The EMBO Journal Vol.16 No.9 pp.2482–2492, 1997

Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6

Marilyn Kozak

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

A primer extension (toeprinting) assay was used to monitor selection by ribosomes of the first versus the second AUG codon as a function of introducing mutations on the 3′ side (positions +4, +5 and +6) of the first AUG codon. Six different flanking codons starting with G (GCG, GCU, GCC, GCA, GAU and GGA) strongly augmented selection of AUG#1 when compared with matched mRNAs that had A or C instead of G in position +4. Augmentation by G in position +4 failed only when it was combined with U in position +5, as in the sequence augGUA. In contrast with the usual enhancing effect of introducing G in position +4, most mutations in position +5 had no discernible effect, as shown with the series augANA (where N = C, A, G or U) and the series augCNA. AUG codon recognition was also unaffected by mutations in position +6, as shown by testing four mRNAs that had augCCN as the start site. Thus the primary sequence context that augments the recognition of AUG start codons does not appear generally to extend beyond G in position +4. When the toeprinting assay was used with mRNAs that initiate translation at CUG instead of AUG, cugGAU was not recognized better than cugGGU, contradicting the hypothesis that initiation at non-AUG codons might be favored by A instead of G in position +5.

Keywords: initiation codon context/mRNA structure/protein synthesis/scanning model/translation

Introduction

Eukaryotic ribosomes appear to select the start site for translation by a scanning mechanism. The working hypothesis is that the small, 40S ribosomal subunit, carrying Met-tRNA$_i^{met}$ and various initiation factors, engages the mRNA at the capped 5′ end and migrates linearly until it encounters the first AUG codon. At the AUG codon, which is recognized by base pairing with the anticodon in Met-tRNA$_i^{met}$, the 40S ribosomal subunit stops, the 60S subunit joins and the 80S ribosome is poised to start protein synthesis. Evidence for this scanning mechanism and for the corollary first-AUG rule is summarized elsewhere (Kozak, 1989a, 1992, 1995).

In higher eukaryotes, sequences flanking the AUG codon modulate its ability to halt the scanning 40S ribosomal subunit. One of the modulating elements is the GCCACC motif in positions –6 to –1, immediately preceding the AUG codon (Kozak, 1987). Mutations that weaken adherence to this consensus motif, especially mutations that substitute a pyrimidine for the A in position –3, cause some 40S ribosomal subunits to bypass the first AUG and to initiate instead at the next AUG downstream (Kozak, 1986a, 1989b; Lin *et al.*, 1993; Ossipow *et al.*, 1993). This context-dependent 'leaky scanning' has also been seen when the highly conserved G in position +4, immediately following the AUG codon, is mutated (Kozak, 1986a, 1989b). Deviations in one or both of these key positions, and the resulting leaky scanning, seem to account for the ability of certain mRNAs to produce two proteins by initiating translation from the first and second AUG codons (Kozak, 1986b, 1991).

Two recent studies have raised the possibility that context effects on initiation might extend into the coding domain beyond position +4. In one case, initiation at GUG appeared to be more efficient when the second codon was GAU instead of GUA (Boeck and Kolakofsky, 1994). A companion study by Grünert and Jackson (1994) reported similarly that initiation at an AUG or CUG start codon was favored by A in position +5 and U in position +6.

However, documenting the involvement of these or other nucleotides on the 3′ side of the initiator codon might be complicated by the fact that mutations introduced in these positions of the mRNA may change the amino acid sequence of the encoded polypeptide. This could be a problem because the identity of the amino acid adjacent to the N-terminal methionine can affect post-translational modifications which, in turn, can affect protein turnover. To circumvent possible complications from post-translational events, an assay that directly monitors ribosome–mRNA initiation complexes was used in the present study to reinvestigate the question of whether nucleotides in positions +4, +5 and +6 affect the recognition of initiator codons.

Correct definition of the context requirements for initiation is important for predicting translational start sites, which is an important aspect of interpreting cDNA sequences (Kozak, 1996).

Results

Preliminary test of mutations in positions +4, +5 and +6

The mRNAs used for these experiments have two start codons and two open reading frames (ORFs), as outlined in Figure 1. ORF1, which extends from AUG#1 to a UAA codon overlapping Leu45 in the chloramphenicol acetyltransferase (CAT) coding sequence, encodes a 70 amino acid polypeptide with a molecular mass of 8 kDa. This polypeptide is designated p8^{out} (meaning out-of-frame with respect to CAT) or simply p8. ORF2 initiates with AUG#2, which is in-frame with the downstream CAT coding sequence. ORF2 thus encodes a 240 amino

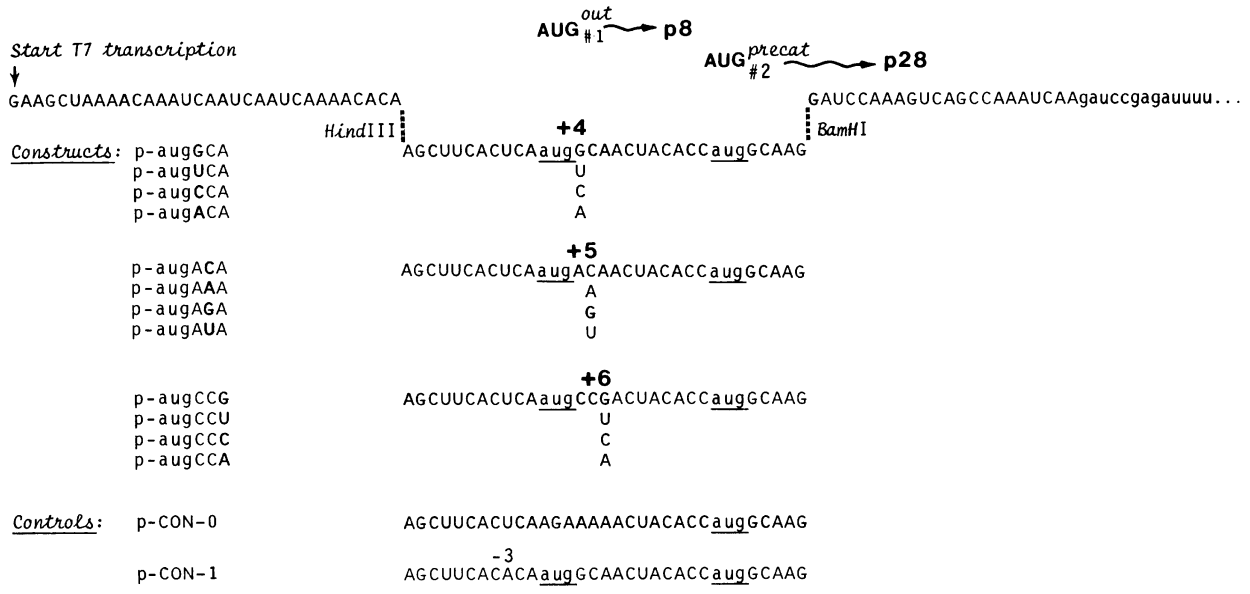

Fig. 1. Sequences of mRNAs used to study the effects of varying the nucleotide in positions +4, +5 and +6. The sequence in the top line is common to all mRNAs in this series. The 3′ end of this sequence indicated by an ellipsis (. . .) leads to the CAT coding domain (Kozak, 1989b). Not shown is the m7G cap at the 5′ end of all mRNAs. Mutations in the indicated positions of particular mRNAs were introduced around the first AUG codon, which initiates translation of an 8 kDa polypeptide (p8). In a different, overlapping reading frame, the second AUG codon initiates translation of a 28 kDa polypeptide (p28) which is an N-terminally extended version of CAT. Because of the suboptimal context preceding AUG#1 (notably the presence of U rather than A in position –3), some ribosomes would be expected to reach AUG#2 by leaky scanning. Thus, each of the 12 test constructs should direct translation of both p8 and p28. The improved context (A in position –3) in the control p-Con-1 should strongly shift translation in favor of p8. With the other control, p-Con-0, p28 should be the sole translation product because the upstream AUG codon is absent. Notice that mRNAs are named by stating the three bases following the first AUG codon.

