Volume 9 Number 20 1981 Nucleic Acids Research

Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes

Marilyn Kozak

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received 3 August 1981

ABSTRACT

Sequences flanking the initiator codon in eukaryotic mRNAs are not random. Out of 153 messages examined, 151 have either a purine in position -3, or a G in position +4, or both. Thus, $\begin{matrix}A\\G\end{matrix}$XXAUGG emerges as the favored sequence for eukaryotic initiation sites. Nucleotides flanking nonfunctional AUG triplets, which occur in the 5'-noncoding region of a few eukaryotic messages, are different from those found at most functional sites. Whereas most authentic initiator codons are preceded by a purine (usually A) in position -3, most nonfunctional AUGs have a pyrimidine in that position. The observed asymmetry suggests that purines in positions -3 and +4 might facilitate recognition of the AUG codon during formation of initiation complexes. To test this idea, in vitro binding studies were carried out with ^{32}P-labeled oligonucleotides. Binding of AUG-containing oligonucleotides to wheat germ ribosomes was significantly enhanced by placing a purine in position -3 or +4. The scanning model, which postulates that 40S ribosomal subunits attach at the 5'-end of a message and migrate down to the AUG codon, is discussed in light of these new observations. A modified version of the scanning mechanism is proposed.

INTRODUCTION

The pivotal role of the AUG codon in defining the start site for protein biosynthesis has long been recognized. Only a small fraction of the AUG triplets in a given message function as ribosome binding sites, however. In prokaryotic messenger RNAs, the major ancillary signal that dictates which AUG codons will be selected by ribosomes is a purine-rich sequence centered about ten nucleotides upstream from the AUG triplet (1). Other sequences located farther to the left (2-4) or right (5,6) of the AUG codon have also been shown to influence translational efficiency, at least in some messages. In prokaryotes, the role of the purine-rich sequence preceding the initiator codon was deduced from comparison of nucleotide sequences among a large number of messages, and from manipulation of messenger RNAs; i.e., altering the "Shine/Dalgarno" region preceding a particular initiator codon lowered or abolished binding of ribosomes to that site (7,8). By contrast, eukaryotic ribosomes appear to be more tolerant of sequence changes within the region of a message preceding the ini-

tiator codon (9-12). Comparison of the 5'-proximal nucleotide sequences of eukaryotic messages reveals remarkable heterogeneity, even among some closely related mRNAs (13-15), reinforcing the impression that eukaryotic ribosomes might not require a particular sequence to demarcate the initiation site. Based in part on observations such as these, I have hypothesized that eukaryotic ribosomes bind initially at the 5'-end of a message and then migrate, scanning the mRNA sequence until they encounter the first AUG triplet which, solely by virtue of its position, is the initiator codon. Considerable evidence has been adduced in favor of this "scanning mechanism" (16,17). A compilation in Dec., 1980, revealed that in 90 out of 99 eukaryotic messages which had been sequenced, translation does indeed begin at the AUG triplet which is closest to the 5'-terminus (17). Although it is gratifying that 90% of the messages examined conform to the prediction, the 9 exceptional messages (a number which has since grown to 11) in which translation does *not* start at the first AUG have been puzzling. The scanning model, in its simplest form, states that the initiator codon is recognized by its position (i.e., first-in-line) irrespective of the flanking sequences. But as sequence information has become available from more and more eukaryotic messages, it has become obvious that the nucleotides flanking the initiator codon are not random. The data compiled in this paper show that nucleotides in positions -3 and +4 are highly conserved. (The numbering system used here is $\overset{-3}{X}pXpX\overset{+1}{p}A\overset{+4}{p}UpGpX$.) To determine whether the conserved nucleotides play a role during initiation, I have constructed a series of AUG-containing oligonucleotides and measured their ability to bind to wheat germ ribosomes in vitro. The binding efficiency of the oligonucleotides was significantly enhanced by placing a purine in position -3 or +4. A slightly more elaborate version of the scanning model, which takes these new data into account, may provide an explanation for those exceptional eukaryotic messages in which initiation is not restricted to the 5'-proximal AUG codon.

MATERIALS AND METHODS

Synthesis and characterization of oligonucleotides

To simplify the representation of nucleotide sequences, the designation $\overset{*}{p}$ is used for [^{32}P]; Y = pyrimidine; R = purine; X = any one of the four common ribonucleotides. All of the di- and trinucleotides used in this work were purchased from P-L Biochemicals except for GpUpG, which was from Boehringer Mannheim, and CpCpC, which was from Collaborative Research.

Stepwise synthesis of oligonucleotides varying in position -3

The oligonucleotide ApUpG$\overset{*}{p}$Cp was first synthesized by ligation of [5'-^{32}P]

pCp to ApUpG. The donor [5'-^{32}P]pCp was prepared in a reaction containing 40 mM Tris·HCl (pH 8.5), 10 mM dithiothreitol, 10 mM $MgCl_2$, 2 mM spermine, 1 µg bovine serum albumin, 0.2 mM 3'-CMP, 300 µCi of γ-^{32}P-ATP (New England Nuclear, specific activity adjusted to 300 Ci/mmol), and 3 U of polynucleotide kinase (P-L Biochemicals). After incubation at 37°C for 40 min, the reaction was terminated by boiling for 2.5 min. Following inactivation of the polynucleotide kinase, the solution containing [5'-^{32}P]pCp was used immediately in a reaction with RNA ligase. The 40 µl reaction mixture contained 20 µl of (boiled) kinase reaction, 0.6 OD_{260} units of ApUpG triplet, 15 U of RNA ligase (from T4 phage-infected E. coli; P-L Biochemicals product 0880), 0.13 mM ATP, 10% dimethylsulfoxide (Mallinckrodt), additional $MgCl_2$ to give a final concentration of 20 mM, and additional Tris·HCl to give a final concentration of 50 mM. Incubation was carried out at 4°C for 18-20 hr, as recommended by Bruce and Uhlenbeck (18). Unless otherwise indicated in the text, these reaction conditions permitted quantitative ligation of the ^{32}P-labeled donor to the acceptor oligonucleotide. The product of the ligase reaction was purified by phenol extraction followed by electrophoresis on Whatman 3MM paper in pyridine/acetate buffer at pH 3.5. ApUpG$\overset{*}{p}$Cp migrates slightly faster than the xylene cyanol marker. The ^{32}P-labeled oligonucleotide was eluted from the paper with water, further purified by chromatography on Bio-Gel P-2, then stored in water at -70°C.

Sequential kinase/ligase reactions were carried out a second time to obtain XpCpCpApUpG$\overset{*}{p}$Cp, where X = C, A or G. The tetranucleotide ApUpG$\overset{*}{p}$Cp from the preceding step was first phosphorylated by incubation with polynucleotide kinase and 0.2 mM (nonradioactive) ATP; the resulting pApUpG$\overset{*}{p}$Cp was used as donor in a reaction with RNA ligase. Three ligase reactions were carried out, using as acceptor either CpCpC, ApCpC or GpCpC; reaction conditions were as described in the preceding paragraph. After phenol extraction, the ^{32}P-labeled heptanucleotides were recovered by precipitation from 70% ethanol.

