAs is inhibited in poliovirus-infected cells. The inhibition mechanism appears to involve a proteolytic inactivation of the p220 subunit of eIF-4F, the cap-binding complex initiation factor required for ribosomes to bind to the 5′ terminus and subsequently to initiate the translation of capped mRNAs (*110, 111*).

Biochemical and genetic analysis of poliovirus mutants generated using a cDNA copy of the viral genome revealed regions within the 5′-noncoding region of poliovirus mRNA important for the efficient translation of the viral mRNA (*101, 112–115*). Surprisingly, mutational analysis of the seven upstream AUG codons present in the 5′-noncoding region of poliovirus mRNA has revealed that the upstream short open reading frames are not essential for virus replication and do not act as barriers to the translation of poliovirus mRNA (*136*). By contrast, analysis of a series of poliovirus deletion mutants identified a functional *cis*-acting element within the 5′-untranslated sequences of poliovirus mRNA that enables it to translate in a cap-independent

manner (*112, 113, 115*). The major determinant of the polio cap-independent translational element maps between nucleotides 320 and 631 from the 5′ end of the poliovirus mRNA and is functional both *in vivo* (*113*) and *in vitro* (*112, 113*). Translation *in vitro* of SP6 transcripts containing additional deletions in the 5′-noncoding region more narrowly focused the *cis*-acting element responsible for cap-independent translation, as measured in a mixed rabbit reticulocyte–HeLa system, to a 60-nucleotide sequence located between positions 567 and 627 (*115*). The poliovirus 5′-noncoding mRNA sequences responsible for the cap-independent synthesis of poliovirus polypeptides can also confer cap-independent translation on heterologous chimeric mRNAs encoding either the bacterial CAT or the herpes virus thymidine kinase as a reporter enzymic activity (*112, 113*). Expression of the reporter enzyme from the chimeric mRNA is extensively augmented by poliovirus-mediated inhibition of cap-dependent protein synthesis.

The cap-independent translation-initiation on poliovirus (*101, 104, 115*) and EMC virus (*102, 105*) mRNAs appears to occur by a mechanism that involves binding of ribosomes to an internal sequence within the 5′-noncoding region. A bicistronic plasmid containing the herpes simplex thymidine kinase gene as the first cistron and the CAT gene as the second cistron does not express the first cistron in poliovirus-infected COS cells but does express the second cistron when the poliovirus 5′-untranslated region is inserted as the intercistronic spacer (*101*). The second cistron is not expressed when the CAT 5′-untranslated region is the intercistronic spacer (*101*). Deletion analysis of the poliovirus 5′-untranslated region suggests that the internal ribosome binding site occupies several hundred nucleotides located between nucleotides 140 and 630 of the poliovirus 5′-untranslated region (*101*). It is unclear whether internal ribosome binding is directed toward a specific AUG or whether the internal binding is followed by scanning. However, because the introduction of a hairpin secondary structure ($\Delta G^\circ = -30$ kcal/mol) at position 631 of the polio RNA dramatically inhibits translation initiation from the downstream AUG, it has been proposed that the ribosomes, following internal binding, are translocated by scanning until they reach the initiator AUG located at position 745 of the poliovirus mRNA (*101*).

Internal ribosome binding can also occur to picornavirus mRNAs *in vitro* (*131*). In HeLa cell extracts, internal binding to the 5′-noncoding region of poliovirus mRNA in a bicistronic context is independent of an upstream open reading frame (*139*). Data obtained

with EMC virus RNA are also consistent with an internal binding of ribosomes to mRNA that is followed by scanning toward the initiator AUG site (*102, 105*). Hybridization of complementary cDNA fragments to different sites within the first 338 nucleotides of the 5′-noncoding region of EMC RNA did not affect translation of the viral mRNA *in vitro,* whereas the binding of cDNA fragments to eight different sites located between nucleotides 450 and the initiator AUG codon at postion 834 caused high degrees of translation inhibition in reticulocyte lysates (*105*). These findings were extended and confirmed by the analysis of artificial bicistronic mRNAs that contained, in order from 5′ to 3′, the 5′-noncoding region of poliovirus connected to the coding region of the *sea* oncogene as the first indicator gene followed by truncated versions of the 5′-noncoding region of EMC virus connected to the poliovirus 2A coding region as the second indicator gene (*102*). The translation *in vitro* of run-off T7 polymerase transcripts of the chimeric polio(*sea*)–EMC (2A) constructs revealed that a specific, internal ribosome entry site probably exists within the 5′-noncoding region of EMC virus RNA. The translational efficiency of the second 2A cistron was not reduced in the presence of a poorly translated first *sea* cistron as long as the second reporter cistron remained under the control of the 5′-noncoding region of EMC virus RNA. Deletion analysis revealed that the EMC RNA 5′-noncoding sequence between nucleotides 260 and 484 plays a critical role in the efficient translation of adjacent coding sequences, in both mono- and bicistronic mRNAs, presumably because the internal ribosomal entry site resides within this region of the EMC RNA (*102*).

It appears that internal initiation sites are also utilized during the translation *in vitro* of vesicular stomatitis virus (VSV) NS mRNA encoding the phosphoprotein NS(P) (*55*), adenovirus mRNA encoding the viral DNA polymerase (*56*), and hepatitis B virus (HBV) pregenomic mRNA encoding the viral reverse transcriptase (*141*). In the case of HBV, combined genetic and biochemical studies both in cell culture and in ducks reveal that the reverse transcriptase is synthesized by a mechanism involving translation initiation at an internal *pol* AUG codon rather than by ribosomal frameshifting within the core-*pol* overlap (*141*). In these cases, the internal initiation occurs at an AUG positioned several hundred nucleotides downstream from, but in-frame with, the 5′-proximal AUG. Synthesis of the VSV 7-kDa (*55*) and the adenovirus 62-kDa (*56*) protein products initiated from the internal AUG is unaffected by hybrid-arrest translation conditions using cDNA fragments complementary to 5′ region viral sequences that inhibit the synthesis of the 5′-proximal AUG-initiated protein

products (NS protein for the VSV mRNA, and the 120-kDa DNA *pol* protein for the adenovirus mRNA). The hybrid-arrest translation results suggest that leaky scanning of ribosomes from the 5′ end of these viral mRNAs to the respective internal initiation sites does not occur to an appreciable extent. However, additional studies are necessary further to support this interpretation. The formation of RNA · DNA duplex structures may inhibit translation by two mechanisms: the duplex may exert a direct steric effect that affects the binding of factors or the movement of ribosomes; alternatively, the mRNA may be cleaved at the site of the duplex by RNase H present in the reticulocyte lysate (*116*).