acid polypeptide (the 219 amino acid CAT protein with a 21 amino acid N-terminal extension), with a molecular mass of 28 kDa. The product of ORF2 is designated $p28^{precat}$ or simply p28. Because the sequence preceding AUG#1 includes U in position –3, which is suboptimal, leaky scanning should allow these mRNAs to produce both polypeptides: p8 from AUG#1 and p28 from AUG#2. The control mRNA in Figure 2A (lane 9) illustrates how this leaky scanning can be modulated by changes in context. Because AUG#1 in the control has the optimal A in position –3, this mRNA produces a much higher yield of p8, and a much lower yield of p28, than any other mRNA in this series. This fits with previous studies of mutations involving sequences on the 5′ side of the AUG codon (Kozak, 1986a, 1989b).

The present study asks whether mutations on the 3′ side of AUG#1 can also modulate the selection of translational start sites. As shown in Figure 2A (lanes 2–8), the yield of p8 initiated from AUG#1 indeed varied at least 5-fold when point mutations were introduced in positions +4 or +5. However, the scanning mechanism predicts that if, for example, p-augAAA really supports initiation better than p-augAUA, as suggested by the 5-fold higher yield of p8 in lane 6 versus lane 8, then the yield of p28 should be proportionately lower in lane 6. That prediction is not met. Instead, the only mRNA in the test series that shows both elevated p8 synthesis and reduced p28 synthesis is p-augGCA (lane 2), the construct that has G instead of U, C or A in position +4.

With the other mRNAs tested in Figure 2A, a possible explanation for the variable yield of p8 without concomitant reduction of p28 is that mutations in positions +4 and +5, which change the subterminal amino acid, thereby alter the turnover of polypeptide p8. In this case, the amount of radiolabeled p8 that accumulates during the hour-long incubation would not reflect the efficiency of initiation at AUG#1 accurately. To circumvent this potential problem, the mRNAs used in Figure 2A were retested using a direct initiation assay.

To examine the effects of mutations in position +6, I chose a codon that specifies the same amino acid regardless of which base occurs in position +6. Thus the N-terminal sequence of the nascent polypeptide is Met–Pro when translation initiates at augCCG, augCCU, augCCC or augCCA. Among these four mRNAs there was no significant difference in the yield of p8 in a standard translation assay (Figure 2B, lanes 1–4). These mRNAs were also retested using the initiation assay described next.

Direct analysis of AUG codon recognition using mRNAs with mutations in positions +4, +5 and +6

By using a reticulocyte lysate supplemented with sparsomycin and cycloheximide to inhibit elongation (see Materials and methods), initiation complexes accumulate in which the ribosome is held at the AUG codon. The particular AUG start site can be identified by using a primer extension inhibition assay in which a ^{32}P-labeled deoxyoligonucleotide primer, annealed to the mRNA downstream from all potential initiator codons, is extended by reverse transcriptase up to the 3′ edge of the bound ribosome. Figure 3 outlines how the assay works in principle.

Two control reactions in Figure 4 illustrate how the assay works in practice. With the control mRNA p-Con-1 in which AUG#1 resides in a nearly optimal context

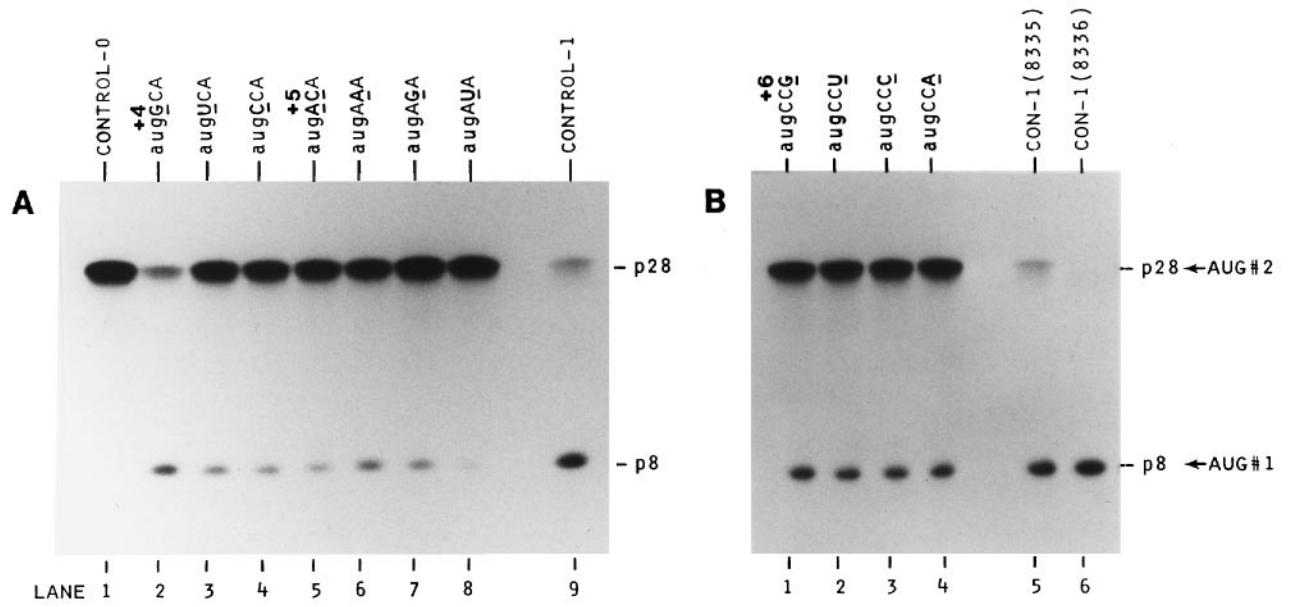

Fig. 2. Translation of mRNAs that vary in positions +4, +5 and +6 flanking the first AUG codon. The autoradiograms show [^{3}H]leucine-labeled proteins produced in a rabbit reticulocyte translation system using mRNAs that have point mutations in positions +4 and +5 (**A**) or position +6 (**B**). The mutations identified above each lane were introduced around AUG#1, which initiates translation of p8. p28 results from initiation at the invariant AUG#2. Figure 1 gives the 5′ end sequences of these mRNAs in full. A control mRNA that lacks AUG#1 (p-Con-0 in lane 1 of A) produced only p28. For the control mRNA p-Con-1 (lane 9 in A; lane 5 in B), the context around AUG#1 was improved by changing position –3 from U to A, thus enhancing synthesis of p8 and greatly reducing synthesis of p28. In (B), the slight residual translation of p28 evident in lane 5 was abolished in lane 6 by introducing downstream the structure-prone sequence 8336, which is thought to slow scanning and thus augment recognition of AUG#1 (Kozak, 1990a). This is shown only as an illustration, inasmuch as all the other mRNAs used in this figure contained the unstructured sequence 8335 at the *Bam*HI site. The conditions used for translation (protein accumulation assay) and subsequent fractionation by polyacrylamide gel electrophoresis are described in Materials and methods.

