Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 2662–2666, March 1995
Biochemistry

Adherence to the first-AUG rule when a second AUG codon follows closely upon the first

(initiation of translation/scanning model/mRNA structure/gene expression)

MARILYN KOZAK

Department of Biochemistry, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, NJ 08854

Communicated by Aaron J. Shatkin, Center for Advanced Biotechnology and Medicine, Piscataway, NJ, December 21, 1994 (received for review October 6, 1994)

ABSTRACT **The rule that eukaryotic ribosomes initiate translation exclusively at the 5′ proximal AUG codon is abrogated under rare conditions. One circumstance that has been suggested to allow dual initiation is close apposition of a second AUG codon. A possible mechanism might be that the scanning 40S ribosomal subunit flutters back and forth instead of stopping cleanly at the first AUG. This hypothesis seems to be ruled out by evidence presented herein that in certain mRNAs, the first of two close AUG codons is recognized uniquely. To achieve this, the 5′ proximal AUG has to be provided with the full consensus sequence; even small departures allow a second nearby AUG codon to be reached by leaky scanning. This context-dependent leaky scanning unexpectedly fails when the second AUG codon is moved some distance from the first. A likely explanation, based on analyzing the accessibility of a far-downstream AUG codon under conditions of initiation versus elongation, is that 80S elongating ribosomes advancing from the 5′ proximal start site can mask potential downstream start sites.**

Eukaryotic mRNAs generally adhere to the first-AUG rule; that is, in most cases the AUG codon nearest the 5′ end is the unique site of initiation of translation. This pattern (1) along with other evidence (2–6) led to formulation of the scanning model, which postulates that the 40S ribosomal subunit enters at the 5′ end of the mRNA and migrates linearly, stopping when it encounters the first AUG codon.

Two escape mechanisms account for most exceptions to the first-AUG rule. Reinitiation at a downstream AUG codon may be possible when the 5′ proximal AUG triplet is followed shortly by a terminator codon (7, 8). A second mechanism that allows access to downstream AUG codons is leaky scanning. The scanning model postulates that 40S ribosomal subunits stop at the first AUG if that codon occurs in a favorable context, which for vertebrates is gccGCCACC**AUG**G (4–6). But if the first AUG codon occurs in a suboptimal context—e.g., in the absence of the critical purine in position −3 or G in position +4—some 40S subunits will bypass the first AUG and initiate instead at a downstream site. Two independently initiated proteins may thus be produced from one mRNA by context-dependent leaky scanning (9–12). There are also a few cases where a too-short 5′ noncoding sequence—e.g., shorter than ≈20 nt—promotes leaky scanning (13–15).

RNA-6 of influenza virus B appears to be an exception to the foregoing exceptions. RNA-6 directs the synthesis of two proteins, NB and NA, initiated respectively at the first [AUG(NB)] and second [AUG(NA)] AUG codons (16). Because the context around the first AUG includes an A in position −3 and because the length of the 5′ noncoding sequence (46 nt) is adequate, this mRNA does not ostensibly meet the requirements for leaky scanning. Williams and Lamb (17) found that, whereas the natural 4-nt spacing between AUG(NB) and AUG(NA) allowed translation of both proteins, translation of NA was precluded when the inter-AUG spacing was increased to 46 nt. The apparent dependence of dual initiation on the proximity of the AUG codons led Williams and Lamb (17) to postulate that once the 40S ribosomal subunit reaches a given region of the mRNA, linear scanning might break down, and it might be a random choice as to which of two close AUG codons is used.

Here, I describe experiments designed to evaluate the possibility that, in addition to reinitiation and context-dependent leaky scanning, the close apposition of two AUG codons might provide a third escape mechanism—a third way around the constraint imposed by the scanning mechanism, which usually limits initiation to the 5′ proximal AUG codon. The question is interesting not only for the practical value of understanding how to construct bicistronic mRNAs but also for theoretical reasons. The question is whether the scanning 40S ribosomal subunit comes to a clean halt or whether the 40S subunit flutters back and forth over a small stretch of mRNA, such that either of two AUG codons within the stop–scan window can initiate translation.

MATERIALS AND METHODS

Plasmid Construction and Nomenclature. The plasmids used here were derived from pSP64 (Promega) into which a bacterial chloramphenicol acetyltransferase (CAT) gene sequence was inserted at the *Bam*HI site. The parental SP64-CAT construct (6) retains unique *Hin*dIII and *Bam*HI cleavage sites into which oligonucleotides can be inserted to introduce ATG (AUG) codons upstream from the start of the CAT coding sequence, henceforth designated AUG^{cat}. An N-terminally extended "preCAT" protein results from initiation at an upstream, in-frame AUG codon (AUG^{precat}). Constructs that carry an o subscript have a single, upstream, out-of-frame AUG codon (AUG^{out}) before the start of the CAT coding sequence. Control constructs that lack AUG^{out} carry a c subscript. For constructs that contain both AUG^{out} and AUG^{precat}, the spacing between the two upstream AUG codons is indicated in parentheses—e.g., K3(2). The sequence downstream from AUG^{out} was varied by introducing either the structure-prone oligonucleotide 8336 (characterized in ref. 18) or the unstructured sequence GAUCCAAAGUCAGCCAAAUCAA (oligonucleotide 8335) at the *Bam*HI site. In all cases where the downstream sequence is not specified, oligonucleotide 8336 was used.