A highly significant sequence similarity extends through the 5′-noncoding region of the three poliovirus serotypes, PV1, PV2, and PV3 (*108, 117*). Furthermore, a comparative sequence analysis of the 5′-noncoding region of several picornaviruses, including coxsackie B3, human rhinoviruses HRV2 and HRV14, and polioviruses PV1, PV2, PV2S, PV3, and PV3S, revealed the conservation of secondary structure predicted to encompass the entire 5′-noncoding regions of the picornaviruses (*117*). The fact that divergence of picornavirus 5′-noncoding sequences occurs in a manner that permits conservation of certain overall structural features—including over 20 stem and loop structures, two pyrimidine-rich regions and long stretches of conserved sequence—suggests important functional roles for the 5′-noncoding region, possibly mediated as much by overall structure as by specific primary sequence. Undoubtedly, the conserved 5′-noncoding structure of picornavirus mRNAs plays important roles in several stages of the virus multiplication cycle, perhaps including protein synthesis. The conserved 5′-noncoding structure may define a region recognized by the protein synthesis initiation factors and ribosome subunits that results in the internal, cap-independent initiation of translation.

Biochemical evidence suggests that protein-synthesis initiation factors eIF-4A and eIF-4B may play important roles in cap-independent, internal initiation of eukaryotic mRNA translation (*118*). These two factors, together with ATP, are normally specifically required for the binding of 5′-capped mRNA to the 43-S ribosome complex (*119–121*). The eIF-4A is an ATP-dependent single-stranded RNA-binding protein that displays mRNA-dependent ATP-ase activity (*119*); eIF-4F is the three-subunit cap-binding protein (CBP) complex that includes eIF-4A, the CBP eIF-4E, and p220 (*120*); and, eIF-4B is a factor whose exact function remains unknown, but which appears to stimulate the activities of eIF-4A and eIF-4F and to

function in the binding of mRNA to the 43-S complex. Both eIF-4A and eIF-4F can function as RNA-unwinding proteins (*121*); however, eIF-4F; is not required for the cap-independent initiation of translation observed in poliovirus-infected cells (*110*).

In the absence of eIF-4F, both eIF-4A and eIF-4B can bind to an uncapped synthetic mRNA lacking secondary structure with essentially the same degree of effectiveness and affinity observed for capped natural mRNA in the presence of all three factors (*118*). Perhaps the conserved structural elements within the 5′-noncoding sequence of picornaviruses include a feature that allows eIF-4A and eIF-4B to bind to uncapped picornavirus mRNA in the absence of eIF-4F, thereby permitting internal, cap-independent translation-initiation. It should be noted that the efficiency of internal initiation *in vitro* can vary with the nature of the cell-free protein-synthesizing system, as revealed from studies with rabbit reticulocyte and HeLa cell-free systems (*102, 104*). Possibly different kinds of cell-free extracts contain varying concentrations of a *trans*-acting protein-synthesis factor(s) which play(s) an important role in affecting the relative efficiency of ribosomal entry at or near the 5′ end of an mRNA as compared to entry at internal sites within the mRNA. Undoubtedly, the efficiency of internal initiation of translation will be modulated by multiple parameters, including both the degree of optimization of *cis*-acting mRNA sequences and/or structures and the relative concentration and form of the *trans*-acting components of the protein-synthesis machinery.

C. Non-AUG Initiation of Translation

It is apparent that codons other than AUG may initiate the synthesis of proteins in animal cells, albeit so far rarely and generally at a reduced efficiency relative to AUG. Utilization of ACG as an initiator codon has been described for polycistronic mRNAs of two animal viruses, adeno-associated virus (AAV) (*24*) and Sendai virus (*21, 22*). In addition, utilization of CUG as an initiation codon has been described for c-*myc* mRNA (*26*). The adeno-associated virus, Sendai virus, and c-*myc* mRNAs that utilize a non-AUG initiation codon display certain similarities: two or more independent translation-initiation sites are utilized on the same mRNA; the non-AUG initiation codon is located upstream from the AUG initiation codon(s); and, the non-AUG initiation codon is generally utilized less efficiently than the downstream AUG initiation codon (*22, 24–26*).

Adeno-associated virus is a defective parvovirus that replicates in the nucleus of human cells in culture coinfected with adenovirus. The

genome of AAV is single-stranded DNA, either plus or minus (*6*). The AAV capsid protein synthesis was the first of non-AUG initiation described for animal cells. From the sequence of the AAV2 genome (*57*), the structure of the mRNAs that encode the capsid proteins (*58, 59*), and the amino-terminal sequence of the AAV capsid protein B (*24*), it has been concluded that the synthesis of the AAV capsid protein B is initiated at an ACG codon (*24, 25*).

The ACG codon responsible for the initiation of translation of the AAV 70-kDa capsid protein B occurs upstream from, and in the same reading frame as, the AUG codon utilized for initiation of translation of the 60-kDa capsid protein C (*24*). The capsid protein C is synthesized in amounts about 10 to 20 times greater than the capsid protein B (*24, 25, 80*). Translation *in vitro* of a synthetic SP6 AAV transcript has definitively established that the AAV capsid proteins B and C are indeed synthesized from a single mRNA species by alternative use of their respective in-frame initiation codons; the main source of the B and C proteins *in vivo* is probably the known spliced 2.3-kb RNA (25). The coordinated synthesis of B and C from the same mRNA may be due to leaky scanning through the non-AUG initiation codon for B. The less efficient ACG initiation of B-protein synthesis as compared to the more efficient AUG initiation of C-protein synthesis would provide a mechanism for the regulation of the amount of B- and C-protein synthesis in a fixed ratio independent of mRNA concentration (*24, 25*).