(ACAaugG, see Figure 1), one prominent primer extension product is evident in Figure 4A (lanes 11 and 12) and the size of this product indicates that it derives from ribosomes bound at AUG#1, the start codon for p8. This primer extension product was absent when p-Con-0 mRNA was used for ribosome binding (Figure 4A, lanes 1 and 2), consistent with the fact that p-Con-0 lacks the upstream AUG codon (see Figure 1). With p-Con-0 the toeprinting assay maps ribosomes instead at the p28 start codon. (The p28 start site is labeled AUG#2 in Figure 4 because it is the second start codon in all mRNAs except p-Con-0.)

The rest of Figure 4A tests the effects of introducing mutations in position +4 flanking AUG#1. Because all four test transcripts (the first four mRNAs in Figure 1) have U instead of the optimal A in position –3, some ribosomes are able to reach AUG#2 by leaky scanning. The question is whether the ratio of initiation at AUG#1 versus AUG#2 differs among these four mRNAs which are identical except for position +4. Quantitation of the data from Figure 4A (Table I, measurement 1, entries 1–4) indeed shows a 2.6-fold shift in favor of AUG#1 when that codon is followed by G in position +4 (henceforth written G^{+4}). There was no real hierarchy among the other three nucleotides in position +4.

The toeprinting experiment was repeated in Figure 4B using four mRNAs that were identical except for position +5. Quantitation of these results showed no significant shift in the AUG#1/AUG#2 ratio (Table I, measurement 1, entries 5–8). Thus there was no evidence that the nucleotide in position +5 affects the selection of translational start sites. Nor was there any significant effect when mutations in position +6 were tested (Figure 4C; Table I, measurement 1, entries 9–12).

Because leaky scanning in cell-free translation systems was shown previously to be sensitive to the concentration of Mg^{2+} (Kozak, 1989b, 1990b), I repeated the test of mutations in positions +4, +5 and +6 at three different Mg^{2+} concentrations. The results of these toeprinting assays are shown in Figure 5 and the quantitation is given in Table I (measurements 2, 3 and 4). As reported previously, when a given mRNA is tested at different Mg^{2+} concentrations, the tendency to scan past AUG#1 and initiate instead at AUG#2 increases as the Mg^{2+} concentration is decreased. This can be seen in Figure 5A, for example, by comparing the translation of p-augGCA in lanes 1, 5 and 9. The point is sustained by comparing any other mRNA in Figure 5A at low, medium and high concentrations of Mg^{2+} (e.g. p-augUCA in lanes 2, 6 and 10; p-augCCA in lanes 3, 7 and 11; or p-augACA in lanes 4, 8 and 12).

The real purpose of this experiment was to compare, at a given concentration of Mg^{2+}, the translation of four mRNAs that differ in a single position downstream of AUG#1. In Figure 5A (lanes 1–4) this four-way comparison reveals that G is the only nucleotide in position +4 that augments recognition of AUG#1 at low Mg^{2+}, and that conclusion holds at moderate (lanes 5–8) and high (lanes 9–12) Mg^{2+} concentrations. In Figure 5B, a similar experiment using four mRNAs that differ only in position +5 shows that, at any given Mg^{2+} concentration, AUG#1 is recognized with equal efficiency irrespective of nucleotide changes in position +5. Figure 5C shows that, at any given Mg^{2+} concentration, recognition of AUG#1 is indifferent also to the nucleotide in position +6. The conclusion from this analysis is that G^{+4} appears

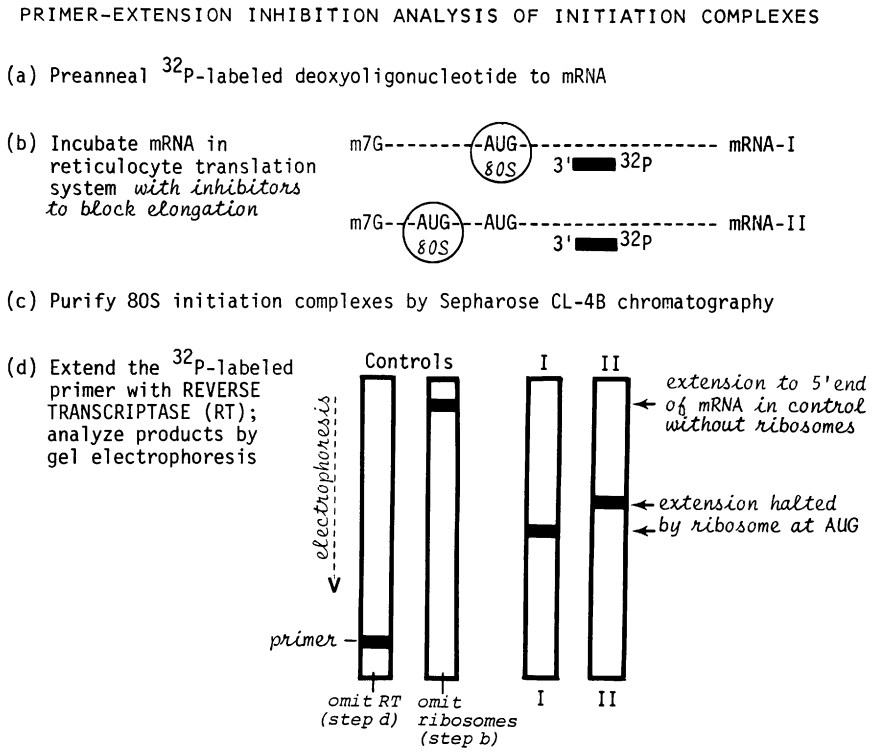

Fig. 3. Schematic representation of a primer extension assay for mapping the position of ribosomes on mRNA. The unextended ^{32}P-labeled primer, represented by the wide black line in step (b), is shown near the bottom of the polyacrylamide gel in step (d). Extension of the primer with reverse transcriptase in the absence of bound ribosomes proceeds to the 5′ end of the mRNA. If ribosomes are allowed to bind to the mRNA before the addition of reverse transcriptase, primer extension halts prematurely; the exact size of the extension product(s) reveals which AUG codon(s) were selected, taking into account that the leading edge of an 80S ribosome extends ~15 nucleotides 3′ of the AUG codon (Kozak and Shatkin, 1977). The basic design of this 'toeprinting' assay was developed by Hartz *et al.* (1988) for studies with prokaryotic ribosomes. The primer used in the present studies was 23 nucleotides long, the full-length extension product was 184 nucleotides and the extension inhibition products obtained when a ribosome was bound at AUG#1 or AUG#2 were 123 and 109 nucleotides, respectively.

to be the only nucleotide on the 3′ side of the AUG codon that augments initiation.

Distinguishing between particular codon effects and generalized context effects

Although G^{+4} augments AUG codon recognition under a variety of reaction conditions, as shown above, in all those studies the G in position +4 was part of the codon GCA. To determine if the augmentation is attributable specifically to G^{+4} or if it is the flanking codon GCA that happens to favor initiation, I tested mRNAs that had six different GNN codons adjacent to AUG#1. In the toeprint analyses shown in Figure 6A, each mRNA was compared with a matched construct that had C or A instead of G in position +4. Quantitation of the results (Table II) reveals that AUG#1 was indeed recognized ~3-fold better in five out of six cases where G^{+4} was the flanking nucleotide. Since five different flanking codons starting with G (GCG, GCU, GCC, GCA and GAU) strongly augmented the recognition of AUG#1, it seems reasonable to attribute the enhancement to the G residue in position +4 rather than to a particular flanking codon.