The first members of the K series were numbered sequentially; these include K3, K4, K6, and K7 (see Fig. 1). Subsequent derivatives were modeled after K4, and therefore they were called K44, K45, K46, and so forth.

***In Vitro* Transcription and Translation.** Using *Ava* I-linearized plasmid DNA as the template, capped mRNAs were syn-

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

Abbreviation: CAT, chloramphenicol acetyltransferase.

thesized with SP6 RNA polymerase as described (19). Synthesis of CAT-related polypeptides was measured *in vitro* by using an mRNA-dependent rabbit reticulocyte translation system from GIBCO/BRL (19). The 30-μl reaction mixture, supplemented with 2.2 mM $Mg(CH_3COO)_2$, 90 mM $K(CH_3COO)$, and 45 mM KCl, contained 10 μl of reticulocyte lysate, 50 μCi of [^{35}S]methionine (>1000 Ci/mmol; 1 Ci = 37 GBq), and 0.2 μg of mRNA. Incubation was at 30°C for 60 min. The justification for these particular conditions is that they support a pattern of context-dependent initiation *in vitro* similar to what occurs *in vivo* (6) and they reproduce physiological differences in translational efficiency between α- and β-globin mRNAs (19). The mRNAs used in the present study produced, in addition to CAT and preCAT polypeptides, a 67-amino acid polypeptide derived from initiation at AUG^{out}. This small product was not quantified because it lacks internal methionine residues.

Primer Extension Analyses. A primer-extension inhibition ("toeprint") assay developed for studying ribosome–mRNA complexes in prokaryotes (20) was adapted for studying initiation complexes in eukaryotes. A gel-purified oligonucleotide primer (ATGTTCTTTACGATGCCATTGGGA, complementary to CAT gene codons 14–21) was labeled at the 5′ end by incubation with T4 polynucleotide kinase and [γ-^{32}P]ATP (3000 Ci/mmol). An equimolar mixture of mRNA and labeled primer (1 pmol each of mRNA and primer in 12 μl of water) was then heated to 65°C for 3 min, transferred to a dry ice/ethanol bath for 1 min, allowed to thaw in wet ice, and maintained at 0°C for 20 to 30 min.

mRNA–ribosome complexes were formed by adding the preannealed mRNA·primer mixture to a standard reticulocyte lysate (lacking [^{35}S]methionine) supplemented with 200 μM sparsomycin. Initiation complexes formed during a 2-min incubation at 30°C were purified by chromatography at 4°C on Sepharose CL-4B. The column elution buffer (50 mM Tris·HCl, pH 8.3/40 mM KCl/6 mM $MgCl_2$/5 mM dithiothreitol) was designed to preserve intact initiation complexes and to support the subsequent reverse transcriptase reaction. The column fractions that contained ^{32}P-labeled initiation complexes were identified by subjecting 5% of each fraction to analytical glycerol-gradient centrifugation under established conditions for resolving mRNA–ribosome complexes (21). Sepharose-column fractions that contained initiation complexes were supplemented with 500 μM dATP, dGTP, dCTP, and dTTP and with murine leukemia virus reverse transcriptase (GIBCO/BRL Superscript II; 4 units per μl) and incubated for 20 min at 37°C. The primer-extended products were extracted and analyzed by electrophoresis through 8% polyacrylamide sequencing gels. For reference, the gels also contained a molecular size ladder of reverse transcriptase products obtained by using a dideoxynucleotide-based RNA sequencing kit (Boehringer Mannheim).