The kinase/ligase reactions were repeated a third time to obtain CpCpCp-XpCpCpApUpG$\overset{*}{p}$Cp. The heptanucleotide ApCpCpApUpG$\overset{*}{p}$Cp, GpCpCpApUpG$\overset{*}{p}$Cp or CpCpCp-ApUpG$\overset{*}{p}$Cp was first phosphorylated in a standard kinase reaction with nonradioactive ATP. In the subsequent ligase reaction, pXpCpCpApUpG$\overset{*}{p}$Cp (X = C, A or G) served as the donor, with CpCpC (0.6 OD_{260} units/25 µl reaction) as acceptor.

The size and purity of various oligonucleotides was checked by homochromatography on DEAE thin layer plates at 60°C using homomixture c (19). Autoradiography was carried out at room temperature using Kodak BB-1 film.

Synthesis of oligonucleotides varying in position +4

^{32}P-Labeled pentanucleotides of the form ApApUpGpX (where X = C, A, G or U)

were synthesized by ligating $\overset{*}{p}$GpX to ApApU. The 5'-phosphorylated dinucleotide donor was first prepared in a reaction with polynucleotide kinase and γ-^{32}P-ATP. Reaction conditions were as described above for synthesis of $\overset{*}{p}$Cp, except that 3'-CMP was replaced with either GpC, GpA, GpG or GpU, each at a concentration of 10 µg/25 µl reaction. After the standard boiling procedure to inactivate kinase, the solution containing [5'-^{32}P]pGpX was used in a ligase reaction, with ApApU (0.6 units/25 µl reaction) as acceptor. These ligase reactions were incomplete, and therefore at the end of the incubation the ^{32}P-labeled pentanucleotides were separated from unreacted donor by preparative electrophoresis on 3MM paper. The pentanucleotides were eluted with water, and a portion of each sample was used directly for ribosome binding. Another aliquot was used as acceptor in a second ligase step, with nonradioactive pCp as donor. This generated a series of hexanucleotides of the form ApApU$\overset{*}{p}$GpXpCp, where X = C, A, G or U.

The structure of the pentanucleotides was confirmed by a series of enzymatic digestions: P1 nuclease yielded 5'-^{32}P-GMP; T2 RNase yielded 3'-^{32}P-UMP; pancreatic RNase yielded a ^{32}P-labeled product that co-migrated with ApApUp on DEAE paper at pH 3.5. The mobilities of the intact dinucleotides and pentanucleotides on 3MM paper at pH 3.5 (see Fig. 2) are consistent with their composition; that is, the mobilities follow the expected gradient U>G>A>C.

Binding of oligonucleotides to wheat germ ribosomes

The ability of ^{32}P-labeled oligonucleotides to form 80S initiation complexes was measured using conditions that were previously employed with natural messenger RNAs (20). The wheat germ S23 extract was supplemented with 1 mM ATP, 0.24 mM GTP, 200 µM sparsomycin, and other components as described (20). The final concentration of magnesium was 2.8 mM. After incubation for 10 min at 19°C, samples were chilled and layered onto 10-30% glycerol gradients. Centrifugation was for 3 hr, 39,000 rpm, 4°C in the Beckman SW41 rotor. Gradient fractions (0.4 ml) were mixed with 0.4 ml of water and 8 ml of Beckman HP scintillation cocktail, and were counted to determine the distribution of radioactivity.

RESULTS

A survey of nucleotides flanking the initiator codon in eukaryotic mRNAs

If nucleotides bordering the AUG codon are involved in defining the ribosome binding site, comparison of a large number of initiation sites might reveal a conserved sequence. The simplest approach is to examine the frequency of occurrence of a given nucleotide in each position to the left and right of

the AUG codon. Figure 1 presents such an analysis for 106 eukaryotic messages. The region selected for analysis encompasses 15 nucleotides preceding the initiator codon, and the first 15 nucleotides of the coding region. This is the portion of the message that would be protected against nuclease by an 80S ribosome bound at the initiator AUG. There are a few remarkable features in the nucleotide distribution shown in Figure 1. Within the *noncoding* region (residues -1 to -15), the greatest disproportion occurs close to the AUG triplet: 80% of eukaryotic messages have an A in position -3, and C residues occur with high frequency in positions -1, -2, -4 and -5. The G content is unusually low throughout region -1 to -15. In the *coding* region depicted in Figure 1, the nucleotide distribution is approximately random except for positions +4 (60% G), +5 (44% C) and +6 (42% U). The significance of these asymmetries is not immediately obvious, nor is the problem easily approached experimentally. To simplify the task, I have focused in this study on the two positions which show the greatest deviation: position -3, which is a purine (most often A) in 94% of the messages examined; and position +4, which is also a purine (most often G) in 83% of eukaryotic messages.

In attempting to assign a function to these conserved nucleotides, one is confronted by the problem that some messages (albeit a minority) lack the "consensus sequence." One solution that comes to mind is that, in order to

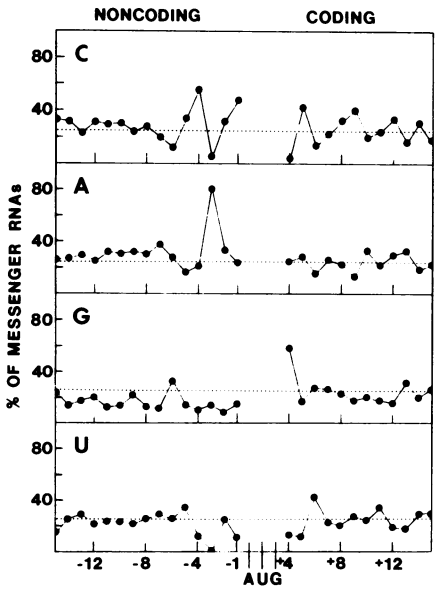

Figure 1. Frequency of occurrence of cytosine (C), adenine (A), guanine (G) and uracil (U) in positions -1 to -15 (preceding the AUG initiator codon), and in coding positions +4 to +15. The AUG triplet is numbered +1 to +3. The dotted line across each panel indicates the 25% value that would be expected on a random basis. For these calculations, I used 90 eukaryotic messages tabulated in reference 17, and 16 additional sequences given in references 21-37. I did not include the human β- and δ-globin sequences, which are very similar to rabbit β-globin; the minor species of mouse β-globin, which is nearly identical to mouse β-globin major; and sequences from the New Jersey strain of vesicular stomatitis virus, which are similar to the Indiana strain. SV40 late 16S mRNA was also omitted because of ambiguity in identifying the initiator codon for VP1.