In Figure 6A, augGUA was the only mRNA in which G^{+4} unexpectedly failed to enhance initiation. To determine whether it is specifically the flanking codon GUA that disfavors initiation or whether the 3′ sequence GU somehow undermines recognition of the preceding AUG codon, I tested initiation at augGUG, augGUU and augGUC along with augGUA. Figure 6B (lanes 2–5) shows equally poor recognition of AUG#1 with all four constructs in this series. Thus the usual stimulatory effect of G^{+4}, seen in Figure 6B with the control transcripts augGAU and augGGA (lanes 1 and 6), fails for some reason when G^{+4} is followed by U^{+5}.

Although U in position +5 prevents the usual stimulatory effect of G^{+4}, U in position +5 is not generally deleterious. Thus, augAUA was not recognized less efficiently than augACA, aug AAA or augAGA in Figure 5B. The point is confirmed in Figure 6C where augCUA (lane 4) was recognized as efficiently as augCCA, augCAA or augCGA (lanes 1–3).

Effects of mutations flanking a CUG start codon

In view of some earlier reports about effects of downstream mutations (Boeck and Kolakofsky, 1994; Grünert and Jackson, 1994), it seemed useful to retest some of the foregoing conclusions with mRNAs that initiate translation at a non-AUG codon. In the mRNAs depicted in Figure 7A, CUG replaces AUG#1 as the start codon for p8. When these mRNAs were used as templates in a standard translation assay, some [^{3}H]leucine-labeled p8 was produced (Figure 7B, lanes 2–5), albeit less than with AUG as the start site for p8 (Figure 7B, lanes 1 and 6). That AUG as the p8 start codon is much stronger than CUG is also evident from the greater inhibition of p28 synthesis in lanes 1 and 6 compared with lanes 2–5 in Figure 7B. The complete absence of p8 when the CUG codon was

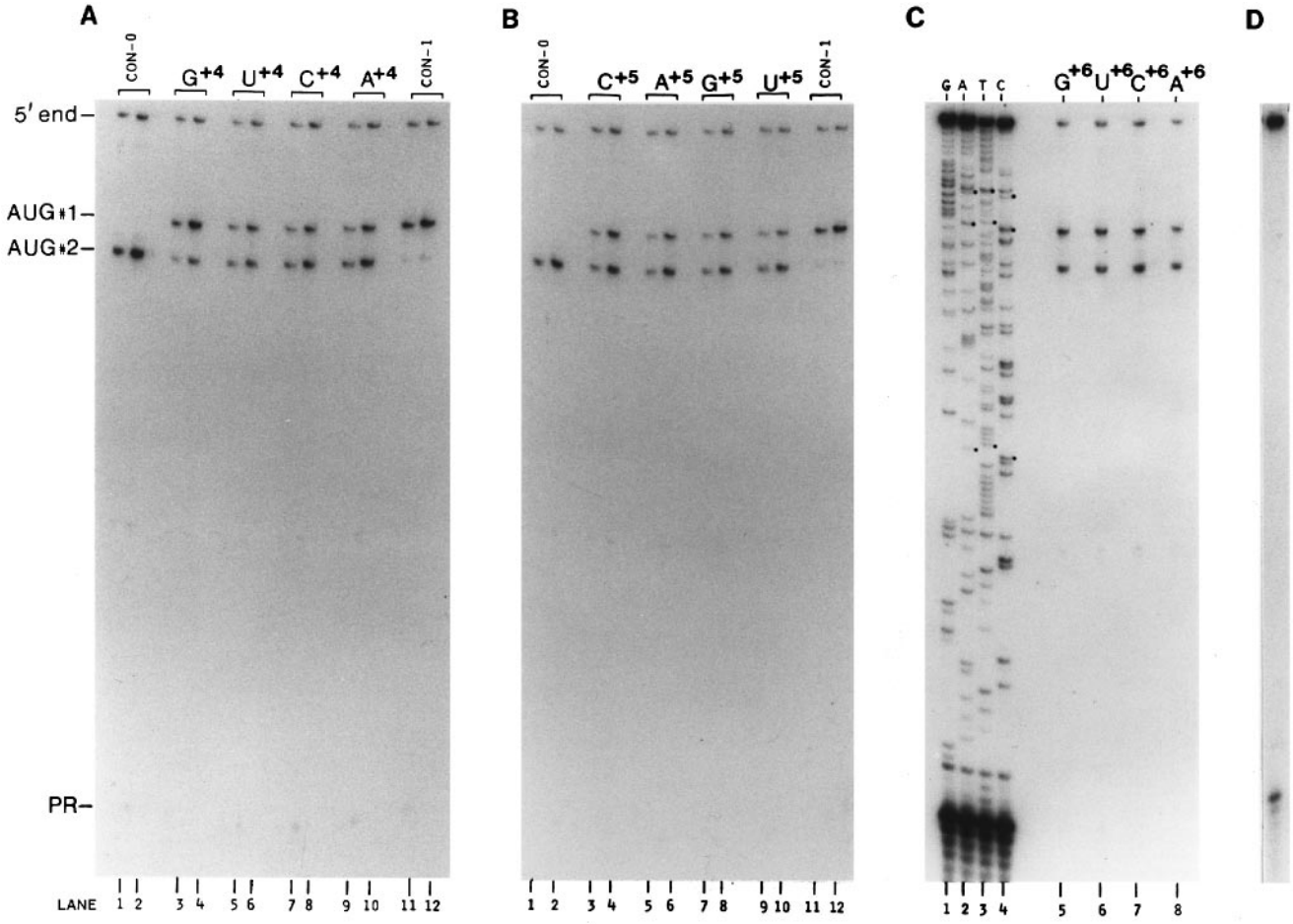

Fig. 4. Primer extension analysis of ribosome–mRNA complexes. Initiation at the first and second AUG codons was monitored as a function of introducing mutations around AUG#1. The assay is explained in Figure 3. The primer (PR) and primer extension products are labeled along the left margin. (**A** and **B**) The mRNAs used in lanes 3–10 varied only in position +4 or position +5, as indicated at the top of each panel. The sequences of these mRNAs as well as the two control transcripts (lanes 1, 2, 11 and 12) are depicted in full in Figure 1. Adjacent bracketed lanes show that, with a given mRNA, the ratio of initiation at AUG#1 versus AUG#2 was the same when low (lanes 1, 3, 5, 7, 9 and 11) and 3-fold higher (lanes 2, 4, 6, 8, 10 and 12) levels of initiation complexes were analyzed. Because of small variations in the amount of radioactivity applied to each lane, the important comparison is not the intensity of the AUG#1 band from lane to lane, but the ratio of AUG#1 to AUG#2 in each lane. These ratios are given in Table I. (**C**) Toeprint analyses with mRNAs that differed in position +6, as indicated above lanes 5–8. Lanes 1–4 display the complementary strand sequence of p-augCCA mRNA. A series of black dots within these sequencing lanes highlight the C–A–T bands that correspond to the first, second and (silent) third AUG codons. When the electrophoresis was run for longer, the foreshortened primer extension products caused by bound ribosomes could be mapped, by reference to the sequencing lanes, 15–16 nucleotides downstream from the first and second AUG codons. In the absence of ribosomes, the primer was extended all the way to the 5′ end of the mRNA, as shown in a control reaction (**D**).

mutated (Figure 7C, lane 1) confirms that CUG is the source of p8 in lanes 2–5.