Interestingly, in the case of the cellular dihydrofolate reductase (DHFR) mRNA, substitution of ACG for the normal AUG translation codon leads to the synthesis of a normal DHFR protein both *in vivo* and *in vitro* (*122*). In addition, a truncated form of DHFR is also produced, apparently by initiation at the next in-frame AUG located downstream from the ACG. Initiation of DHFR mRNA translation at the ACG codon depends upon a favorable −3/+4 sequence context (*122*).

Sendai virus possesses a nonsegmented, negative-strand RNA genome and replicates in the cytoplasm of infected cells (*6*). The Sendai virus P/C mRNA is polycistronic. Deletion and site-directed point mutants of the P/C mRNA indicate that it codes for five proteins in two overlapping reading frames, utilizing both ACG and AUG as the initiation codons (*21–23, 29–33*). The C reading frame is responsible for the synthesis of four proteins, C′, C, Y1, and Y2, each of which is initiated at an independent site, but all of which appear to terminate at the same site. The P reading frame is responsible for the synthesis of a single product, the P protein (*21, 22*). The P protein is a

phosphoprotein that appears to be associated with the virion-associated RNA-dependent RNA polymerase activity (57); the C, C′, Y1, and Y2 proteins all appear to be nonstructural proteins and their functions are not well established (*21–23, 32*).

The 5′-proximal initiation codon utilized in the Sendai virus P/C mRNA is an ACG in the C reading frame (*21–23*). The ACG codon at nucleotide position 81 is used to initiate the synthesis of the C′ protein. The succeeding initiation codons of the P/C mRNA are all AUG codons. The first AUG downstream from the ACG is in a different frame, the P reading-frame, and is used to initiate the synthesis of P protein at nucleotide position 104. The further downstream initiation AUG codons are all in the C reading-frame and are used to initiate the synthesis of C protein at nucleotide 114, Y1 protein at 181, and Y2 protein at 175 (*21, 22*).

The ACG codon responsible for the initiation of the Sendai virus C′ protein is in a context similar to that of the ACG codon responsible for the initiation of the AAV capsid protein B (*21, 22, 24*). Both ACG codons are in a favored context for efficient ribosome initiation with a purine at position −3 and a guanine at position +4; in addition, positions +5 to +10 are also identical with the exception of +7 which is a uracil in the Sendai mRNA and a cytosine in the AAV mRNA (*21, 22, 24*).

The efficiency of synthesis *in vivo* of the C′ protein from the 5′-proximal ACG differs less than 2-fold from that of the P protein expressed from the downstream AUG, whereas the amount of C protein synthesized is four to five times that of P protein (*21, 32*). These results suggest that an ACG in an otherwise favorable −3/+4 context can function almost as efficiently for initiation of translation as an AUG codon in a less favorable −3/+4 context, by only about 10 to 20% as efficiently as an AUG in a more favored context. The ACG initiator codon of the P/C mRNA appears to be even more efficiently utilized *in vitro* than it is *in vivo* (*32, 60*). Proteins Y1 and Y2 are both expressed *in vivo,* although less efficiently than P protein (N. Gupta, personal communication). Neither the Y1 nor the Y2 AUG initiation codon is in a favorable context for efficient initiation, as pyrimidines are present in the −3 and +4 positions of both AUGs (*21, 22*).

A non-AUG translational initiation also is utilized by the cellular mRNA that encodes c-myc proteins. The c-*myc* gene comprises three exons with a single large AUG-initiated open reading frame extending from exon 2 through exon 3; exon 1 lacks any AUG codons. Two major forms of C-myc proteins have been identified that, depending upon the species of cell analyzed, differ by 2 to 4 kDa in apparent mass (*123,*

124). These two proteins are derived from alternative translational initiations in the same reading frame in exons 1 and 2. Site-specific mutagenesis results show that initiation of translation of capped SP6 or T7 transcripts *in vitro* occurs at an AUG codon in exon 2 and at a CUG codon near the 3′ end of exon 1. The initiation of translation from the exon-1 CUG and from the downstream exon-2 AUG results in the production of c-myc proteins with distinct amino-termini (*26*). Analysis of c-myc proteins synthesized in Burkitt's lymphoma cell lines containing different chromosomal translocations, and in Epstein–Barr virus (EBV)-immortalized lymphoblastoid cell lines that do not have a rearranged c-*myc* locus, suggests that alternative CUG and AUG translation-initiation sites are used *in vivo* as well as *in vitro* (*26*).

D. Ribosomal Frameshifting during Elongation

Ribosomal frameshifting during the elongation stage of mRNA translation provides a mechanism for the synthesis of a single protein from two different reading frames on an RNA template. The coupling of the reading frames requires that the ribosome shifts correctly from one reading frame to the other at a discrete position on the mRNA so as to avoid termination of polypeptide synthesis (*125*). Frameshifting permits the production of two unique polypeptide products from a single mRNA by initiation of translation at a single site, but subsequent termination of translation at different sites in different reading frames. Termination of protein synthesis at different sites is the result of translational frameshifting that causes a fraction of the ribosomes to change reading frame at a discrete position on the mRNA. Several retroviruses utilize such a frameshifting strategy (*61–68*).

Most, if not all, retroviruses express the *gag*-encoded viral core proteins in the form of a *gag*–polyprotein precursor that is processed by a virus-encoded protease; the *pol*-encoded reverse transcriptase and integrase proteins are normally expressed at much lower levels by a similar proteolytic processing of a large, fused *gag–pol* precursor polypeptide (*103, 126*). However, the genetic structure of the *gag-pol* domains of the Rous sarcoma virus (RSV) genome (*61*), the human T-cell leukemia virus (HTLV-I and -II) genomes (*62*), the bovine leukemia virus genome (*63*), the human immunodeficiency virus (HIV-1) genome (*64*), and the mouse mammary tumor virus (MMTV) genome (*65*) seemingly would preclude the synthesis of a *gag–pol* fusion protein, because the *gag* and *pol* genes are in different reading frames. The *pol* gene of RSV and HIV is in the −1 reading frame with

respect to the *gag* gene; however, the products of *pol* do not arise from independent translation-initiation events or from the translation of a spliced mRNA, but rather because of a ribosomal frameshifting event during *gag–pol* expression (*61*, *64*). In MMTV, two −1 frameshift events are required for the synthesis of the *gag–pol* fusion protein, one at the *gag–x/pro* overlap and the other at the *x/pro–pol* overlap (*65*, *66*). The frameshifting model is supported by synthesis of both the *gag* protein and the *gag–pol* fusion protein in an animal cell-free protein-synthesizing system *in vitro* using a single RSV, HIV, or MMTV mRNA template synthesized *in vitro* from cloned viral cDNA by SP6 polymerase.