RESULTS

Initiation from Two Close AUG Codons *in Vitro*. A preliminary experiment was undertaken to determine whether the ability to initiate at two AUG codons in an mRNA resembling influenza virus B RNA-6 could be reproduced *in vitro* and whether access to a second nearby AUG codon depends on the particular sequence of RNA-6. Our test for whether ribosomes can initiate at the second AUG codon involved placing the first AUG codon out of frame with respect to the CAT reporter sequence. Thus, if initiation were limited to the first AUG codon, no CAT-related proteins would be synthesized. And that is the result obtained with a control construct K_o, in which the second AUG codon (AUG^{cat}) resides 65 nt downstream from AUG^{out} (Fig. 1, lane 2). In contrast, K3(2) and K4(2) mRNAs, in which the second AUG resides only 2 nt downstream from AUG^{out}, were able to support synthesis of an elongated preCAT protein (Fig. 1, lanes 3 and 4). The ability to initiate from the second of two close AUG codons apparently did not depend strictly on the intervening sequence, since the sequence AC in K4(2) worked as well as AA in K3(2). The yields of preCAT protein from K3(2) and K4(2) were much lower than from a control mRNA, K_c, that lacks the upstream AUG^{out} barrier (Fig. 1, lane 1); but the main point is that some translation occurred with K3(2) and K4(2) mRNAs, unlike K_o, in which the far-downstream AUG^{cat} codon was silent.

To determine whether access to the second AUG codon in K3(2) and K4(2) depends strictly on the 2-nt spacing in those mRNAs, I next constructed K6(1) and K7(4), with inter-AUG spacings of 1 and 4 nt, respectively (Fig. 1). K_o6 is a matched control for K6(1). The 2 nt spacing in the original K3(2) and K4(2) constructs had the advantage of positioning AUG^{out} in a reading frame that terminates 130 nt downstream from AUG^{cat}, thus precluding synthesis of CAT by reinitiation. In contrast, the new spacings of 1 or 4 nt between the two AUG codons position AUG^{out} in a reading frame that terminates 10 nt upstream from AUG^{cat}, which enables the new constructs to produce some CAT protein by reinitiation (Fig. 1, lanes 5–7, lower band). Despite this complication, one could still ask whether another AUG codon ($AUG^{precat}_{\#2}$) introduced close to $AUG^{out}_{\#1}$ would be accessible to ribosomes. The answer was that AUG^{precat} was used detectably in both K6(1) and K7(4) (Fig. 1, lanes 6 and 7, upper band).

In K7(4) a 19-nt segment encompassing the two upstream AUG codons (GCCAAAAAUGAACAAUGCU) is identical to the initiation region of influenza virus B RNA-6. The experiment in Fig. 1 thus establishes that the bifunctionality of RNA-6 can be reproduced *in vitro* in the context of a chimeric test transcript. The ability to initiate at two close AUG codons is not unique to RNA-6, since the phenomenon also occurred with several synthetic leader sequences and with various inter-AUG spacings. Because the changes in spacing also changed the context and secondary structure

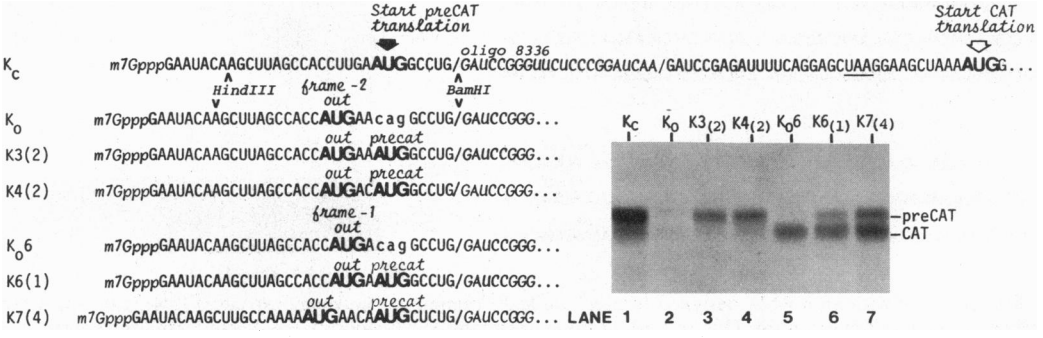

FIG. 1. A preliminary test of the ability to discriminate between two nearby AUG codons. The autoradiogram shows [^{35}S]methionine-labeled protein products obtained from a reticulocyte translation system. Downstream from the *Bam*HI site all mRNAs are identical to K_c, which in line 1 is given in full up to AUG^{cat}. In the control constructs K_o and K_o6, nucleotides substituted for upstream AUG codons are typed in lowercase.

around AUG^{precat}, variability in the yield of preCAT protein is understandable. Thus, it is best to score the constructs in Fig. 1 only as allowing (lanes 3, 4, 6, and 7) or disallowing (lane 2) initiation from the second AUG codon. In experiments described below, where the variables are manipulated one at a time, preCAT yields will be seen to follow a rational pattern.