Because earlier studies showed that appropriately positioned downstream secondary structure aids the recognition of weak initiator codons (Kozak, 1990a), I tested the CUG mRNAs with both an unstructured sequence (oligonucleotide 8335) and a structure-forming sequence (oligonucleotide 8336) downstream. The inclusion of oligonucleotide 8336 significantly elevated initiation from the CUG codon in Figure 7B (compare the yield of p8 in lane 4 versus lane 2, or in lane 5 versus lane 3); therefore, this downstream sequence was retained in the mRNAs used in Figure 7C and D.

In Figure 7C, I examined the effects of mutations in position +5 flanking the CUG start codon for p8. I specifically tested A versus G in this position because Grünert and Jackson (1994) reported the biggest effect when A^{+5}, which they considered optimal, was changed to G^{+5}. Although I too observed a higher yield of p8 with the mRNA that initiates at cugG<u>A</u>U instead of cugG<u>G</u>U (lanes 4 and 5 in Figure 7B and C), closer inspection of the results argues against concluding that A^{+5} enhances initiation at CUG. Notice, for example, that the yield of p8 was not augmented by A in position +5 when the flanking codons were AGU versus AAU (Figure 7C, lanes 2 and 3) instead of GGU and GAU.

Because differences in protein stability might distort the results of the protein accumulation assays in Figure 7B and C (for example, N-terminal acetylation might stabilize the form of p8 initiated from cugGAU), the critical test was whether mutations in position +5 would affect recognition of the CUG codon when initiation was monitored directly, using the primer extension assay. As shown in Figure 7D, although G in position +4 augmented recognition of the upstream CUG codon (compare cug<u>A</u>GU with cug<u>G</u>GU, lanes 4 and 8; or compare

Table I. Effects of mutations in positions +4, +5 and +6 as monitored by primer extension analysis of ribosome–mRNA initiation complexes

Entry No.	Sequence flanking AUG#1[a]		Ratio of initiation at AUG#1 versus AUG#2[b]							
			Measurement #1 (2.0 mM Mg^{2+})		Measurement #2 (1.4 mM Mg^{2+})		Measurement #3 (2.0 mM Mg^{2+})		Measurement #4 (2.7 mM Mg^{2+})	
			Actual	Normalized	Actual	Normalized	Actual	Normalized	Actual	Normalized
1	Position +4 is	G	2.6	2.6	0.46	2.4	2.2	2.2	6.3	3.9
2		U	1.0	1.0	0.19	1.0	1.0	1.0	1.6	1.0
3		C	0.8	0.8	0.16	0.8	0.8	0.8	1.6	1.0
4		A	0.9	0.9	0.19	1.0	0.8	0.8	1.7	1.1
5	Position +5 is	C	0.8	1.1	0.13	1.0	0.6	1.2	1.4	1.1
6		A	0.7	1.0	0.14	1.1	0.5	1.0	1.5	1.2
7		G	0.7	1.0	0.15	1.2	0.5	1.0	1.5	1.2
8		U	0.7	1.0	0.13	1.0	0.5	1.0	1.3	1.0
9	Position +6 is	G	1.1	1.1	0.23	1.0	1.2	1.0	2.2	0.9
10		U	1.0	1.0	0.24	1.0	1.2	1.0	2.4	1.0
11		C	0.9	0.9	0.16	0.7	0.9	0.8	2.0	0.8
12		A	0.8	0.8	0.19	0.8	1.0	0.8	1.9	0.8

[a]The codon following AUG#1 was NCA for the +4 series, ANA for the +5 series and CCN for the +6 series. The mRNA sequences are given in full in Figure 1.

[b]Because the sequence flanking AUG#2 was constant, an increase in the AUG#1/AUG#2 ratio indicates improved recognition of AUG#1. The data were obtained by densitometric scanning of the autoradiograms in Figure 4 (for measurement 1) and Figure 5 (for measurements 2–4). Because the mRNAs were tested four at a time (i.e. in Figure 5, four mRNAs were tested at three Mg^{2+} concentrations in each of three experiments), the strictly valid comparison is among mRNAs 1–4 or among mRNAs 5–8 but not between mRNAs 1–4 and mRNAs 5–8. However, by using the entries in columns 4, 6, 8 and 10, wherein the values have been normalized to the mRNA in each series that has U in the test position, it is possible to compare mRNAs between experiments.

cugAAU with cugGAU, lanes 6 and 10), there was no convincing difference between matched mRNAs that had G versus A in position +5 (compare cugAGU with cugAAU in lanes 4 and 6; or compare cugGGU with cugGAU in lanes 8 and 10).

Notice that the extended context in these CUG-containing mRNAs (GACAUAcugRRU) is the same as that used by Grünert and Jackson (1994).

Discussion

The optimal context for initiation does not extend beyond G in position +4

These experiments refute the suggestion that the recognition of initiator codons is strongly favored by A in position +5 and U in position +6, as proposed by Boeck and Kolakofsky (1994) and Grünert and Jackson (1994). By using an assay that directly monitors the initiation step of translation, I found no effect on recognition of the first AUG codon when position +5 was varied in the series ACA, AAA, AGA, AUA (Table I, entries 5–8) or the series CCA, CAA, CGA, CUA (Figure 6C). The efficiency of initiation at AUG#1 was also affected very little when position +6 was varied in the series CCG, CCU, CCC, CCA (Table I, entries 9–12) or the series GCG, GCU, GCC, GCA (Table II, entries 1, 3, 5 and 7). Because I did not test all possible combinations, these experiments do not rule out the possibility that, as part of some particular sequence, A^{+5} and U^{+6} might be preferable to some alternative sequence (see below); but the experiments do preclude generalizing the optimal context for initiation to include A^{+5} and U^{+6}.

The experiments herein do suggest, on the other hand, that the positive effect of G^{+4} is generalizable. As illustrated in Figure 6, recognition of AUG#1 improved in response to six different codons that introduced G in position +4: GCG, GCU, GCC, GCA, GAU and GGA. A strong contribution of G^{+4} was also seen in the experiment in which a CUG codon was used in place of AUG#1 (Figure 7D). Speculation that the frequent

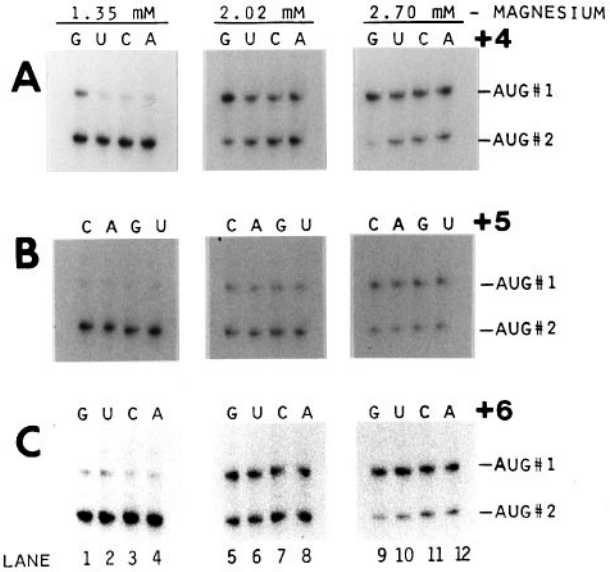

Fig. 5. Selection of AUG start sites as a function of magnesium concentration and sequence variation following AUG#1. (**A**) Twelve toeprinting reactions using mRNAs that differed from one another only in position +4 following AUG#1. The identity of the base in position +4 is marked above each lane. The primer extension reactions were carried out after 4 min incubation in a reticulocyte translation system in which the magnesium concentration had been adjusted to 1.35 (lanes 1–4), 2.02 (lanes 5–8) or 2.70 mM (lanes 9–12). The experiment was repeated using mRNAs that differed from one another only in position +5 (**B**) or position +6 (**C**). The mRNA sequences are given in full in Figure 1. Autoradiograms (similar to Figure 4) have been cropped. Quantitation of these results is given in Table I.