Site-directed mutagenesis and amino-acid sequencing located the site of HIV *gag–pol* frameshifting at a leucine UUA codon within a U-UUA sequence near the 5′ end of the *gag–pol* overlap region (*64*). The same sequence also appears in the *gag–pol* overlap of RSV (*61*) and the *pro–pol* overlap of MMTV (*65*, *66*). The exact molecular mechanism of translational frameshifting in animal cell systems is not yet known, but in RSV, HIV and MMTV, the frameshifting events may involve a slippage of the leucyl-tRNA reading the UUA codon back to the −1-frame UUU codon (*64*, *137*). However, a larger sequence context, possibly including secondary structure within the region flanking the site of the frameshift event, also appears necessary for frameshifting, because synthetic oligonucleotides containing either of the MMTV overlap regions inserted into novel contexts do not induce frameshifting (*66*). Indeed, the effects of deletion and site-directed mutations best correlate with the potential to form an RNA stem–loop structure adjacent to the frame-shift site. A 147-nucleotide sequence from RSV RNA containing the frameshift site and stem–loop structure is sufficient to direct frameshifting in a novel genetic context (*137*). The efficiency of ribosomal frameshifting during translation of the mRNA encoding *gag* and the *gag–pol* fusion is about 5% within the RSV *gag–pol* overlap (*61*) and about 10% within the HIV *gag–pol* overlap (*64*). By contrast, in MMTV, about 25% of the ribosomes traversing the *gag–x/pro* overlap and 10% traversing the *x-pro/pol* overlap frameshift in the −1 direction (*66*).

E. Suppression of Termination

The genetic structure of the *gag–pol* domains of the murine leukemia virus (MuLV) (*67*) and the feline leukemia virus (FeLV) (*68*) genomes precludes the synthesis of the *gag–pol* fusion protein because *gag* and *pol* are separated by an in-frame nonsense codon. To

circumvent this apparent block to synthesis of the *gag–pol* leukemia virus fusion proteins, MuLV and FeLV utilize a strategy of termination codon suppression in which ribosomes "read-through" the in-frame stop codon. The MuLV protease is located at the 5′ end of the *pol* gene and is synthesized within the precursor gag–pol fusion protein by suppression of an amber (UAG) termination codon located at the 3′ end of the *gag* gene (*67*). The first four amino-terminal amino acids of the protease are derived from the *gag* gene; the fifth residue, glutamine, is derived through suppression of the in-frame amber termination codon located at the *gag–pol* junction (*67*). Similar to MuLV (*67*), the FeLV protease is likewise synthesized through in-frame suppression of the *gag* amber termination codon by insertion of a glutamine residue at the UAG codon (*68*). Based on precedents from prokaryotic systems, natural suppression of the nonsense termination codon present in the MuLV and FeLV mRNAs presumably occurs by a mechanism involving suppressor tRNAs that possess altered anticodons (*127*). Normal mouse cells contain a natural UAG suppressor glutamine tRNA, representing about 1–2% of the total glutamine tRNA (*133, 134*). The supressor glutamine tRNA may be increased in some cells infected with MuLV (*133*) and selectively decreased in cells treated with Avarol,[1] a sesquiterpenoid hydroquinone displaying anti-MuLV activity (*134*). Indeed, nonsense suppressor tRNAs active in mammalian cells have been constructed using site-specific mutagenesis of cloned tRNA genes. The biological activity of these recombinant suppressor tRNAs was directly demonstrated by suppression of termination of nonsense codons introduced into cDNA clones encoding animal virus mRNAs (*128–130*). However, in principle, either the primary or a higher ordered structure of the MuLV or FeLV mRNA templates could also affect the efficiency of suppression of termination by affecting the competition between the protein-synthesis release factors promoting termination and the tRNA species mediating suppression by promoting read-through. Analysis of recombinant fusion genes consisting of the viral *gag–pol* junction upstream and in the same reading frame as the *E. coli lac*Z gene revealed that *gag* amber codon suppression in these constructs does not require augmented levels of suppressor tRNA species (*98*). Rather, suppression is caused by an intrinsic *cis*-acting component of the viral mRNA (*98*).

[1] Avarol is a highly reduced, highly methylated nephthalenylmethyl-1,4-benzenediol (Chem. Abstr. No. 55303-98-5) [Eds.].

F. Reinitiation following Termination

In polycistronic mRNAs with open reading frames or cistrons that do not overlap, reinitiation without prior dissociation of the 40-S subunit from the mRNA following termination of translation of the upstream cistron would provide a possible mechanism for the subsequent initiation of translation of the downstream cistron. Such reinitiation following termination by animal cell ribosomes has been shown to occur with several viral and cellular mRNAs (*69–77, 79*). Furthermore, termination-reinitiation of translation has been implicated as a possible negative regulatory mechanism in the expression of certain viral mRNAs [for example, the B19 parvovirus VP1 RNA (*70*), a cytomegalovirus (CMV) β mRNA (*79*), SV40 early mRNA (*69*), and the RSV *src* mRNA (*72*)].

The B19 parvovirus produces two capsid proteins in strikingly different quantities—VP1 (<4%) and VP2 (>96%)—from overlapping RNAs that are derived from the same transcriptional unit (*71*). Translation of the VP1 RNA is very inefficient compared to VP2 RNA in the reticulocyte cell-free system. The region about 250 nucleotides upstream from the VP1 initiation AUG codon contains seven AUG codons followed by in-frame termination codons that create multiple minicistrons; this region is removed by splicing and is not present in the efficiently translated VP2 RNA. The multiple upstream AUG codons of the VP1 RNA negatively regulate VP1 synthesis (*70*). Removal of the upstream AUG codons from VP1 RNA greatly increases the efficiency of VP1 RNA translation. Conversely, the addition of the same VP1 RNA upstream sequence containing multiple AUG codons to a position upstream from the initiation codon of VP2 markedly decreases VP2 translation to a level of about 5% of the original level (*70*). Although the upstream AUG-rich region of the VP1 RNA behaves as a negative regulatory element in translational control, no significant difference is observed when the termination site of the last minicistron precedes rather than overlaps the authentic VP1 AUG initiation codon (*70*). Because some of the upstream AUG codons in the B19 VP1 RNA are present in an appropriate $-3/+4$ context for translation-initiation (*71*), it was proposed that reinitiation rather than leaky scanning is mainly responsible for the reduced efficiency of VP1 translation (*70*).