In the aforementioned experiment, initiation occurred from a second nearby AUG codon even though AUG$^{out}_{\#1}$ was in a fairly strong context. Thus, K3(2), K4(2), and K6(1) had the optimal GCCACC motif in positions −6 to −1. Although K7(4) lacked most of this motif, it did have the critical A residue in position −3. At first glance, these favorable context elements might seem to eliminate leaky scanning as the explanation for preCAT synthesis, but closer inspection suggests that leaky scanning is not ruled out *a priori*. (*i*) The leader sequence on all these mRNAs was short, a feature that has been shown to allow some 40S ribosomes to slip past the first AUG codon (15). (*ii*) In none of these mRNAs did AUG^{out} carry the full consensus sequence, including a G residue in position +4. (*iii*) A third feature that has been shown to suppress leaky scanning is a stem–loop structure positioned 13 to 15 nt downstream from the intended initiator codon (18). Although the constructs in Fig. 1 had the structure-prone oligonucleotide 8336 downstream, the stem–loop structure was slightly too close to AUG^{out} to augment initiation at that site and thus to suppress leaky scanning.

In the next section I undertook to correct the structural deficiencies that might have allowed leaky scanning with the mRNAs used in Fig. 1 and thus to determine whether leaky scanning or a flutter mechanism underlies dual initiation from closely apposed AUG codons. The flutter mechanism (see the Introduction) postulates that the stop-scan step is inherently imprecise, and therefore initiation from two close AUG codons should persist despite optimization of the context at AUG$^{out}_{\#1}$.

Leaky Scanning Is What Enables Dual Initiation. A new set of mRNAs was patterned after K4(2) from the first experiment. As shown for K44(5) (Fig. 2, line 4), the 5′ noncoding sequence has been lengthened to 48 nt and the hairpin-forming oligonucleotide 8336 is now in a near-optimal position 13 nt downstream from AUG^{out}. In K44(5), both AUG$^{out}_{\#1}$ and AUG$^{precat}_{\#2}$ have the optimal G residue in position +4. Indeed, the extended sequence GNCACCAUGGCA is common to the first and second AUG codons in this mRNA.

Despite the nearly identical context and the proximity of AUG^{precat} to AUG^{out}, initiation was apparently restricted to AUG^{out} in K44(5): this mRNA produced virtually no preCAT protein in the standard reticulocyte translation system (Fig. 2*A*, lane 4). The result was confirmed by using a wheat germ translation system (6) (Fig. 2*B*, lane 4). The failure of K44(5) to support preCAT translation could not be blamed on AUG^{precat} having been moved too far from AUG^{out}, inasmuch as K44(2), with an inter-AUG spacing of only 2 nt, also produced no preCAT protein (Fig. 2 *A* and *B*, lane 5). A matched control showed that AUG^{precat} supported translation efficiently in the absence of the upstream AUG barrier (K$_c$44; Fig. 2 *A* and *B*, lane 3). In the absence of an upstream AUG codon, AUG^{cat} also functioned (K$_c$440; Fig. 2 *A* and *B*, lane 1), but AUG^{cat} was silenced by the upstream out-of-frame start site in K$_o$44 (Fig. 2 *A* and *B*, lane 2). The last result is not surprising, inasmuch as AUG^{out} had previously been shown to silence the distant AUG^{cat} site in K$_o$ (Fig. 1, lane 2). The novel result in Fig. 2 compared with Fig. 1 is that the nearby AUG^{precat} codon, which was able to support initiation in K4(2) (Fig. 1, lane 4), has been silenced in K44(2) and K44(5).

To test the interpretation that suppression of leaky scanning was responsible for the failure of K44(5) to initiate at AUG^{precat}, I introduced small sequence changes near AUG$^{out}_{\#1}$ that would be expected to restore leaky scanning. Initiation at AUG^{precat} was indeed slightly restored with K45(5), in which AUG^{out} is flanked by an A residue at position +4 instead of the optimal G residue at +4 (Fig. 2*C*, lane 6), and the yield of preCAT protein was even higher with K46(5) in which the upstream GCCACC motif has been shifted relative to AUG^{out}. Earlier studies had shown that GCCACC augments initiation only when the motif abuts the AUG codon (5).

For the experiment in Fig. 3*A*, K47(5), K46(5), and K44(5) were retested along with matched constructs (K$_o$47, K$_o$46, K$_o$44) in which the second potential start site is farther downstream from AUG^{out}. The experiment also includes a pair of constructs, K48(5) and K$_o$48, in which the consensus sequence around AUG^{out} extends all the way to position −9. It was not surprising that the strong context around AUG$^{out}_{\#1}$, which virtually precluded access to the nearby AUG$^{precat}_{\#2}$ codon in K44(5) and K48(5) (Fig. 3*A*, lanes 7 and 9), had a similar restrictive effect on the distant AUG$^{cat}_{\#2}$ site in K$_o$44 and K$_o$48 (Fig. 3*A*, lanes 8 and 10). A more surprising result was that the weaker context around AUG$^{out}_{\#1}$, which allowed some access to AUG$^{precat}_{\#2}$ in K47(5) and K46(5) (Fig. 3*A*, lanes 3 and 5), did not allow a comparable level of initiation from AUG$^{cat}_{\#2}$ in K$_o$47 and K$_o$46 (Fig. 3*A*, lanes 4 and 6). The next section takes up the question of why context-dependent leaky scanning appears to be attenuated as the second potential initiator codon is moved farther from the first.