TABLE 1. GROUPING OF EUKARYOTIC mRNAs ACCORDING TO SEQUENCES FLANKING THE AUG CODON

	Distribution of Functional AUG Initiator Codons[a]		Distribution of nonfunctional AUGs in 5'-Noncoding Region[b]
AXXAUGG	-chicken ovomucoid	-4 actin genes in Dictyostelium (42)	none
	-chicken ovalbumin	-yeast actin, glyceraldehyde dehydro-	
	-rabbit α- and β-globins	genase, iso-2-cytochrome c (14),	
	-mouse α-globin	his-4 (43) and enolase (44)	
	-mouse β-globin, major & minor	-human leukocyte interferons A B D (41)	
	-human α, β, δ, and ε-globins	-human Ig V_H gene segment (46)	
	-chicken α-globin (38)	-2 mouse Ig V_H gene segments (21,23)	
	-rat preproinsulin I	-mouse Ig $V_{\lambda II}$ gene segment (25)	
	-chicken preproinsulin	-human chorionic gonadotropin, β (32)	
	-anglerfish preproinsulin	-chicken VLD lipoprotein II (36)	
	-mouse dihydrofolate reductase	-adenovirus-2 hexon protein	
	-S. purpuratus histones H1 H2B H3	-adenovirus-2 protein VI (47)	
	-P. miliaris histone H3	-Rous sarcoma *gag* & *src* (28) proteins	
	-silkmoth chorion gene 10a	-hepatitis B virus surface antigen (31)	
	-rat cytochrome c (45)	-human flu virus proteins NP (48), NS,	
	-satellite tobacco necrosis virus	and HA (subtype H2) (49)	
	-turnip yellow mosaic coat protein	-3 late proteins of polyoma virus	
	-tobacco mosaic virus genome	-VSV NS and L proteins	
	-reovirus proteins σNS and μ1	-VSV (New Jersey strain) N protein	
	-T antigens of polyoma (39),	-poliovirus (26,27)	
	BK virus (40), and SV40	-rat calcitonin precursor (77)	
AXXAUGA	-chicken conalbumin	-alfalfa mosaic virus coat protein	none
	-rat liver α_{2u} globulin (50)	-one of the large reovirus proteins	
	-silkworm fibroin	-VSV M and G proteins	
	-bovine preproparathyroid hormone	-adenovirus-2 fiber protein	
	-P. miliaris histone H1	-Ad-5 and Ad-12 early region E_1a	
	-human fibroblast interferon	-Ad-2 early region 3, 14.5 K (53)	
	-mouse α-amylase, salivary & panc.	and 16K (54) proteins	
	-chicken lysozyme	-Semliki Forest virus capsid (30)	
	-human (51) and rat prolactin	-Sindbis virus capsid protein (29)	
	-yeast iso-1-cytochrome c	-human flu NA: subtypes N2 (55), N1 (56)	
	-silkmoth chorion gene 401a	-human flu HA: subtypes H3 (57), H1 (58)	
	-mouse Ig V_H gene segment (22)	-fowl plague virus matrix protein	
	-flounder antifreeze peptide (52)	-fowl plague virus HA protein (59)	

AXXAUGY	-goat β-globin -chicken β-tubulin -S. purpuratus histones H2A & H4 -P. miliaris histones H2A & H4 -Drosophila 70K heat shock protein -rainbow trout protamine (61) -6 actin genes in Drosophila (62)	-bovine corticotropin/lipotropin (35) -2 histone H2B genes in yeast (60) -2 discoidin I genes in Dictyostelium -2 silkmoth chorion genes, A-family -tobacco mosaic virus coat protein -brome mosaic virus coat protein -VSV N protein (Indiana strain)	none
GXXAUGG	-human γ-globin -chicken β-globin -Xenopus β-globin (65) -human preproinsulin -human and rat growth hormones	-human chorionic gonadotropin α-subunit -reovirus proteins σ2, σ3, μNS, μ2 -Moloney MSV oncogene (63,64) -mouse metallothionein-I (78) -SV40 VP3 (66) and agnogene (66,67)	-SV40 16S and 19S mRNA leaders -poliovirus #6 (ref. 27)
GXXAUGA	-mouse Ig V_K(K2) gene segment (24) -mouse α-fetoprotein (34)	-Ad-2 late polypeptide IX -Ad-2 early region 3, 14K protein(53)	none
GXXAUGY	-rat relaxin (33) -human histone H4 (37)	-chicken type I α_2 collagen (68)	-poliovirus #1
YXXAUGG	-human histone H3 (69) -alfalfa mosaic virus RNA-3 (70) -reovirus σ1 protein	-Ad-5 early region E_1b -human leuko. interferons C,F,H (41) -herpes thymidine kinase; SV40 VP2	-poliovirus #5, #6 (ref. 26), and #8 -Rous sarcoma virus #1
YXXAUGA	-SV40 VP1 ?		-poliovirus #4 -mouse α-amylase, liver -mouse α-amylase, salivary -Rous sarcoma virus #2 and #3 -chicken preproinsulin
YXXAUGY	-brome mosaic virus RNA-3 ?		-poliovirus #2, #3 and #7 -α_2-collagen #1 and #2 -Semliki Forest virus genome

[a]Messages for which no reference is indicated were included in an earlier compilation (17). The abbreviations used are VSV, vesicular stomatitis virus; MSV, murine sarcoma virus; Ad, adenovirus; HA, hemagglutinin; NA, neuraminidase; Ig, immunoglobulin. [b]The rightmost panel deals with mRNAs in which translation does not begin at the 5'-proximal AUG. Eleven such messages have been identified: SV40 late 16S and 19S mRNAs (74,75); the poliovirus genome (26,27); the genome of Rous sarcoma virus (R. Swanstrom & J.M. Bishop, pers.commun.) and two subgenomic mRNAs of RSV (76; J.M. Bishop, pers.commun.); mouse α-amylase mRNAs from salivary gland and liver (71); chicken α_2-collagen (68); chicken preproinsulin (72); and the Semliki Forest virus genome (73). *Nonfunctional* upstream AUGs in these 11 messages are listed on the right; the *functional* initiator codon from each message is included in the left side of the table. Some of these mRNAs have only one AUG preceding the initiator codon; others have 2 (collagen) or 3 (Rous); the poliovirus genome has 8 AUGs preceding the presumptive start site for translation. The upstream AUGs are designated #1 (nearest the 5'-terminus), #2, etc. The poliovirus genome was sequenced independently in two laboratories (26,27). Each group reported a different sequence flanking AUG #6, which is therefore listed twice.

function as an efficient initiation signal, an AUG codon must be flanked by *either* a purine in position -3, or a G in position +4. In other words, at least one of the favored flanking nucleotides is required. To evaluate this idea, I have surveyed and grouped 153 eukaryotic messages on the basis of nucleotides occurring in positions -3 and +4. The data are presented in Table 1. (I attempted to include all eukaryotic messages for which adequate sequence data have been published, and in which the functional initiator codon has been identified. Although sequences bordering the AUG codon have been determined for many messages, in some cases the remainder of the 5'-untranslated sequence is not known. Thus, the number of mRNAs used in compiling Table 1 is larger than the number used in Figure 1. In the case of closely related genes, only one member of the set was chosen for Figure 1, whereas Table 1 is more inclusive.) Of the 153 messages listed in Table 1, 120 have an A in position -3. The largest group (64 messages) has the sequence AXXAUG*G*, but AXXAUG*A* (32 mRNAs) and AXXAUG*Y* (24 mRNAs) also occur with high frequency at functional initiation sites. The table lists 15 messages in which the initiation site has the sequence GXXAUGG, and another 4 with the sequence GXXAUGA. Thus, in 91% of the messages surveyed (139 out of 153), one finds either an A in position -3 (with G, A or Y in position +4), or one finds a G in position -3 and a purine in position +4. Judging from their high frequency of occurrence, those sequences must function well. The infrequency of finding GXXAUGY or YXXAUGX, on the other hand, might mean that an AUG triplet flanked by those nucleotides functions poorly as an initiation signal.[1]