SV40 (*69*) and CMV (*79*), similar to B19 papovavirus (*70*), also possess upstream "leader" minicistrons that display a negative regulatory activity for mRNA translation. The SV40 early region encodes two proteins, "large T" and "small t," from mRNAs that share

common 5′-noncoding regions (*132*). During the course of SV40 infection of permissive cells, a change occurs in the position of the start sites for the synthesis of SV40 early RNA. Prior to DNA replication, the start site of early–early RNA is at map position 5235–5237, whereas following DNA replication two start sites are observed, one for late–early RNA located at m.p. 22–28 and the other for far upstream late–early RNA at m.p. 35–43. During the switch from early–early to late–early transcription, the shift in the start site of the SV40 early RNAs results in the introduction of AUG initiation codons in the 5′-noncoding region of the late–early mRNA. The introduction of the upstream AUG codons contributes to a reduced translational efficiency of the mRNAs encoding the tumor antigens. When SP6 transcripts are prepared and translated in a wheat system, the early–early RNA is a 3- to 6-fold more efficient mRNA for synthesis of t antigen, and presumably also T antigen, than is the late–early mRNA; furthermore, the early–early RNA is about a 10-fold better mRNA than the far upstream late–early mRNA. If AUG initiation codons followed by in-frame termination codons occur in the 5′-noncoding region upstream from the authentic initiation codon, minicistrons are created that, in principle, may be translated. Such a situation appears to occur in the case of the SV40 late–early RNA. A 23-residue polypeptide product of the SV40 late–early minicistron has been identified in infected cells; its function is unknown (*69*). In the case of the human CMV β gene transcript, two short minicistrons of 7 and 35 codons in length occur within the 5′-noncoding leader region of a 170-codon cistron. Transcripts carrying the 5′ leader are inefficiently translated as a consequence of the two upstream minicistrons, and the presence of either of the two minicistrons reduces expression of a downstream cistron during viral infection (*79*).

The three RSV mRNAs encoding *gag* and *gag–pol*, *env*, and *src* contain a common 5′-noncoding leather segment (*72*, *126*). With the spliced RSV *src* mRNA, the AUG codon used to initiate *src* protein synthesis is positioned 90 nucleotides downstream from the AUG codon present in the common 5′-leader region used to initiate *gag* and *gag–pol* protein synthesis. In addition, an in-frame UGA termination codon lies between the *gag* and *gag–pol* AUG and the *src* AUG in the spliced *src* mRNA. Mutation of the upstream minicistron termination UGA codon to CGA virtually eliminates initiation of translation at the downstream "authentic" initiator AUG used for synthesis of the 60-kDa *src* oncoprotein (*72*). Perhaps the absence of 60-kDa *src* synthesis is caused by eliminating the possibility of reinitiation because of the absence of minicistron termination. Conceptually

similar results were obtained when the expression of the hepatitis-B surface antigen (HB_sAg) gene was studied in transfected COS cells (*76*). The insertion of AUG initiation codons upstream from the "authentic" AUG used to initiate translation of the downstream HB_sAg gene could severely depress the initiation of translation at the "authentic" AUG codon. Such inhibition could, however, be at least partially suppressed by the presence of a translational termination codon in-frame with the upstream AUG (*76*).

These results are consistent with the possibility that animal cell ribosomes can reinitiate translation at an AUG codon after previously initiating, and subsequently terminating, at an upstream site. The fact that the presence of an upstream termination codon can be shown to modulate directly the efficiency of initiation of translation at the "authentic" downstream internal AUG codons of *src* mRNA (*72*) and HB_2Ag mRNA (*76*) strongly suggests that the internal initiation is by a reinitiation mechanism and is not mediated by ribosomes that escaped initiation at the upstream site.

The translational analysis in stably transfected Chinese hamster ovary (CHO) cells and transiently transfected COS cells of polycistronic mRNAs constructed to contain two or more cellular genes in non-overlapping open reading frames likewise demonstrate that an upstream open reading frame usually (*74*, *75*, *77*), but not always (*75*, *77*), markedly reduces the efficiency of translation of a downstream cistron. Furthermore, the relative position of the terminator codon of the upstream cistron's open reading frame appears to be a critical parameter in determining expression from the downstream, internal cistron. Some evidence suggests that efficient reinitiation of translation can occur at an internal downstream AUG if the translation that initiated from the upstream AUG codon terminates about 80 to 150 nucleotides before the downstream AUG (*77*). However, other evidence suggests that reinitiation can occur at an internal AUG if the translation that initiated from the upstream AUG codon terminates within about 20 to 50 nucleotides either before or after the downstream AUG (*74*, *75*). Although the maximum intercistronic distance over which an animal cell ribosome may reinitiate has not been precisely resolved (*75*, *77*), these latter results (*74*, *75*) imply that ribosomes may also be able to "reach back" or scan bidirectionally.

In summary, the results of many studies designed to assess the effects of upstream initiation and termination codons on downstream translation-initiation are consistent collectively with the notion that internal initiation of translation may occur by a mechanism involving translation termination-reinitiation. Translation termi-

nation-reinitiation may occur when the termination codon of the upstream cistron either precedes or is in the close vicinity of the initiation site of the downstream second cistron. The evidence supporting reinitiation following termination is derived from a number of different translational systems, both *in vitro* and *in vivo*, and for both viral and cellular mRNAs (69–77, 79).