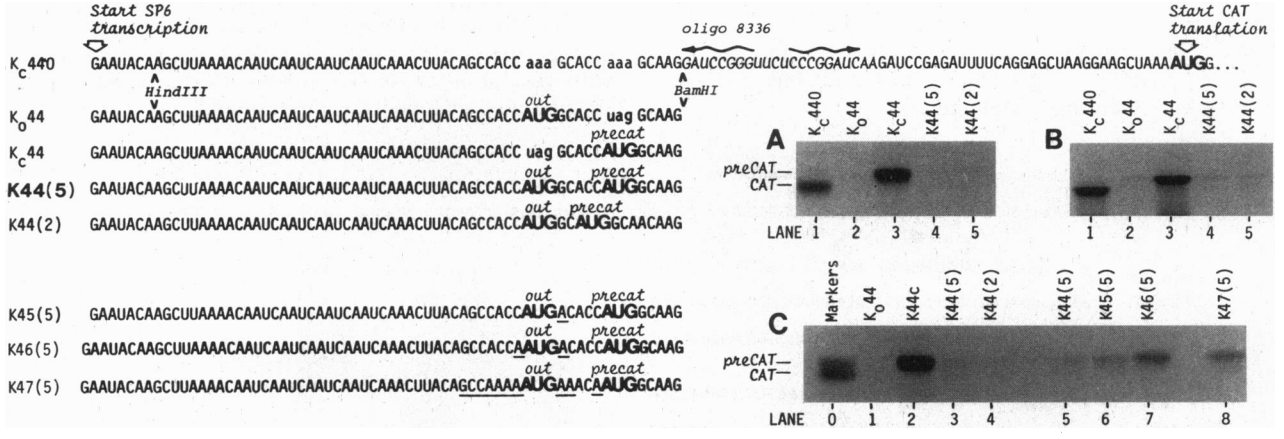

FIG. 2. Test for dual initiation with an mRNA designed to preclude leaky scanning. The key construct is K44(5) in which both AUG^{out} and the nearby AUG^{precat} occur in an optimal context. One or both of these upstream AUG codons were deleted from the control constructs K$_c$440, K$_o$44, and K$_c$44. In K45(5), K46(5), and K47(5), nucleotides that differ from K44(5) are underlined. Downstream from the *Bam*HI site, all constructs are identical to K$_c$440. Divergent arrows mark a stem–loop structure within oligonucleotide 8336. The autoradiograms show protein products from a reticulocyte (*A* and *C*) or wheat germ (*B*) translation system.

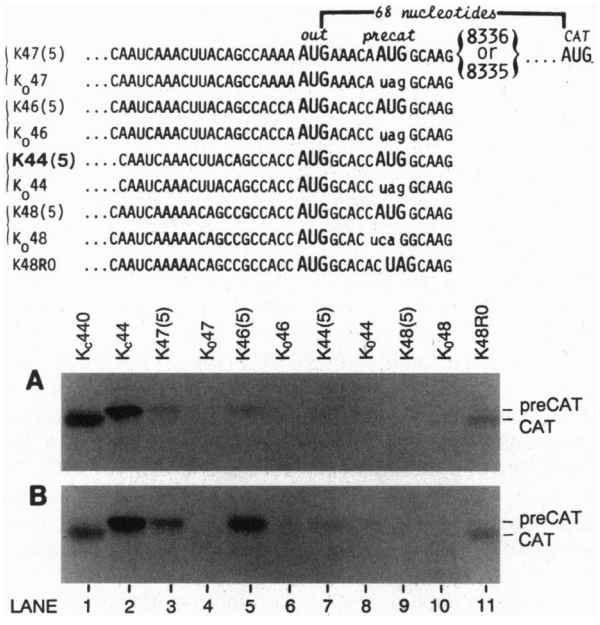

FIG. 3. Modulation of initiator codon selection by context and downstream secondary structure. For each mRNA in which the second AUG codon (AUG^{precat}) occurs close to AUG^{out}, there is a matched control (K_o47, K_o46, and so forth) in which the second AUG codon (AUG^{cat}) occurs far downstream from AUG^{out}. Each mRNA was tested with either a structured sequence (oligonucleotide 8336) (*A*) or an unstructured sequence (oligonucleotide 8335) (*B*) downstream from AUG^{out}. Upstream from the ellipsis (. . .) and downstream from oligonucleotide 8336 or 8335 the sequences were the same as in Fig. 2. Protein yields in *B* may be directly compared with those seen in *A*, inasmuch as aliquots of a common transcription reaction mixture were used to synthesize 21 of the mRNAs tested here [excluding K_c440(8335), which had to be remade], and aliquots of a master mix were used for all 22 translation assays.