It is intriguing to note that, in the eleven unusual messages in which translation does *not* begin at the *first* AUG, the pattern of nucleotides flanking the nonfunctional upstream AUG triplets (right side of Table 1) is different from that found at most functional initiation sites. The nonfunctional AUG triplets which occur within the 5'-noncoding regions are clustered in the lower right quadrant of Table 1; i.e., they are bordered by nucleotides which are rarely seen around functional initiator codons. The only serious exceptions are SV40 16S and 19S mRNAs, in which the upstream AUG triplets in the leader are flanked by GXXAUGG--a sequence frequently found at functional initiation sites. But the first AUG in SV40 16S mRNA really should not be listed with the "nonfunctional" AUGs in the right side of Table 1. It was recently shown that ribosomes do initiate at that site, translating the so-called agnogene (67). The mechanism by which ribosomes are also able to initiate at a downstream AUG triplet to make the VP1 protein is not understood; but the ideas proposed in this paper are not contradicted by finding "good" sequences flank-

ing the first AUG in SV40 16S mRNA. The upstream AUG in that message *is functional*. The only other entry in the GXXAUGG column is the sixth AUG in the poliovirus genome. As explained in the footnote to Table 1, two different sequences were reported for that site. Until the ambiguity can be cleared up, it seems fair to omit AUG #6 from further discussion. With these caveats, the data presented in Table 1 suggest the generalization that nonfunctional AUG triplets, which are found in the 5'-noncoding region of a few eukaryotic messages, are bordered by sequences (GXXAUGY or YXXAUGX) which differ from those found around most functional initiator codons.

The main conclusion from this survey is that nucleotides flanking functional initiator codons in eukaryotic messages are not random. Purines occur with very high frequency in positions -3 and +4. As a first step in determining whether the conserved flanking nucleotides play a role during initiation, I have carried out in vitro ribosome binding studies using various synthetic oligonucleotides.

Binding of ^{32}P-labeled oligonucleotides to wheat germ ribosomes

Effect of varying the nucleotide in position +4

Oligonucleotides of the form ApApUp*GpX (where X = C, A, G or U) were constructed by ligating p*GpX to the triplet ApApU. The efficiency of ligation was approximately 40% for reactions with p*GpC and p*GpA, and somewhat less than that for reactions with p*GpG and p*GpU, as shown in Figure 2. The ^{32}P-labeled pentanucleotides, purified by electrophoretic fractionation, were incubated with wheat germ ribosomes under conditions that permitted formation of 80S initiation complexes. As shown in Table 2 (experiments 1 and 2), the efficiency of binding varied from 0.5% for ApApUp*GpU, to 7-10% for ApApUp*GpG.

The pentanucleotide series was converted to hexanucleotides by ligating pCp to the 3'-terminus. The chromatographic analysis shown in Figure 3 reveals that the ligation was quantitative: all ^{32}P-radioactivity was converted to the slower-migrating hexanucleotide position. The overall efficiency of binding to ribosomes was higher with the hexanucleotide series than with the pentanucleotides. Table 2 (experiment 3) shows that, within the hexanucleotide series ApApUp*GpXpCp, varying the nucleotide in position +4 produced a gradient in binding efficiency: G>A>C>U. A similar gradient was *not* observed with the control series ApApUp*GpUpXp (X = C, A, G or U; experiment 4). Thus, the efficiency of binding was markedly enhanced by placing a purine in position +4, but not in position +5.

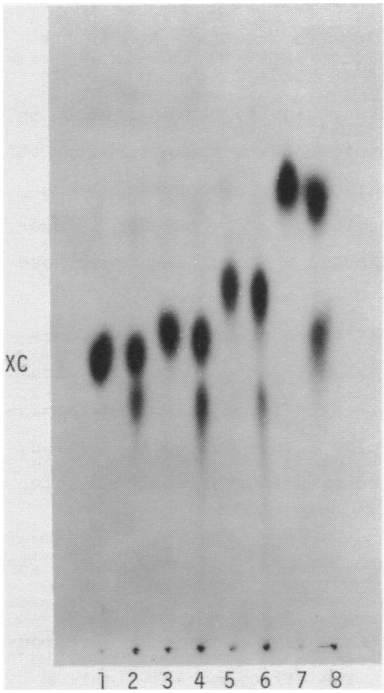

Figure 2. Synthesis of ApApUpGpX series of oligonucleotides. Odd-numbered lanes show the ^{32}P-labeled dinucleotides formed in the first-step kinase reactions: lane 1, pGpC; lane 3, pGpA; lane 5, pGpG; lane 7, pGpU. The even-numbered lanes show the products of the second-step ligase reactions, with nonradioactive ApApU as acceptor: lane 2, ApApUpGpC; lane 4, ApApUpGpA; lane 6, ApApUpGpG; lane 8, ApApUpGpU. The pentanucleotide is the slower spot in lanes 2, 4, 6 and 8; the faster spot in each lane represents residual unligated dinucleotide. Products were fractionated by electrophoresis on Whatman 3MM paper at pH 3.5. An autoradiogram is shown. XC - xylene cyanol marker. The origin is at the bottom.

TABLE 2. Binding of ^{32}P-labeled oligonucleotides to wheat germ 80S ribosomes: effect of varying the nucleotide in position +4 or +5

Experiment	^{32}P-labeled pentanucleotide	Percent bound[a]	Experiment	^{32}P-labeled hexanucleotide	Percent bound[a]
1	ApApUpGpC	3	3	ApApUpGpCpCp	12
	ApApUpGpA	4		ApApUpGpApCp	18
	ApApUpGpG	7		ApApUpGpGpCp	23
	ApApUpGpU	0.5		ApApUpGpUpCp	6
2	ApApUpGpU	0.5	4	ApGpUpGpGpCp (control)	0.3
	ApApUpGpG	9.5		ApApUpGpGpCp	24
	ApGpUpGpG (control)	0		ApApUpGpUpCp	5
				ApApUpGpUpAp	1
				ApApUpGpUpGp	3
				ApApUpGpUpUp	1

[a]After incubation for 10 min at 19°C, each 50 µl reaction mixture was centrifuged through a glycerol gradient. The ^{32}P-radioactivity co-sedimenting with 80S ribosomes is shown as a percent of the total radioactivity recovered. Each sample contained 30,000 cpm of ^{32}P-oligonucleotide.

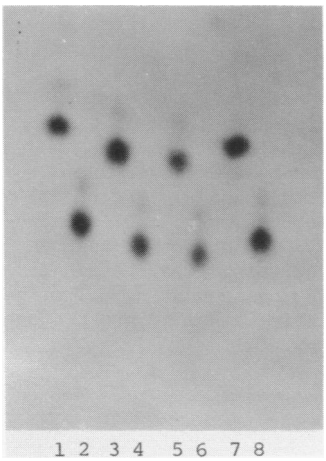

Figure 3. Addition of pCp to the 3'-end of the pentanucleotide series ApApUp*GpX. The ^{32}P-labeled pentanucleotides ApApUp*GpC, ApApUp*GpA, ApApUp*GpG and ApApUp*GpU are shown in lanes 1, 3, 5 and 7, respectively. The products obtained after ligation to (nonradioactive) pCp are shown in lanes 2, 4, 6 and 8. Fractionation was by homochromatography on a DEAE-cellulose thin layer plate. The origin is at the bottom.