III. Conclusions

Among the best-characterized examples of naturally occurring polycistronic mRNAs are those specified by animal viruses. Careful genetic and biochemical analyses have firmly established that the protein-synthesizing machinery of animal cells can indeed recognize and translate polycistronic viral mRNAs. The translation of polycistronic mRNAs involves multiple initiation and/or termination codons in a manner that permits the synthesis of multiple primary translational products from a single, mature mRNA. Several different biochemical mechanisms appear to allow the translation of polycistronic mRNAs by the protein-synthesizing machinery of animal cells. Indeed, different mechanisms can best account for the translational expression of different polycistronic mRNAs. For example, for some polycistronic viral mRNAs, translation-initiation appears to occur by the binding of the 40-S ribosomal subunit at or near the mRNA 5′ terminus and a subsequent "leaky" scanning in which alternative initiation codons are recognized as functional initiation sites. For other mRNAs, translation-initiation appears to occur by the binding of ribosomal subunits to internal sequences, or by reinitiation following termination without prior dissociation of the ribosomal subunits from the mRNA. Certain polycistronic mRNAs utilize a non-AUG codon, for example, ACG, to initiate translation of one of the cistrons. Finally, the translation of polycistronic mRNAs, in some cases, is by mechanisms in which alternative termination codons are used to terminate translation, either as the result of ribosomal frameshifting during elongation, or by suppression of an in-frame termination signal.

Many of the molecular genetic mechanisms of eukaryotic gene expression first elucidated from studies of animal viruses have subsequently been shown to be used for the expression of cellular genes. Hopefully, future studies will establish which of the multiple mechanisms so far described for the translation of polycistronic animal virus mRNAs are used likewise for the translation of polycistronic cellular mRNAs. Future studies may also better identify and characterize the parameters important in the regulated translation of polycistronic

mRNAs, parameters including *cis*-acting primary sequences or higher ordered structures of the mRNA, and *trans*-acting protein factors of the protein-synthesis machinery.

ACKNOWLEDGMENTS

I would like to thank the many authors who generously provided preprints and reprints of their work. The work from the author's laboratory was supported in part by research grants from the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

REFERENCES

1. J. R. Nevins, *ARB* **52,** 441 (1983).
2. R. A. Padgett, P. J. Grabowski, M. M. Konarska, S. Seiler and P. A. Sharp, *ARB* **55,** 1119 (1986).
3. K. Moldave, *ARB* **54,** 1109 (1985).
4. M. Kozak, *Microbiol. Rev.* **47,** 1 (1983).
5. M. Kozak, *NARes* **12,** 857 (1984).
6. B. N. Fields, D. M. Knipe, R. M. Chanock, J. L. Melnick, B. Roizman and R. E. Shope (eds.), "Virology." Raven, New York, 1985.
7. M. Kozak, *Adv. Virus Res.* **31,** 229 (1986).
8. R. J. Schneider and T. Shenk, *ARB* **56,** 317 (1987).
9. H.-G. Kräusslich and E. Wimmer, *ARB* **57,** 701 (1988).
10. M. Kozak, *Curr. Top. Microbiol. Immunol.* **93,** 81 (1981).
11. H. A. Lutcke, K. C. Chow, F. S. Mickel, K. A. Moss, H. F. Kern and G. A. Scheele, *EMBO J.* **6,** 43 (1987).
12. M. Kozak, *Cell* **47,** 481 (1986).
13. S. A. Sedman and J. E. Mertz, *J. Virol.* **62,** 954 (1988).
14. C. Dabrowski and J. C. Alwine, *J. Virol.* **62,** 3182 (1988).
15. P. J. Good, R. C. Welch, A. Barkan, M. B. Somasekhar and J. E. Mertz, *J. Virol.* **62,** 944 (1988).
16. C. M. Preston and D. J. McGeoch, *J. Virol.* **38,** 593 (1981).
17. L. Haarr, H. S. Marsden, C. M. Preston, J. R. Smiley, W. C. Summers and W. P. Summers, *J. Virol.* **56,** 512 (1985).
18. J. D. Rosenblatt, A. J. Cann, D. J. Slamon, I. S. Smalberg, N. P. Shah, J. Fujii, W. Wachsman and I. S. Y. Chen, *Science* **240,** 916 (1988).
19. W.-K. Chan, M. E. Penaranda, S. E. Crawford and M. K. Estes, *Virology* **151,** 243 (1986).
20. L. McGinnes, C. McQuain and T. Morrison, *Virology* **164,** 256 (1988).
21. J. Curran and D. Kolakofsky, *EMBO J.* **7,** 245 (1988).
22. K. C. Gupta and S. Patwardhan, *JBC* **263,** 8553 (1988).
23. S. Patwardhan and K. C. Gupta, *JBC* **263,** 4907 (1988).
24. S. P. Becerra, J. A. Rose, M. Hardy, B. M. Baroudy and C. W. Anderson, *PNAS* **82,** 7919 (1985).
25. S. P. Becerra, F. Koczot, P. Fabisch and J. A. Rose, *J. Virol.* **62,** 2745 (1988).
26. S. R. Hann, M. W. King, D. L. Bentley, C. W. Anderson and R. N. Eisenman, *Cell* **52,** 185 (1988).
27. J. L. Bos, L. J. Polder, R. Bernards, P. I. Schrier, P. J. van den Elsen, A. J. van der Eb and H. van Ormondt, *Cell* **27,** 121 (1981).