To determine whether the downstream secondary structure, which was arbitrarily included in all mRNAs in Fig. 3*A*, actually influenced the results, the experiment was repeated in Fig. 3*B* with an unstructured sequence in place of the structure-prone oligonucleotide 8336. The results show that when AUG^{out} was in a less-than-perfect context, as in K46(5) and K47(5), access to AUG^{precat} increased when downstream secondary structure was eliminated.

As an aside, to see whether reinitiation can occur in this system, I inserted a UAG terminator codon 6 nt downstream from AUG^{out} (59 nt upstream from AUG^{cat}), producing a construct called K48R0, which indeed did support CAT protein synthesis (Fig. 3 *A* and *B*, lane 11).

Elongational Occlusion Suppresses Initiation from a Far-Downstream AUG Codon. If leaky scanning is the explanation for synthesis of preCAT protein from constructs described above in which $AUG^{out}_{\#1}$ is in a good but not perfect context, then a new puzzle is created: why does leaky scanning allow initiation from a nearby AUG codon (AUG^{precat}) but not from the more distant AUG^{cat} codon? The answer might be that a queue of 80S elongating ribosomes advancing from AUG^{out} can occlude potential downstream initiation sites. It would seem reasonable to expect such occlusion to become more severe as the distance between the first and second AUG codons, and thus the length of the ribosomal queue, increases.

If elongational masking is indeed the explanation for the discrepant result wherein the second AUG codon is accessible in K46(5) but not K_o46 (Fig. 3*B*, lanes 5 and 6), for example, the discrepancy should disappear when these mRNAs are tested under conditions of initiation. This can be accomplished by studying mRNA–ribosome complexes in the presence of an inhibitor of elongation, such as sparsomycin. To probe sparsomycin-blocked initiation complexes, I used an extension–inhibition assay in which a ^{32}P-labeled oligodeoxynucleotide was annealed to the mRNA downstream from all potential initiation sites and, after allowing ribosomes to engage the mRNA, reverse transcriptase was used to extend the primer up to the edge of the bound ribosome. Earlier ribosome protection experiments in which initiation complexes were trimmed with RNase established that the leading edge of a ribosome extends 12–15 nt 3′ of the AUG codon (21), and primer-extension products would be expected to terminate in a similar position.

The left side of Fig. 4 illustrates results obtained with control transcripts. When the mRNA–primer complex was incubated with reverse transcriptase in the absence of ribosomes, the ^{32}P-labeled primer was extended all the way to the 5′ end of the mRNA (lane 0). All other mRNAs used in Fig. 4 were allowed to engage ribosomes before the addition of reverse transcriptase. With K_c440 mRNA, in which there were no AUG codons upstream from AUG^{cat}, the primer-extension assay revealed the expected 80S initiation complexes at AUG^{cat} (Fig. 4, lanes 1 and 2). With K_o44 mRNA, which has an upstream, out-of-frame AUG codon in a near-optimal