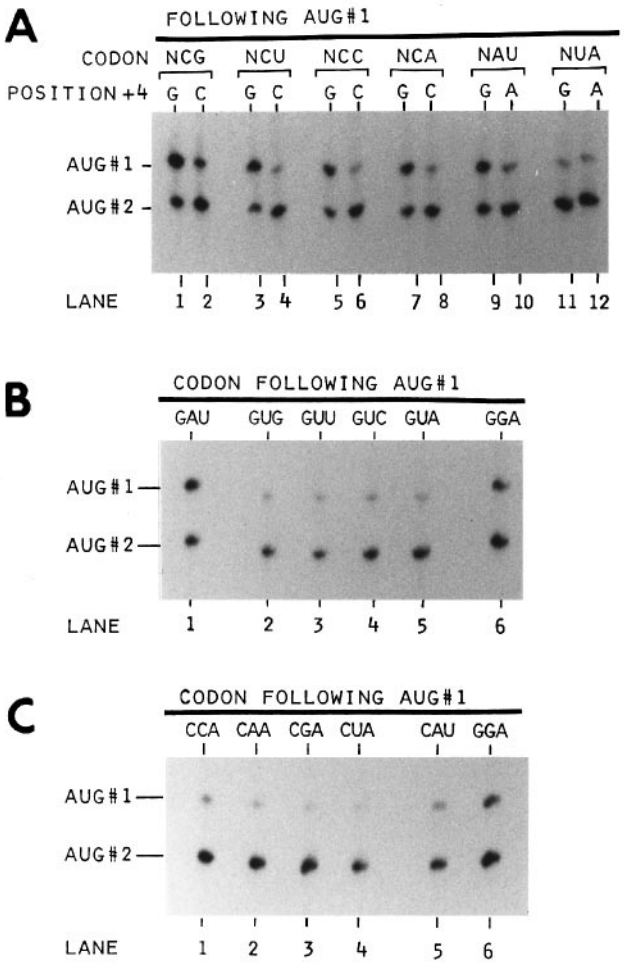

Fig. 6. Additional testing of various codons following AUG#1. Primer extension assays were carried out using a reticulocyte translation system with the Mg^{2+} concentration adjusted to 1.7 mM. The codon flanking AUG#1 is identified above each lane of the gel. Except for this first codon variation, the mRNA sequences were as given in Figure 1. (**A**) The positive effect of G^{+4} is seen with a variety of codons flanking AUG#1. The first eight lanes, compared two at a time, show the shift in initiation when AUG#1 is followed by G versus C in position +4. Thus the flanking codons are GCG versus CCG in lanes 1 and 2, GCU versus CCU in lanes 3 and 4, GCC versus CCC in lanes 5 and 6 and GCA versus CCA in lanes 7 and 8. The last four lanes show the shift when AUG#1 is followed by G versus A in position +4. The flanking codons are GAU, lane 9; AAU, lane 10; GUA, lane 11; and AUA, lane 12. Quantitation of these results is given in Table II. (**B**) The stimulatory effect of G^{+4}, evident when AUG#1 is followed by GAU (lane 1) or GGA (lane 6), fails when the flanking codon is GUN (lanes 2–5). (**C**) Variations in position +5 do not significantly affect recognition of AUG#1, as shown with CNA as the flanking codon (lanes 1–4). The mRNAs used in lanes 5 and 6 are controls. Warming during electrophoresis slightly retards the mobility of samples near the edges of the gel.

occurrence of G in position +4 might simply reflect selection for Ala, Gly and Val as the penultimate amino acids (Flinta *et al.*, 1986), rather than a role for G^{+4} in initiation, no longer seems valid. The experiments herein, using an assay that directly monitors ribosome–mRNA initiation complexes, show that recognition of AUG start codons is stimulated strongly by G^{+4}.

augGU is not a favorable context for initiation

The one interesting situation in which G in position +4 failed to augment recognition of AUG#1 involves the sequence augGUN. As shown in Figure 6A, for example, augGUA (lane 11) was recognized with only the same low efficiency as augAUA (lane 12). At first glance, the data in Figure 6A (e.g. lane 9 versus lane 11) seem to confirm an earlier report that (gug)GAU is a much stronger initiation site than (gug)GUA (Boeck and Kolakofsky, 1994). That observation contributed to the idea that A^{+5} and U^{+6} might be part of the optimal context for initiation. However, the more extensive set of data in the present study shows it is not that A^{+5} and U^{+6} augment initiation, but that the usual stimulatory effect of G^{+4} fails in the case of augGUA. Table II shows, for example, that augAAU (entry 10) is not recognized significantly better than augAUA (entry 12). Instead, Table II shows unexpectedly low recognition of AUG#1 when the flanking codon is GUA (entry 11) compared with every other mRNA that has G in position +4 (entries 1, 3, 5, 7 and 9).

This unexpected deficiency is not limited to the sequence augGUA. Figure 6B shows that the usual stimulatory effect of G^{+4} fails with every codon that begins with GU. The simplest interpretation is that the sequence GU in positions +4/+5 somehow distorts the mRNA and thus impairs AUG codon recognition by the scanning 40S ribosomal subunit.

Because the deleterious effect seems to be attributable to a particular flanking sequence (augGU) rather than to a particular flanking codon, it is not likely that the defect occurs after assembly of the 80S ribosome when a tRNA tries to enter the A site. It seems unlikely, for example, that augGUA is a poor initiation site because the complementary Val-tRNA is scarce (Zhang *et al.*, 1991) or because Val-tRNA is structurally incompatible with Met-$tRNA_i^{met}$ when both tRNAs line up on the 80S ribosome (Irwin *et al.*, 1995). Those explanations might be tenable if the defect were limited to augGUA; but augGUG, augGUU and augGUC were equally poor start sites. Experiments described herein specifically contradict the idea that scarcity of the elongator tRNA required to form the first peptide bond at AUG#1 might shift initiation to a downstream site. That hypothesis appears to be ruled

Table II. The positive effect of G in position +4 occurs with a variety of codons flanking AUG#1

Entry No.	Sequence at AUG#1	Ratio of initiation at AUG#1/AUG#2[a]	
		Actual[b]	Normalized[c]
1	augGCG	2.00	3.3
2	augCCG	0.61	1.0
3	augGCU	1.70	4.2
4	augCCU	0.41	1.0
5	augGCC	1.60	3.4
6	augCCC	0.47	1.0
7	augGCA	1.40	3.0
8	augCCA	0.46	1.0
9	augGAU	1.20	2.4
10	augAAU	0.50	1.0
11	augGUA	0.35	0.9
12	augAUA	0.40	1.0

[a]Because the sequence flanking AUG#2 was constant, an increase in this ratio indicates improved recognition of AUG#1.
[b]The autoradiogram in Figure 6A was used for these measurements.
[c]To normalize the results, the AUG#1/AUG#2 ratio for each mRNA that has G in position +4 is expressed relative to the matched C^{+4} or A^{+4} mRNA, which is set at 1.0.