28. W. S. M. Wold, S. L. Deutscher, N. Takemori, B. M. Bhat and S. C. Magie, *Virology* **148,** 168 (1986).
29. R. A. Lamb, B. W. J. Mahy and P. W. Choppin, *Virology* **69,** 116 (1976).
30. C. Giorgi, B. M. Blumberg and D. Kolakofsky, *Cell* **35,** 829 (1983).
31. L. Dethlefsen and D. Kolakofsky, *J. Virol.* **46,** 321 (1983).
32. J. A. Curran, C. Richardson and D. Kolakofsky, *J. Virol.* **57,** 684 (1986).
33. J. A. Curran and D. Kolakofsky, *J. Gen. Virol.* **68,** 2515 (1987).
34. D. Luk, A. Sańchez and A. Banerjee, *Virology* **153,** 318 (1986).
35. M. S. Galinski, M. A. Mink, D. M. Lambert, S. L. Wechsler and M. W. Pons, *Virology* **155,** 46 (1986).
36. W. J. Bellini, G. Englund, S. Rozenblatt, H. Arnheiter and C. D. Richardson, *J. Virol.* **53,** 908 (1985).
37. F. Fuller, A. S. Bhown and D. H. L. Bishop, *J. Gen. Virol.* **64,** 1705 (1983).
38. M. Bouloy, P. Vialat, M. Girard and N. Pardigon, *J. Virol.* **49,** 717 (1984).
39. S. Gerbaud, P. Vialat, N. Pardigon, C. Wychowski, M. Girard and M. Bouloy, *Virus Res.* **8,** 1 (1987).
40. L. Nagata, S. A. Masri, D. C. W. Mah and P. W. K. Lee, *NARes* **12,** 8699 (1984).
41. L. W. Cashdollar, R. A. Chmelo, J. R. Wiener and W. K. Joklik, *PNAS* **82,** 24 (1985).
42. S. M. Munemitsu, J. A. Atwater and C. E. Samuel, *BBRC* **140,** 508 (1986).
43. B. L. Jacobs and C. E. Samuel, *Virology* **143,** 63 (1985).
44. H. Ernest and A. J. Shatkin, *PNAS* **82,** 48 (1985).
45. G. Sarkar, J. Pelletier, R. Bassel-Duby, A. Jayasuriya, B. N. Fields and N. Sonenberg, *J. Virol.* **54,** 720 (1985).
46. B. L. Jacobs, J. A. Atwater, S. M. Munemitsu and C. E. Samuel, *Virology* **147,** 9 (1985).
47. M. W. Shaw and P. W. Choppin, *Virology* **139,** 178 (1984).
48. M. W. Shaw, P. W. Choppin and R. A. Lamb, *PNAS* **80,** 4879 (1983).
49. J. L. Leibowitz, S. Perlman, G. Weinstock, J. R. DeVries, C. Budzilowicz, J. M. Weissemann and S. R. Weiss, *Virology* **164,** 156 (1988).
50. C. J. Budzilowicz and S. R. Weiss, *Virology* **157,** 509 (1987).
51. M. A. Skinner, D. Ebner and S. G. Siddell, *J. Gen. Virol.* **66,** 581 (1985).
52. S. M. Munemitsu and C. E. Samuel, *Virology* **163,** 643 (1988).
53. R. A. Katz, B. R. Cullen, R. Malavarca and A. M. Skalka, *MCBiol* **6,** 372 (1986).
54. C.-C. Liu, C. C. Simonsen and A. D. Levinson, *Nature* **309,** 82 (1984).
55. R. C. Herman, *J. Virol.* **58,** 797 (1986).
56. D. Hassin, R. Korn and M. S. Horwitz, *Virology* **155,** 214 (1986).
57. A. Srivastava, E. W. Lusby and K. I. Berns, *J. Virol.* **45,** 555 (1983).
58. M. R. Green and R. G. Roeder, *J. Virol.* **36,** 79 (1980).
59. C. A. Laughlin, H. Westphal and B. J. Carter, *PNAS* **76,** 5567 (1979).
60. K. C. Gupta and D. W. Kingsbury, *BBRC* **131,** 91 (1985).
61. T. Jacks and H. Varmus, *Science* **230,** 1237 (1985).
62. K. Shimotohno, Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa and T. Sugimura, *PNAS* **82,** 3101 (1985).
63. Y. Yoshinaka, I. Katoh, T. D. Copeland, G. W. Smythers and S. Oroszlan, *J. Virol.* **57,** 826 (1986).
64. T. Jacks, M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr and H. E. Varmus, *Nature* **331,** 280 (1988).
65. R. Moore, M. Dixon, R. Smith, G. Peters and C. Dickson, *J. Virol.* **61,** 480 (1987).
66. T. Jacks, K. Townsley, H. E. Varmus and J. Majors, *PNAS* **84,** 4298 (1987).
67. Y. Yoshinaka, I. Katoh, T. D. Copeland and S. Oroszlan, *PNAS* **82,** 1618 (1985).
68. Y. Yoshinaka, I. Katoh, T. D. Copeland and S. Oroszlan, *J. Virol.* **55,** 870 (1985).

69. K. Khalili, J. Brady and G. Khoury, *Cell* **48,** 639 (1987).
70. K. Ozawa, J. Ayub and N. Young, *JBC* **262,** 10922 (1988).
71. K. Ozawa J. Ayub, H. Yu-Shu, G. Kurtzman, T. Shimada and N. Young, *J. Virol.* **61,** 2395 (1987).
72. S. Hughes, K. Mellstrom, E. Kosik, F. Tamanoi and J. Brugge, *MCBiol* **4,** 1738 (1984).
73. R. J. Kaufman, P. Murtha and M. V. Davies, *EMBO J.* **6,** 187 (1987).
74. D. S. Peabody and P. Berg, *MCBiol* **6,** 2695 (1986).
75. D. S. Peabody, S. Subramani and P. Berg, *MCBiol* **6,** 2704 (1986).
76. C.-C. Liu, C. C. Simonsen and A. D. Levinson, *Nature* **309,** 82 (1984).
77. M. Kozak, *MCBiol* **7,** 3438 (1987).
78. G. Jay, S. Nomura, C. W. Anderson and G. Khoury, *Nature* **291,** 346 (1981).
79. A. P. Geballe and E. S. Mocarski, *J. Virol.* **62,** 3334 (1988).
80. R. M. L. Buller and J. A. Rose, *J. Virol.* **25,** 331 (1978).
81. M. Kozak, *NARes* **9,** 5233 (1981).
82. M. Kozak, *Nature* **308,** 241 (1984).
83. M. Kozak, *Cell* **44,** 283 (1986).
84. A. J. Shatkin, *Cell* **9,** 645 (1976).
85. K. H. Levin and C. E. Samuel, *Virology* **77,** 245 (1977).
86. T. G. Lawson, M. H. Chadaras, B. K. Ray, K. A. Lee, R. D. Abramson, W. C. Merrick and R. E. Thach, *JBC* **263,** 7266 (1988).
87. M. Kozak, *Nature* **280,** 82 (1979).
88. M. Konarska, W. Filipowicz, H. Domdey and H. J. Gross, *EJB* **114,** 221 (1981).
89. M. Kozak and A. J. Shatkin, *JBC* **253,** 6568 (1978).
90. M. Kozak, *Cell* **22,** 459 (1980).
91. M. Kozak, *Cell* **19,** 79 (1980).
92. J. Pelletier and N. Sonenberg, *Cell* **40,** 515 (1985).
93. M. Kozak, *PNAS* **83,** 2850 (1986).
94. D. Chevrier, C. Vézina, J. Bastille, C. Linard, N. Sonenberg and G. Boileau, *JBC* **263,** 902 (1988).
95. L. P. van Duijn, S. Holsappel, M. Kasperaitis, H. Bunschoten, D. Konings and H. O. Voorma, *EJB* **172,** 59 (1988).
96. M. Kozak, *NARes* **12,** 3873 (1984).
97. C.-C. Liu, C. C. Simonsen and A. D. Levinson, *Nature* **309,** 82 (1984).
98. A. T. Panganiban, *J. Virol.* **62,** 3574 (1988).
99. R. J. Kaufman, P. Murtha and M. V. Davies, *EMBO J.* **6,** 187 (1987).
100. G. Sarkar, I. Edery, R. Gallo and N. Sonenberg, *BBA* **783,** 122 (1984).
101. J. Pelletier and N. Sonenberg, *Nature* **334,** 320 (1988).
102. S. K. Jang, H.-G. Kräusslich, M. J. H. Nicklin, G. M. Duke, A. C. Palmenberg and E. Wimmer, *J. Virol.* **62,** 2636 (1988).
103. M. R. Roner, R. K. Gaillard and W. K. Joklik, *Virology* **168,** 292 (1989).
104. A. J. Dorner, B. L. Semler, R. J. Jackson, R. Hanecak, E. Duprey and E. Wimmer, *J. Virol.* **50,** 507 (1984).
105. D. S. Shih, I.-W. Park, C. L. Evans, J. M. Jaynes and A. C. Palmenberg, *J. Virol.* **61,** 2033 (1987).
106. N. Kitamura, B. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. Aemini, R. Hanecak, J. J. Lee, S. Werf, C. W. Anderson and E. Wimmer, *Nature* **291,** 547 (1981).
107. V. R. Racaniello and D. Baltimore, *PNAS* **78,** 4887 (1981).
108. H. Toyoda, M. Kohara, Y. Kataoka, T. Suganuma, T. Omata, N. Imura and A. Nomoto, *JMB* **174,** 561 (1984).