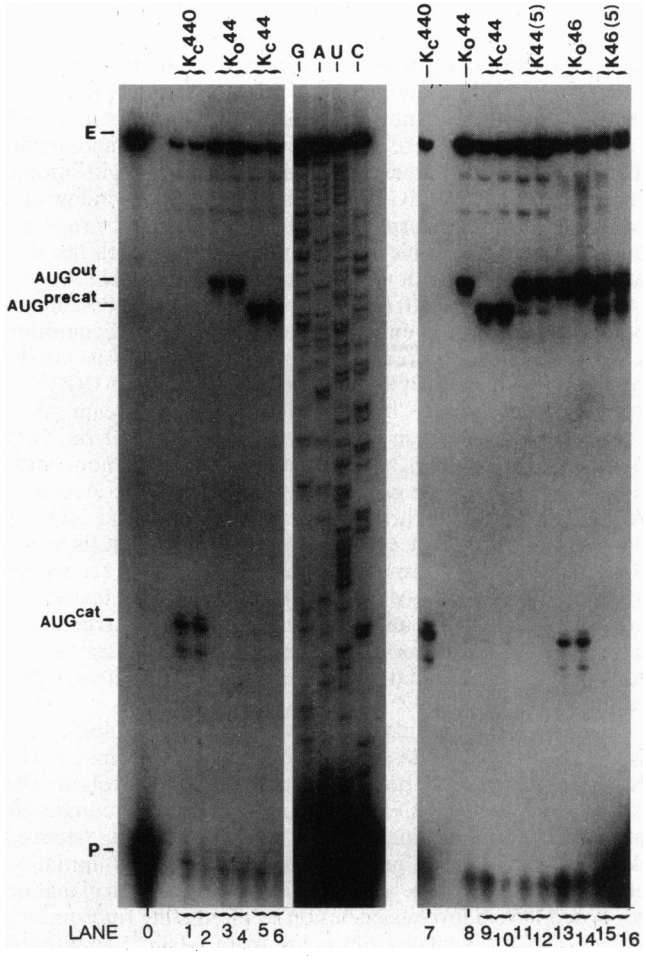

FIG. 4. Primer-extension analysis of initiation complexes. The mRNAs indicated at the top of the autoradiogram were incubated in a reticulocyte lysate supplemented with sparsomycin. Bracketed lanes show analyses on adjacent fractions from a sepharose column used to purify ribosome–mRNA complexes. A ^{32}P-labeled primer (P) annealed within the CAT coding domain was extended by reverse transcriptase as described in *Materials and Methods*. Extension stop-sites labeled AUG^{cat}, AUG^{precat}, and AUG^{out} occur 15 or 16 nt downstream from the stated AUG codon. A control reaction (lane 0) lacking ribosomes shows only the full-length extension product (E). Reference lanes labeled G, A, U, or C depict the minus-strand sequence of construct K460(8334). All other mRNAs used here had oligonucleotide 8335 downstream.

context, the primer-extension assay revealed 80S initiation complexes exclusively at AUG^{out} (Fig. 4, lanes 3 and 4). And with K_c44, in which AUG^{precat} in a near-optimal context is the first potential initiation site, complexes were detected exclusively at AUG^{precat} (Fig. 4, lanes 5 and 6). Thus, the toeprint assay can identify initiation complexes formed at three different positions in K-series mRNAs.

The right side of Fig. 4 depicts an experiment in which the aforementioned control transcripts were retested (Fig. 4, lanes 7 to 10) along with more interesting mRNAs. In the case of K44(5), although a second AUG codon occurs 5 nt downstream from the first, the near-optimal context around $AUG^{out}_{\#1}$ virtually precluded access to $AUG^{precat}_{\#2}$ (Fig. 4, lanes 11 and 12). In contrast, the suboptimal context around AUG^{out} in K46(5) allowed initiation from both AUG^{out} and the nearby AUG^{precat} (Fig. 4, lanes 15 and 16), as the analysis of ^{35}S-methionine labeled proteins had predicted. The key question was whether the toeprint assay would detect initiation from the far-downstream AUG^{cat} codon in K_o46; and the answer, as shown in Fig. 4, lanes 13 and 14, is that ribosomes indeed utilized both AUG^{out} and the downstream AUG^{cat} site in this mRNA.

DISCUSSION

One explanation I considered for the use of two initiator codons in mRNAs patterned after influenza virus B RNA-6 is that the stop-scan step might be inherently imprecise: if the 40S ribosomal subunit were to stutter (flutter back and forth) instead of stopping precisely, protein synthesis might initiate from either of two AUG codons that lie within a window of a certain size. But the results described above seem to rule this out, since it was possible to design mRNAs in which the first of two close AUG codons was recognized uniquely.

To achieve this result, the 5′ proximal AUG^{out} codon had to be provided with the entire consensus initiation recognition sequence. Thus, whereas access to AUG^{precat} was almost completely blocked by the upstream GCCACCAUGG sequence in K44(5), the barrier effect of the upstream AUG codon was weaker when position +4 was changed from G to A—e.g., K45(5) in Fig. 2*C*—especially when the 5′ noncoding sequence was short—e.g., K3(2) and K4(2) in Fig. 1. Access to $AUG^{precat}_{\#2}$ increased further as the context around $AUG^{out}_{\#1}$ deviated more from the consensus sequence—e.g., K46(5) and K47(5) in Fig. 2*C*—especially when downstream secondary structure was eliminated (Fig. 3*B*). Since all these features—context, leader length, and downstream secondary structure—have been shown to modulate leaky scanning, it seems reasonable to conclude that leaky scanning is what allows access to the nearby AUG^{precat} codon in these mRNAs.

In that case, the puzzle is not why K45(5), K46(5), and K47(5) produce a modest amount of preCAT protein, but why K_o45, K_o46, and K_o47 produce virtually no CAT protein. The structure of these mRNAs rules out occlusion by competing initiation complexes, inasmuch as the 68-nt spacing between AUG^{out} and AUG^{cat} provides enough room for initiation complexes to assemble at both sites. Thus, the explanation proposed herein invokes occlusion by elongating ribosomes: a queue of 80S ribosomes advancing from AUG^{out} apparently masks the downstream AUG codon that would otherwise be accessible by leaky scanning. The prediction that the far-downstream AUG^{cat} start site in K_o46 *should become accessible when elongation is blocked* was indeed confirmed by primer-extension analysis of ribosome–mRNA complexes formed in the presence of sparsomycin (Fig. 4).