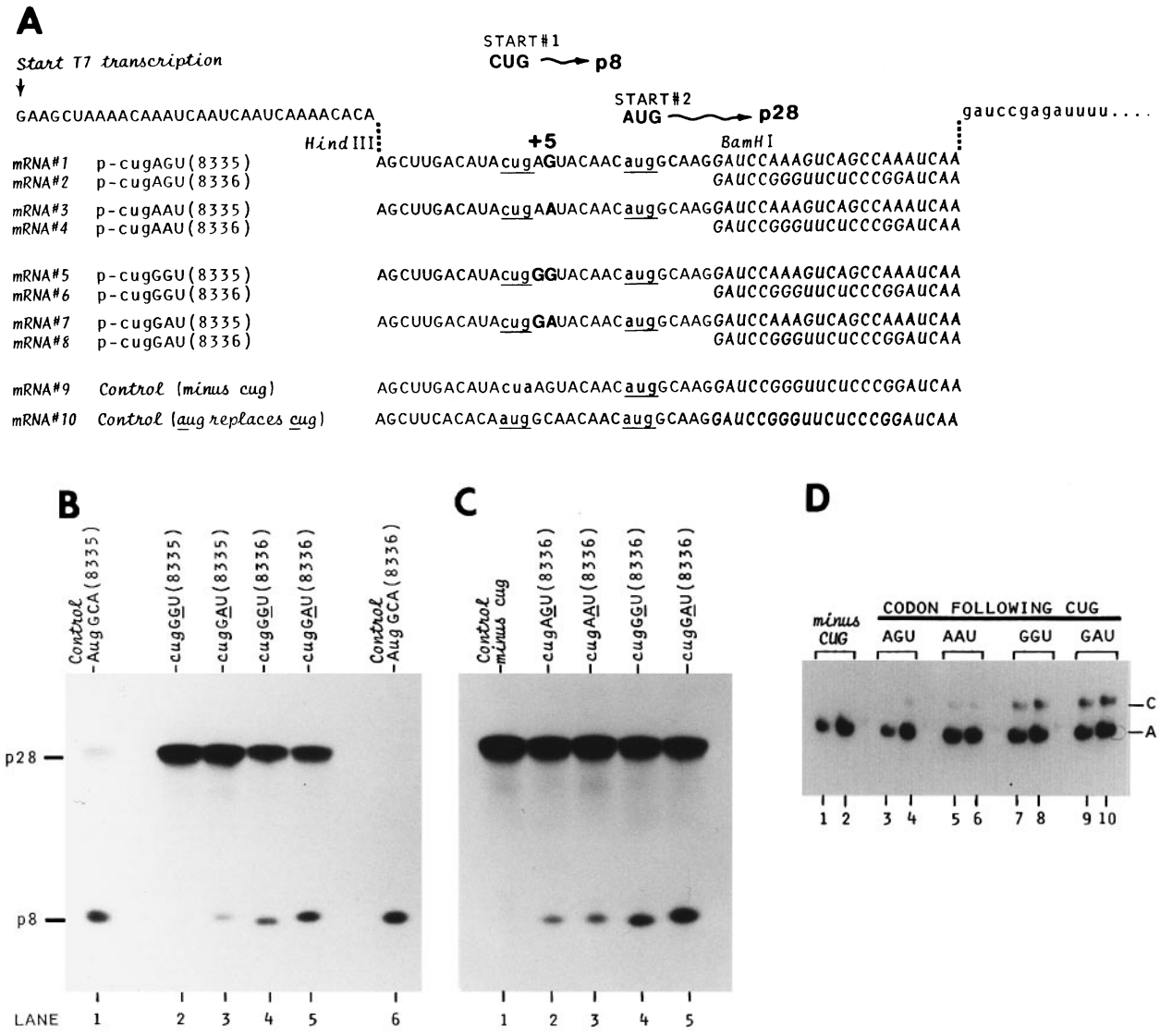

Fig. 7. Effects of mutations in positions +4 and +5 on recognition of a CUG initiator codon. (**A**) Sequences of the mRNAs used for translation. The major difference from Figure 1 is that AUG#1 has been replaced here by a CUG start codon. To compensate for the weakness of the CUG codon, the preceding sequence includes the optimal A in position –3. Each construct was tested with an unstructured sequence (8335) and a secondary structure-forming sequence (8336) downstream. (**B** and **C**) Autoradiograms of [^{3}H]leucine-labeled proteins resulting from translation under standard conditions (2 mM Mg^{2+}) of the mRNA indicated above each lane. The autoradiogram in (C) was exposed twice as long as that in (B). (**D**) Cropped autoradiograms of toeprinting reactions carried out with p-cugAGU mRNA (lanes 3 and 4), p-cugAAU (lanes 5 and 6), p-cugGGU (lanes 7 and 8) and p-cugGAU (lanes 9 and 10). Lanes 1 and 2 show a control reaction with mRNA #9 which lacks the upstream CUG start site. For each mRNA, adjacent bracketed lanes show toeprinting reactions carried out using the first two ^{32}P-containing fractions eluted from the Sepharose CL-4B column. Because of difficulty in synchronizing the collection when several columns are run at the same time, the first fraction (lanes 1, 3, 5, 7 and 9) contains less radioactivity in some cases. Consequently, the mRNAs are most easily compared by focusing on lanes 2, 4, 6, 8 and 10, where equal radioactivity was applied.

out by the results shown in Figure 6, where augGCG (Figure 6A, lane 1) and augGGA (Figure 6B, lane 6) were recognized efficiently despite the low abundance in reticulocytes of tRNAs corresponding to GCG and GGA (Hatfield *et al.*, 1979, 1982).

Initiation is best studied with an initiation-limited assay

The initiation-limited assay used here obviates two problems that can confuse assessment of the degree to which leaky scanning, caused by mutations around AUG#1, allows access to AUG#2. One problem with standard protein synthesis assays is that elongating ribosomes advancing from the upstream start site can block access to a second start site downstream, thus making AUG#1 appear stronger (less leaky) than it really is (Fajardo and Shatkin, 1990; Kozak, 1995). This sort of distortion, called elongational occlusion, was avoided in the present study by using inhibitors that prevent ribosomes from advancing beyond the initiation step.

A second potential problem is that varying bases +4, +5 and +6, and hence varying the penultimate amino acid, might change the stability of the test protein. This was argued not to be relevant in other studies (Boeck and Kolakofsky, 1994; Grünert and Jackson, 1994) because the amino acid changes that would result from the mutations in positions +5 and +6 should not have rendered the polypeptide unstable according to the N-end rule. How-

ever, the beautifully elucidated N-end rule pathway for protein turnover applies to proteins derived by proteolytic processing (Bachmair *et al.*, 1986; de Groot *et al.*, 1991; Gonda *et al.*, 1989; Varshavsky, 1995). One should not necessarily expect the predictions of the N-end rule to apply to nascent polypeptides in which the subterminal amino acid is varied. Unlike proteolytically derived polypeptides, the N-terminus of nascent polypeptides is subject to modification by methionine aminopeptidases, acetyltransferases, N-terminal amidases and other enzymes that may affect protein stability (Moerschell *et al.*, 1990; Kendall and Bradshaw, 1992; Stewart *et al.*, 1994; Baker and Varshavsky, 1995). It is not known whether the extent of these N-terminal modifications varies among batches of reticulocyte lysate, or whether the high level synthesis of some proteins *in vitro* might exceed the capacity of endogenous modifying enzymes. Because of these uncertainties, it seems dangerous to assume that differences in protein accumulation in response to mutations in positions +4, +5 and +6 reflect an effect of these nucleotides on the initiation of translation.