Effect of varying the nucleotide in position -3

The range of oligonucleotides that could be constructed for this study was limited by availability of the required trinucleotide precursors and, more importantly, by the specificity of RNA ligase. As reported previously (79,80), the enzyme has marked sequence preferences with respect to both donor and acceptor molecules. Molecules with 5'-terminal cytidine are, by far, the most efficient donors in reactions catalyzed by T4 RNA ligase. The best trinucleotide acceptor is ApApA; ApCpC functions adequately as acceptor if the enzyme and substrate concentrations are high; ApUpU and UpUpU were very inefficient in preliminary experiments, and therefore they were abandoned. Acceptor activity depends not only on the 3'-terminal residue, but also on the adjacent nucleotides.

To test the idea that a purine in position -3 might enhance binding of AUG-containing oligonucleotides, I first constructed a series of heptanucleotides of the form XpCpCpApUpGp*Cp, where X = C, A or G. These were prepared by ligating ^{32}P-labeled pApUpGp*Cp to either CpCpC, ApCpC or GpCpC. Figure 4 (lanes 2-4) shows that all of the ^{32}P-labeled donor was converted to slower-migrating heptanucleotides. When these oligonucleotides were tested for ability to bind to wheat germ ribosomes, ApCpCpApUpGp*Cp and GpCpCpApUpGp*Cp were slightly more efficient than CpCpCpApUpGp*Cp, as shown in Table 3.

It seemed possible that the stabilizing effect of a purine in position -3 might be more obvious if the purine were not right at the end of the oligonucleotide. To obtain longer templates, I ligated pXpCpCpApUpGp*Cp (X = C, A or G) to CpCpC. The reactions with pCpCpCpApUpGp*Cp and pApCpCpApUpGp*Cp proceeded to

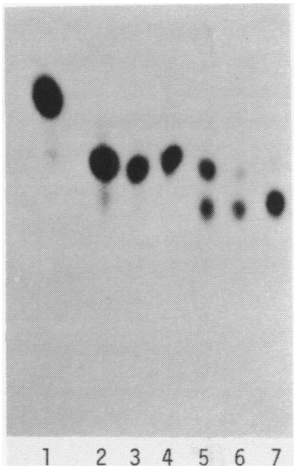

Figure 4. Sequential joining of ^{32}P-labeled pApUpGp*Cp to XpCpC (X = G, A or C), and then to CpCpC. Lane 1, ApUpGp*C; lane 2, GpCpCp-ApUpGp*C; lane 3, ApCpCpApUpGp*C; lane 4, CpCpCp-ApUpGp*C; lane 5, CpCpCpGpCpCpApUpGp*C (slower spot) and residual GpCpCpApUpGp*C; lane 6, CpCpCpApCpCpApUpGp*C; lane 7, $(Cp)_6$ApUpGp*C. The reactions catalyzed by RNA ligase are described in Materials and Methods. Fractionation was by homochromatography on a DEAE-cellulose thin layer plate. The origin is at the bottom. To enhance the resolution, terminal phosphates were removed by incubating all oligonucleotides with bacterial alkaline phosphatase prior to chromatography.

completion, as shown in Figure 4, lanes 6 and 7. With pGpCpCpApUpGp*Cp as donor, however, the extent of joining was only 40% after the first ligation (Figure 4, lane 5). The reaction with pGpCpCpApUpGp*Cp was complete after a second incubation with RNA ligase (data not shown). The structure of the oligonucleotides $(Cp)_6$ApUpGp*Cp and $(Cp)_3$ ApCpCpApUpGp*Cp was confirmed by analyzing the partial T2 RNase digestion products (Figure 5). Ribosome binding experiments were carried out with $(Cp)_6$ ApUpGp*Cp, $(Cp)_3$ApCpCpApUpGp*Cp and $(Cp)_3$GpCpCpApUpGp*Cp. As shown in Table 3, the extent of binding was 7 to 9-fold higher for oligonucleotides with a purine in position -3. Although this effect was readily demon-

TABLE 3. Binding of ^{32}P-labeled oligonucleotides to wheat germ 80S ribosomes: effect of varying the nucleotide in position -3

^{32}P-labeled oligonucleotide	Percent of input oligonucleotide bound[a]		
	Experiment 1	Experiment 2	Experiment 3
ApUpGp*Cp	3	n.d.	n.d.
CpCpCpApUpGp*Cp	3	3	3
ApCpCpApUpGp*Cp	5	5	5
GpCpCpApUpGp*Cp	6	6	n.d.
ApCpCpApUpUp*Cp			0.2
ApCpCpGpUpGp*Cp			0.2
CpCpCpCpCpCpApUpGp*Cp		1	1
CpCpCpApCpCpApUpGp*Cp		9	7
CpCpCpGpCpCpApUpGp*Cp		n.d.	9

[a]The experiments were carried out as described in the footnote to Table 2. Three independent preparations of oligonucleotides were tested in experiments 1, 2 and 3. n.d. - not done.

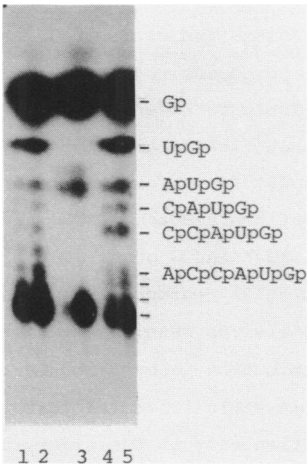

Figure 5. Autoradiogram of products derived by partial hydrolysis of $(Cp)_6ApUpG\overset{*}{p}$ (lanes 1,2) and $(Cp)_3ApCpCpApUpG\overset{*}{p}$ (lanes 4,5). Samples were incubated with T2 RNase (5 U/ml) for 15 min (lanes 1,4) or 45 min (lanes 2,5). Partial digestion products were fractionated by homochromatography on DEAE-cellulose. The uniform spacing between spots 3 through 9 in lane 2 is as expected for $(Cp)_6ApUpG\overset{*}{p}$. In lanes 4 and 5, the larger shift between spot 5 ($CpCpApUpG\overset{*}{p}$) and spot 6 ($ApCpCpApUpG\overset{*}{p}$) is consistent with the proposed sequence. [In a homologous partial digestion series, the distance between two oligonucleotides which differ by a purine nucleotide is always larger than that between two oligonucleotides which differ by a pyrimidine nucleotide (81).] The markers in lane 3 are $G\overset{*}{p}$, $ApUpG\overset{*}{p}$ and $(Cp)_6ApUpG\overset{*}{p}$. Prior to carrying out the analysis shown in this figure, the oligonucleotides were digested with T_1 RNase to remove 3'-terminal Cp, leaving ^{32}P directly at the 3'-end of the T_1-derived product. This simplified identification of partial digestion products subsequently obtained with T2 RNase.

strated using the small oligonucleotides described above, with longer templates, such as $(Cp)_{12}ApUpGpCp$, it was more difficult to show dependence on a purine in position -3. The problem is that longer oligonucleotides bound with considerably higher efficiency (25 to 45%) *seemingly due to just their length.* Not surprisingly, it is difficult to demonstrate a stabilizing effect due to changing one nucleotide when binding has already been dramatically enhanced by the length-effect.