Extrapolating from these results, leaky scanning is the most likely explanation for initiation from two AUG codons in influenza virus B RNA-6—leaky scanning due to a less-than-perfect context around AUG(NB) and to the absence of secondary structure in the A+U-rich sequence downstream. The proximity of the second initiation site to the first is important only because this arrangement minimizes elongational masking, which compensates for the first AUG codon being in a somewhat better context than is usually found in mRNAs that employ leaky scanning. In other mRNAs that use two initiation sites (9), the inter-AUG spacing varies from 4 nt to more than 100 nt. In the latter cases, the severity of elongational masking is probably offset by the extremely weak context at the upstream AUG codon. The pattern of codon usage might also modulate the severity of elongational masking (22).

A related issue raised earlier (22) and again here is that, with mRNAs that employ leaky scanning, the introduction of a terminator codon might augment downstream initiation by relieving elongational occlusion rather than by allowing reinitiation. The experiment with K48RO (Fig. 3) was undertaken, therefore, as an unequivocal test for reinitiation: because the optimal context around the first AUG codon precludes leaky scanning, as shown with K48(5), reinitiation is the only tenable explanation for use of the second AUG codon in K48RO. With some other mRNAs (23, 24) in which the 5′ proximal AUG codon was in a suboptimal context, the conclusion that a downstream cistron was translated by reinitiation might require rethinking.

The experiments described herein provide an especially rigorous test of the first-AUG rule. With appropriate constructs—those in which the first AUG codon was in an optimal context—initiation occurred almost exclusively from the first AUG codon. This was true even when the second AUG was in the same favorable context and even when the potential masking effect of elongating ribosomes was precluded by positioning the second AUG codon close to the first AUG or by using an initiation-only assay. An earlier *in vivo* study had provided strong support for the first-AUG rule by showing that all translation initiated at the first AUG codon when the start site from preproinsulin mRNA was tandemly repeated (25); but that study had a loophole, in that exclusive use of the 5′ proximal AUG codon might have resulted from a moderate preference for the first AUG augmented by elongational occlusion of downstream sites. The present study eliminates the loophole and thus strengthens the evidence for a linear scanning mechanism in which the direction of movement is strictly 5′ to 3′.

Research in my laboratory was supported by Grant GM33915 from the National Institutes of Health.

1. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148.
2. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241.
3. Kozak, M. (1992) *Crit. Rev. Biochem. Mol. Biol.* **27,** 385–402.
4. Kozak, M. (1986) *Cell* **44,** 283–292.
5. Kozak, M. (1987) *J. Mol. Biol.* **196,** 947–950.
6. Kozak, M. (1989) *Mol. Cell. Biol.* **9,** 5073–5080.
7. Kozak, M. (1987) *Mol. Cell. Biol.* **7,** 3438–3445.
8. Hinnebusch, A. G. (1990) *Trends Biochem. Sci.* **15,** 148–152.
9. Kozak, M. (1991) *J. Cell Biol.* **115,** 887–903.
10. Lin, F., MacDougald, O. A., Diehl, A. M. & Lane, M. D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9606–9610.
11. Muralidhar, S., Becerra, S. P. & Rose, J. A. (1994) *J. Virol.* **68,** 170–176.
12. Kozak, M. (1994) *Biochimie* **76,** 815–821.
13. Schutze, M.-P., Peterson, P. & Jackson, M. (1994) *EMBO J.* **13,** 1696–1705.
14. Slusher, L., Gillman, E., Martin, N. C. & Hopper, A. K. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9789–9793.
15. Kozak, M. (1991) *Gene Expression* **1,** 111–115.
16. Shaw, M., Choppin, P. W. & Lamb, R. A. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4879–4883.
17. Williams, M. A. & Lamb, R. A. (1989) *J. Virol.* **63,** 28–35.
18. Kozak, M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8301–8305.
19. Kozak, M. (1994) *J. Mol. Biol.* **235,** 95–110.
20. Hartz, D., McPheeters, D. S., Traut, R. & Gold, L. (1988) *Methods Enzymol.* **164,** 419–425.
21. Kozak, M. & Shatkin, A. J. (1977) *J. Biol. Chem.* **252,** 6895–6908.
22. Fajardo, J. E. & Shatkin, A. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 328–332.
23. Peabody, D. S., Subramani, S. & Berg, P. (1986) *Mol. Cell. Biol.* **6,** 2704–2711.
24. Thomas, K. R. & Capecchi, M. R. (1986) *Nature (London)* **324,** 34–38.
25. Kozak, M. (1983) *Cell* **34,** 971–978.