109. A. J. Dorner, L. F. Dorner, G. R. Larsen, E. Wimmer and C. W. Anderson, *J. Virol.* **42,** 1017 (1982).
110. N. Sonenberg, *Adv. Virus Res.* **33,** 175 (1987).
111. D. Etchison, S. Milburn, I. Edery, N. Sonenberg and J. W. B. Hershey, *JBC* **257,** 14806 (1982).
112. J. Pelletier, G. Kaplan, V. R. Racaniello and N. Sonenberg, *MCBiol* **8,** 1103 (1988).
113. D. Trono, J. Pelletier, N. Sonenberg and D. Baltimore, *Science* **241,** 445 (1988).
114. D. Trono, R. Andino and D. Baltimore, *J. Virol.* **62,** 2291 (1988).
115. K. Bienkowska-Szewczyk and E. Ehrenfeld, *J. Virol.* **62,** 3068 (1988).
116. R. Y. Walder and J. A. Walder, *PNAS* **85,** 5011 (1988).
117. V. M. Rivera, J. D. Welsh and J. V. Maizel, Jr., *Virology* **165,** 42 (1988).
118. R. D. Abramson, T. E. Dever and W. C. Merrick, *JBC* **263,** 6016 (1988).
119. R. D. Abramson, T. E. Dever, T. G. Lawson, B. K. Ray, R. E. Thach and W. C. Merrick, *JBC* **262,** 3826 (1987).
120. K. Moldave, *ARB* **54,** 1109 (1985).
121. B. K. Ray, T. G. Lawson, J. C. Kramer, M. H. Cladaras, J. A. Grifo, R. D. Abramson, W. C. Merrick and R. E. Thach, *JBC* **260,** 7651 (1985).
122. D. S. Peabody, *JBC* **262,** 11847 (1987).
123. G. Ramsay, G. I. Evan and J. M. Bishop, *PNAS* **81,** 7742 (1984).
124. S. R. Hann and R. N. Eisenman, *MCBiol* **4,** 2486 (1984).
125. W. J. Craigen and C. T. Caskey, *Cell* **50,** 1 (1987).
126. R. Weiss, N. Teich, H. Varmus and J. Coffin, "Molecular Biology of Tumor Viruses: TNA Tumor Viruses." CSH Laboratory, Cold Spring Harbor, New York, 1985.
127. J. E. Celis and J. D. Smith, "Nonsense Mutations and tRNA Suppressors." Academic Press, London, 1979.
128. F. A. Laski, R. Belagaje, U. L. RajBhandary and P. A. Sharp, *PNAS* **79,** 5813 (1982).
129. J. F. Young, M. R. Capecchi, F. A. Laski, U. L. RajBhandary, P. A. Sharp and P. Palese, *Science* **221,** 873 (1983).
130. J. M. Sedivy, J. P. Capone, U. L. RajBhandary and P. A. Sharp, *Cell* **50,** 379 (1987).
131. M. L. Celma and E. Ehrenfeld, *JMB* **98,** 761 (1975).
132. J. Tooze, "Molecular Biology of Tumor Viruses: DNA Tumor Viruses." CSH Laboratory Cold Spring Harbor, New York, 1980.
133. Y. Kuchino, H. Beier, N. Akita and S. Nishimura, *PNAS* **84,** 2669 (1987).
134. Y. Kuchino, S. Nishimura, H. C. Schröder, M. Rottmann and W. E. G. Müller, *Virology* **165,** 518 (1988).
135. A. Barkan and J. E. Mertz, *MCBiol* **4,** 813 (1984).
136. J. Pelletier, M. E. Flynn, G. Kaplan, V. Racaniello and N. Sonenberg, *J. Virol.* **62,** 4486 (1988).
137. T. Jacks, H. D. Madhani, F. R. Masiarz and H. E. Varmus, *Cell* **55,** 447 (1988).
138. M. A. Williams and R. A. Lamb, *J. Virol.* **63,** 28 (1989).
139. J. Pelletier and N. Sonenberg, *J. Virol.* **63,** 441 (1989).
140. P. J. Dillon and K. C. Gupta, *J. Virol.* **63,** 974 (1989).
141. L-J. Chang, P. Pryciak, D. Ganem and H. E. Varmus, *Nature* **337,** 364 (1989).