The controls listed in Tables 2 and 3 indicate that, under the conditions of these experiments, only AUG-containing oligonucleotides were able to bind to ribosomes. ApGpU$\overset{*}{p}$GpG and ApGpU$\overset{*}{p}$GpGpCp (Table 2), as well as ApCpCpGpU$\overset{*}{p}$Cp and ApCpCpApU$\overset{*}{p}$Cp (Table 3) showed negligible binding.

DISCUSSION

A survey of sequences flanking the initiator codon in eukaryotic messenger RNAs reveals that almost all functional AUG triplets are preceded by a purine (usually A) in position -3. A high proportion of functional initiation sites also have a purine (usually G) following the AUG codon. From the data summarized in Table 1, $\begin{matrix}A\\G\end{matrix}$XXAUGG emerges as the favored sequence for eukaryotic initiation sites. Of the 153 messages included in the survey, only 11 have a pyrimidine in position -3, and in 9 of those the AUG codon is followed by G. Thus, only 2 putative initiation sites[2]--those of SV40 VP1 and brome mosaic virus

RNA-3--lack *both* a purine in position -3 and a G residue in position +4. The sequence $\begin{matrix}A\\G\end{matrix}$XXAUGG which characterizes most functional initiation sites is *never* observed among the nonfunctional AUGs found in the 5'-noncoding region of eukaryotic messages. It is obvious from the data in Table 1, however, that differences in sequences flanking the AUG triplet do not categorically distinguish functional from nonfunctional sites. Although most functional initiator codons are preceded by a purine in position -3 and most nonfunctional AUGs have a pyrimidine in that position, the sequences GXXAUGY and YXXAUGG occur at a small number of functional sites as well as at (presumably) nonfunctional AUGs within the leader region of poliovirus and Rous sarcoma virus RNAs. A mechanism of initiation compatible with this nonunique distribution is outlined below. The main conclusions from the survey presented in Table 1 are (a) that nucleotides flanking *functional* initiator codons (particularly in positions -3 and +4) are not random; and (b) that nucleotides flanking *nonfunctional* AUG triplets which occur within the 5'-untranslated region of a few eukaryotic messages are different from those bordering most functional initiator codons.

This asymmetry suggests that purines in positions -3 and +4 might facilitate recognition of the AUG codon during formation of initiation complexes. The idea gains support from the oligonucleotide binding studies described above. The extent of binding to wheat germ ribosomes was increased several-fold by placing a purine, rather than a pyrimidine, in position +4 (Table 2). The facilitating effect of a purine in position -3 was also readily demonstrated (Table 3), particularly with the series CpCpCpXpCpCpApUpG$\overset{*}{p}$Cp (X = C, A or G). Since only a small number of permutations were tested in this study, it might be that nucleotides in positions other than those tested also contribute to the stability of initiation complexes. But it is encouraging that differences in oligonucleotide binding can be detected upon varying the component in position -3 or +4; a purine seems to be preferred in both positions.

These results cannot be interpreted in isolation. Although the data in this paper suggest that recognition of the initiator codon is influenced by the flanking sequences, the *position* of the AUG triplet (i.e., near the 5'-end of the message) still seems to be the primary determinant of a functional initiation site (16,17). Within the interior of eukaryotic messenger RNAs the sequence $\begin{matrix}A\\G\end{matrix}$XXAUGG occurs many times; ribosomes do not initiate at these internal AUGs despite the favorable flanking sequence. In apparent contradiction to the studies described above, some previous experiments seemed to indicate that eukaryotic ribosomes select the AUG initiator codon *without regard to the flanking sequences*. Sherman and colleagues (9) showed, for example, that when the

normal initiator codon in the yeast cytochrome c gene was inactivated by mutation, introduction of a new AUG triplet almost anywhere within a 37-nucleotide region restored translation. The sequences flanking the ectopic AUG initiator codons in the pseudorevertants varied widely, and often did not correspond to the optimal sequences defined above. How can those experiments be reconciled with the new data described herein? The interpretation I favor is a modified version of the scanning mechanism; namely, that flanking sequences (nucleotides -3 and +4) *modulate the efficiency* with which the migrating 40S ribosomal subunit recognizes an AUG triplet as a "stop signal." *Some* 40S subunits will stop at the first AUG *irrespective of the flanking sequences;* if the nucleotides bordering an AUG codon are *optimal*, virtually *all* 40S subunits will stop at that AUG. Sherman's data are compatible with such a mechanism: *some* cytochrome c is made in the pseudorevertants, indicating (only) that *some* 40S subunits stop at the first AUG, even when it is flanked by suboptimal sequences. Those experiments do not indicate that the ectopic AUG triplets in the pseudorevertants function as efficiently as the wild-type sequence AUAAUGA presumably does. Thus, the genetic experiments are not incompatible with a modified scanning mechanism in which flanking sequences[3] affect the efficiency with which an AUG codon is recognized as a "stop signal."

The proposed mechanism has some interesting implications:

(a) The scanning model in its simplest form (see Introduction) predicts that spurious AUG triplets cannot occur in the region preceding the functional initiation site; recognition of the authentic initiator codon depends on its being first-in-line. This prediction is upheld by most, but not all, eukaryotic messages. The modified scanning model, on the other hand, admits that ribosomes can initiate at a downstream site, provided that all of the upstream AUGs are flanked by "unfavorable" sequences such that *some* 40S ribosomes can get through. The data in the right side of Table 1 provide encouragement for this idea. In those messages in which translation does not begin at the first AUG, almost all of the AUGs which are bypassed have a pyrimidine in position -3.

(b) Although upstream AUGs are not an absolute barrier to initiating downstream, it seems reasonable to expect that occurrence of AUG triplets within the 5'-noncoding region of a message would reduce translational efficiency, since *some* 40S ribosomes would stop at each upstream AUG irrespective of the flanking sequences. Consistent with this idea, poliovirus RNA is a notoriously inefficient message in vitro (84), as is the Rous sarcoma virus genome (85). The notion that upstream AUGs impair translational efficiency would explain why 90% of eukaryotic mRNAs have *no* AUGs in the 5'-noncoding region. Even if the

flanking sequences are such that some ribosomes can get through, upstream AUGs probably have a deleterious effect. Parenthetically, the idea that some 40S subunits stop and initiate at each upstream AUG rationalizes the finding that, in those few messages which have AUG triplets in the "5'-noncoding region" (the semantic difficulty is obvious), the upstream AUGs are almost always followed closely by in-phase terminator codons (17). Thus, ribosomes which initiate prematurely at an upstream AUG are returned quickly to circulation.
(c) The apparent inability of eukaryotic ribosomes to bind to sites in the interior of a message generally means that a eukaryotic mRNA can direct synthesis of only one protein--that encoded nearest the 5'-end of the template. But a single message might direct synthesis of two proteins if the AUG triplet at the start of the first coding region were flanked by unfavorable sequences, such that only some 40S subunits stop and initiate at that site while others advance to the next AUG. Two examples come to mind: SV40 late 19S mRNA is believed to direct synthesis of both VP2 and VP3 (75, 86); and the mRNA encoding herpes thymidine kinase has been shown to direct synthesis of a second smaller protein (87). In both messages the upstream initiator codon, which seems to be "leaky," is preceded by a pyrimidine in position -3: the initiation site for SV40 VP2 is UCCAUGG, and the putative initiation site for herpes thymidine kinase is CGUAUGG.

In summary, it seems likely that nucleotides in positions -3 and +4 influence recognition of the AUG codon by eukaryotic ribosomes, or one of the ribosome-associated components involved in initiation. Binding of synthetic oligonucleotides to wheat germ ribosomes was enhanced 5- to 15-fold by placing a purine in either of those positions. This preference mirrors the observed frequency of nucleotides flanking the initiator codon in natural mRNAs. Although there is no direct evidence concerning the mechanism by which nucleotides bordering the AUG codon facilitate initiation, the mechanism must be compatible with a large body of evidence which suggests that ribosomes attach to eukaryotic mRNAs at an upstream site, and migrate down to the AUG. I have therefore proposed a modified version of the scanning model which postulates that flanking nucleotides modulate the efficiency with which an AUG triplet is recognized as a stop-signal by the migrating 40S subunit. The modified scanning mechanism accounts for those few messages (eleven, to date) in which translation does not begin at the AUG triplet closest to the 5'-terminus, and also for rare messages which seem to direct synthesis of two independently-initiated proteins.

ACKNOWLEDGMENTS

I am grateful to Drs. J.M. Bishop, R. Swanstrom, B. deCrombrugghe and V. Racaniello for providing preprints of their unpublished sequences. The ubiquitous occurrence of adenosine 3 nucleotides before the initiator codon was first pointed out to me by Dr. David Baltimore. This work was supported by grant AI 16634 and Career Development Award AI 00380 from the National Institutes of Health.

FOOTNOTES

[1]The suggested cut-off between efficient and inefficient initiation signals has been set rather arbitrarily between GXXAUGA (which I have called efficient) and GXXAUGY (which I have called inefficient). This division obviously does not follow from the number of functional initiation sites which have those sequences. Both sequences occur infrequently, and therefore both might be viewed as likely-to-be-inefficient. On the other hand, there is no *evidence* that the 4 messages initiating at GXXAUGA or the 3 messages initiating at GXXAUGY (see Table 1) are defective; thus, one might argue that--despite their infrequent occurrence--both GXXAUGA and GXXAUGY should be regarded as efficient. To some extent, I have been guided by the oligonucleotide binding studies which show a purine in position +4 to be much better than a pyrimidine in that site.

[2]There is some uncertainty in pinpointing the initiator codon in these 2 messages. In the case of brome mosaic virus RNA-3, the putative initiation site was identified by the open reading frame that follows; but only a limited portion of that RNA has been sequenced (82). The difficulty in identifying the initiator codon for SV40 VP1 has been discussed previously (17). It is not known whether translation of VP1 in vivo begins at the first or second AUG within the sequence CUUAUGAAGAUGGCC.

[3]Although this paper emphasizes the role of the flanking *primary* sequence in modulating recognition of the AUG codon, other experiments (83) suggest that the secondary/tertiary structure of eukaryotic messages also contributes to the fidelity of initiation. With *denatured* reovirus mRNA as template, 40S ribosomes tend to migrate beyond the first AUG codon--despite the favorable flanking sequences.

REFERENCES

1. Steitz, J.A. (1979) in Biological Regulation and Development, Goldberger, R.F., Ed., pp. 349-399, Plenum Press, New York.
2. Borisova, G.P., Volkova, T.M., Berzin, V., Rosenthal, G. and Gren, E.J. (1979) Nucleic Acids Res. 6, 1761-1774.
3. Cannistraro, V.J. and Kennell, D. (1979) Nature 277, 407-409.
4. Fiil, N.P., Friesen, J.D., Downing, W.L. and Dennis, P.P. (1980) Cell 19, 837-844.
5. Taniguchi, T. and Weissmann, C. (1978) J. Mol. Biol. 118, 533-565.
6. Atkins, J.F., Steitz, J.A., Anderson, C.W. and Model, P. (1979) Cell 18, 247-256.
7. Dunn, J.J., Buzash-Pollert, E. and Studier, F.W. (1978) Proc. Natl. Acad. Sci. USA 75, 2741-2745.
8. Schwartz, M., Roa, M. and Debarbouille, M. (1981) Proc. Natl. Acad. Sci. USA 78, 2937-2941.
9. Sherman, F., Stewart, J.W. and Schweingruber, A.M. (1980) Cell 20, 215-222.
10. Ghosh, P.K., Lebowitz, P., Frisque, R.J. and Gluzman, Y. (1981) Proc. Natl. Acad. Sci. USA 78, 100-104.
11. Barkan, A. and Mertz, J.E. (1981) J. Virol. 37, 730-737.
12. Solnick, D. (1981) Cell 24, 135-143.

13. Firtel, R.A., Timm, R., Kimmel, A.R. and McKeown, M. (1979) Proc. Natl. Acad. Sci. USA 76, 6206-6210.
14. Montgomery, D.L., Leung, D.W., Smith, M., Shalit, P., Faye, G. and Hall, B.D. (1980) Proc. Natl. Acad. Sci. USA 77, 541-545.
15. Jones, C.W. and Kafatos, F.C. (1980) Cell 22, 855-867.
16. Kozak, M. (1980) Cell 22, 7-8.
17. Kozak, M. (1981) in Current Topics in Microbiology and Immunology, Shatkin, A.J., Ed., Vol. 93, pp. 81-123, Springer-Verlag, Berlin.
18. Bruce, A.G. and Uhlenbeck, O.C. (1978) Nucleic Acids Res. 5, 3665-3677.
19. Barrell, B.G. (1971) in Procedures in Nucleic Acid Research, Cantoni, G.L. and Davies, D.R., Eds. Vol.2, pp. 751-779.
20. Kozak, M. and Shatkin, A.J. (1976) J. Biol. Chem. 251, 4259-4266.
21. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S. (1980) Nature 286, 676-683.
22. Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) Cell 19, 981-992.
23. Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1981) Cell 24, 625-637.
24. Nishioka, Y. and Leder, P. (1980) J. Biol. Chem. 255, 3691-3694.
25. Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 1485-1489.
26. Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emini, E.A., Hanecak, R., Lee, J.J., van der Werf, S., Anderson, C.W. and Wimmer, E. (1981) Nature 291, 547-553.
27. Racaniello, V.R. and Baltimore, D. (1981) Proc. Natl. Acad. Sci. USA, in press.
28. Czernilofsky, A.P., Levinson, A.D., Varmus, H.E., Bishop, J.M., Tischer, E. and Goodman, H.M. (1980) Nature 287, 198-203.
29. Rice, C.M. and Strauss, J.H. (1981) Proc. Natl. Acad. Sci. USA 78, 2062-2066.
30. Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H. and Delius, H. (1980) Proc. Natl. Acad. Sci. USA 77, 6376-6380.
31. Valenzuela, P., Gray, P., Quiroga, M., Zaldivar, J., Goodman, H.M. and Rutter, W.J. (1979) Nature 280, 815-819.
32. Fiddes, J.C. and Goodman, H.M. (1980) Nature 286, 684-687.
33. Hudson, P., Haley, J., Cronk, M., Shine, J. and Niall, H. (1981) Nature 291, 127-131.
34. Law, S.W. and Dugaiczyk, A. (1981) Nature 291, 201-205.
35. Nakanishi, S., Teranishi, Y., Watanabe, Y., Notake, M., Noda, M., Kakidani, H., Jingami, H. and Numa, S. (1981) Eur. J. Biochem. 115, 429-438.
36. Wieringa, B., Geert, A.B. and Gruber, M. (1981) Nucleic Acids Res. 9, 489-501.
37. Heintz, N., Zernik, M. and Roeder, R.G. (1981) Cell 24, 661-668.
38. Richards, R.I. and Wells, J.R.E. (1980) J. Biol. Chem. 255, 9306-9311.
39. Friedmann, T., LaPorte, P. and Esty, A. (1978) J. Biol. Chem. 253, 6561-6567.
40. Yang, R.C.A. and Wu, R. (1979) Virology 92, 340-352.
41. Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss, R., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W. (1981) Nature 290, 20-26.
42. McKeown, M. and Firtel, R.A. (1981) Cell 24, 799-807.
43. Farabaugh, P.J. and Fink, G.R. (1980) Nature 286, 352-356.
44. Holland, M.J., Holland, J.P., Thill, G.P. and Jackson, K.A. (1981) J. Biol. Chem. 256, 1385-1395.
45. Scarpulla, R.C., Agne, K.M. and Wu, R. (1981) J. Biol. Chem. 256, 6480-6486.

46. Matthyssens, G. and Rabbitts, T.H. (1980) Proc. Natl. Acad. Sci. USA 77, 6561- 6565.
47. Akusjarvi, G. and Persson, H. (1981) J. Virol. 38, 469-482.
48. Van Rompuy, L., Min Jou, W., Huylebroeck, D., Devos, R. and Fiers, W. (1981) Eur. J. Biochem. 116, 347-353.
49. Air, G.M. (1979) Virology 97, 468-472.
50. Drickamer, K., Kwoh, T.J. and Kurtz, D.T. (1981) J. Biol. Chem. 256, 3634-3636.
51. Cooke, N.E., Coit, D., Shine, J., Baxter, J.D. and Martial, J.A. (1981) J. Biol. Chem. 256, 4007-4016.
52. Lin, Y. and Gross, J.K. (1981) Proc. Natl. Acad. Sci. USA 78, 2825-2829.
53. Hérissé, J. and Galibert, F. (1981) Nucleic Acids Res. 9, 1229-1240.
54. Hérissé, J., Courtois, G. and Galibert, F. (1980) Nucleic Acids Res. 8, 2173-2192.
55. Blok, J. and Air, G.M. (1980) Virology 107, 50-60.
56. Fields, S., Winter, G. and Brownlee, G.G. (1981) Nature 290, 213-217.
57. Min Jou, W., Verhoeyen, M., Devos, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., Barber, C., Carey, N. and Emtage, S. (1980) Cell 19, 683-696.
58. Winter, G., Fields, S. and Brownlee, G.G. (1981) Nature 292, 72-75.
59. Porter, A.G., Barber, C., Carey, N.H., Hallewell, R.A., Threlfall, G. and Emtage, J.S. (1979) Nature 282, 471-477.
60. Wallis, J.W., Hereford, L. and Grunstein, M. (1980) Cell 22, 799-805.
61. Jenkins, J.R. (1979) Nature 279, 809-811.
62. Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S. and Davidson, N. (1981) Cell 24, 107-116.
63. Reddy, E.P., Smith, M.J., Canaani, E., Robbins, K.C., Tronick, S.R., Zain, S. and Aaronson, S.A. (1980) Proc. Natl. Acad. Sci. USA 77, 5234-5238.
64. VanBeveren, C., Galleshaw, J.A., Jonas, V., Berns, A.J.M., Doolittle, R.F., Donoghue, D.J. and Verma, I.M. (1981) Nature 289, 258-262.
65. Williams, J.G., Kay, R.M. and Patient, R.K. (1980) Nucleic Acids Res. 8, 4247-4258.
66. Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978) Science 200, 494-502.
67. Jay, G., Nomura, S., Anderson, C.W. and Khoury, G. (1981) Nature 291, 346-349.
68. Vogeli, G., Ohkubo, H., Sobel, M.E., Yamada, Y., Pastan, I. and deCrombrugghe, B. (1981) Proc. Natl. Acad. Sci. USA, in press.
69. Clark, S.J., Krieg, P.A. and Wells, J.R.E. (1981) Nucleic Acids Res. 9, 1583-1590.
70. Pinck, M., Fritsch, C., Ravelonandro, M., Thivent, C. and Pinck, L. (1981) Nucleic Acids Res. 9, 1087-1100.
71. Hagenbüchle, O., Tosi, M., Schibler, U., Bovey, R., Wellauer, P.K. and Young, R.A. (1981) Nature 289, 643-646.
72. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555-566.
73. Wengler, G., Wengler, G. and Gross, H.J. (1979) Nature 282, 754-756.
74. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Choudary, P.V., Lebowitz, P. and Weissman, S.M. (1978) J. Biol. Chem. 253, 3643-3647.
75. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Lebowitz, P. and Weissman, S.M. (1978) J. Mol. Biol. 126, 813-846.
76. Cordell, B., Weiss, S.R., Varmus, H.E. and Bishop, J.M. (1978) Cell 15, 79-91.
77. Jacobs, J.W., Goodman, R.H., Chin, W.W., Dee, P.C., Habener, J.F., Bell, N.H. and Potts, J.T., Jr. (1981) Science 213, 457-459.
78. Glanville, N., Durnam, D.M. and Palmiter, R.D. (1981) Nature 292, 267-269.

79. Ohtsuka, E., Nishikawa, S., Fukumoto, R., Tanaka, S., Markham, A.F., Ikehara, M. and Sugiura, M. (1977) Eur. J. Biochem. 81, 285-291.
80. England, T.E. and Uhlenbeck, O.C. (1978) Biochemistry 17, 2069-2076.
81. Silberklang, M., Gillum, A.M. and RajBhandary, U.L. (1977) Nucleic Acids Res. 4, 4091-4108.
82. Ahlquist, P., Dasgupta, R., Shih, D.S., Zimmern, D. and Kaesberg, P. (1979) Nature 281, 277-282.
83. Kozak, M. (1980) Cell 19, 79-90.
84. Shih, D.S., Shih, C.T., Kew, O., Pallansch, M., Rueckert, R. and Kaesberg, P. (1978) Proc. Natl. Acad. Sci. USA 75, 5807-5811.
85. Pawson, T., Martin, G.S. and Smith, A.E. (1976) J. Virol. 19, 950-967.
86. Piatak, M., Ghosh, P.K., Reddy, V.B., Lebowitz, P. and Weissman, S.M. (1979) in Extrachromosomal DNA. ICN-UCLA Symposia on Molecular and Cellular Biology, Cummings, D.J., Borst, P., Dawid, I.B., Weissman, S.M. and Fox, C.F., Eds.,Vol. XV, pp. 199-215, Academic Press, New York.
87. Preston, C.M. and McGeoch, D.J. (1981) J. Virol. 38, 593-605.