The present study gets around this concern by replacing the customary protein accumulation assay with one that directly monitors the initiation step of translation. Indeed, had I relied simply on measurement of protein yields, I might have concluded that initiation at augAAA was 5-fold more efficient than at augAUA (Figure 2A, lanes 6 and 8). However, those two start sites functioned with identical efficiencies when initiation was assayed directly (Figure 5B; Table I, entries 6 and 8).

Non-AUG start sites have the same flanking sequence requirements as AUG start sites

There is no compelling evidence for *cis*-acting elements in mRNAs that act uniquely at CUG, ACG and GUG start sites. Instead, non-AUG start codons seem to show a stronger dependence on the same ancillary features that augment AUG codon recognition. In some studies, for example, mutation of A^{-3} nearly abolished initiation from a CUG or ACG codon (Peabody, 1987; Portis *et al.*, 1994). In the experiments described herein, initiation at a CUG codon was barely detectable in the absence of G^{+4} (Figure 7C and D). The enhancing effect of downstream secondary structure, previously demonstrated for AUG start sites (Kozak, 1990a), was also evident with CUG start sites in Figure 7B. The strong dependence on these ancillary features probably follows from the fact that alternative start codons can form only two, instead of the usual three, base pairs with the anticodon in eukaryotic Met-$tRNA_i^{met}$. Prokaryotes are similar in the sense that initiation at a weak UUG start site requires an unusually strong Shine–Dalgarno interaction (Weyens *et al.*, 1988).

In an earlier study, Boeck and Kolakofsky (1994) postulated that A^{+5} and U^{+6} specifically augment initiation at non-AUG start sites, but they did not include tests with AUG in place of GUG. A companion paper (Grünert and Jackson, 1994) reported, on the other hand, that the effects of mutating positions +5 and +6 around a CUG start codon were qualitatively similar to the effects at an AUG codon.

In the present study, most of the mutations were introduced around AUG codons because the poor initiation at non-AUG codons, even in the best of circumstances, makes quantitation difficult. However, the experiment shown in Figure 7D using a CUG start site confirms the conclusion reached for AUG start sites: that recognition of the initiator codon improves when G is substituted for A in position +4, while substitutions in position +5 have no discernible effect.

There is also no compelling evidence for *trans*-acting factors in eukaryotes that specifically recognize non-AUG start codons. Some interesting experiments in yeast in which certain mutations in eIF-2 were shown to activate a silent UUG codon (Donahue *et al.*, 1988; Dorris *et al.*, 1995) are sometimes cited as evidence for a UUG-specific initiation factor. However, augmented initiation at UUG could be explained if the mutations in eIF-2 simply enhance its non-specific binding to mRNA. This could allow erroneous initiation events (i.e. use of a codon that only partially matches the anticodon in Met-$tRNA_i^{met}$) in the same way that streptomycin induces errors during polypeptide elongation by strengthening non-specific contacts between ribosomes and tRNA, and thus decreasing dependence on specific codon–anticodon contacts.

Materials and methods

Construction of plasmids

Plasmids used herein were derived from Riboprobe vector pSP64 (Promega Corp.) into which a CAT cartridge (Pharmacia Biotech) was previously inserted at the *Bam*HI site (Kozak, 1989b). The vector had been modified previously by introducing a T7 phage promoter (Kozak, 1994) followed by the sequence GAAGCTAAAACAAATCAATCAATCAAAACACAAGCTT. This synthetic 5′ non-coding sequence, which is devoid of secondary structure, was chosen because it supports efficient translation when an appropriate initiator codon is introduced downstream. Between the *Hin*dIII site (AAGCTT underlined above) and a nearby *Bam*HI site marked in Figure 1, I inserted synthetic deoxyoligonucleotides that contain two ATG (AUG) codons, as illustrated in Figure 1. Using the cassette mutagenesis technique depicted in Figure 1, I systematically varied the codon on the 3′ side of AUG#1. The plasmids and resulting mRNAs were named according to the sequence following the first start codon for translation: p-augGCA, etc., in Figure 1; p-cugAGU, etc., in Figure 7.

Because the presence of secondary structure appropriately positioned downstream from an AUG or CUG codon can augment initiation (Kozak, 1990a), two different sequences were used downstream. Beginning at the *Bam*HI site marked in Figure 1, the sequence was either GAUCCAAAGUCAGCCAAAUCAA (oligonucleotide 8335) or GAUCCGGGUUCUCCCGGAUCAA (oligonucleotide 8336). The latter sequence can form a stem–loop structure with a stability of –19 kcal/mol (Kozak, 1990a). Constructs that contain oligonucleotide 8336 are identified explicitly in the text and figures. All mRNAs discussed without mentioning the downstream sequence contained the structure-free oligonucleotide 8335, as in the mRNAs depicted in Figure 1.

Standard recombinant techniques used for these constructions were described previously (Kozak, 1989b). Plasmids were propagated using *Escherichia coli* RR1 (Gibco/BRL). The structures of all plasmids were confirmed by appropriate dideoxy chain-termination sequencing reactions using Sequenase-2 (U.S. Biochemical Corp.).

Synthesis of capped mRNAs

CsCl-purified plasmid DNA, linearized by digestion with *Ava*I, was used as the template for transcription by T7 RNA polymerase. Transcription reactions were carried out at 37°C as described previously (Kozak, 1989b) except that, after a 12 min incubation with m7GpppG caps (10 U/ml, Pharmacia Biotech), the GTP concentration was increased to 500 μM and incubation was continued for another 60 min. The reactions contained RNase inhibitor (150 U/ml, Gibco/BRL).

To ensure uniformity, all transcripts intended for a given translation experiment were synthesized using aliquots from a common reaction mixture, which included a trace of [^{3}H]UTP to facilitate quantification. mRNAs were extracted with phenol and purified by application to pre-spun Sephadex G50 columns (Boehringer Mannheim).

Complete translation assay

For the standard protein accumulation assay, a rabbit reticulocyte translation system supplemented with [^{3}H]leucine (140 μCi/ml, sp. act. 180 Ci/mmol) was programed with mRNA (12 μg/ml) and incubated for 1 h at 30°C. The Flexi reticulocyte lysate (Promega Corp.), which constituted 50% of the final reaction volume, was supplemented with 60 mM KCl and 19 non-radioactive amino acids at 20 μM each. In addition to the endogenous Mg^{2+} (stated by the supplier for each batch of lysate), reactions were supplemented with $Mg(CH_3COO)_2$ to give a final Mg^{2+} concentration of 2 mM, unless otherwise stated in the text or figure legends. A standard Mg^{2+} concentration of 2 mM was chosen because it was shown previously to support a pattern of context-dependent initiation *in vitro* similar to what is seen *in vivo* (Kozak, 1989b). To minimize variation, aliquots from a common reaction mixture were used for translation of all mRNAs in a given experiment.

Translation products were analyzed by polyacrylamide gel electrophoresis as described previously (Kozak, 1989b). The gels were impregnated with Entensify (DuPont NEN) before autoradiography with Kodak X-omat AR film at –70°C.

Primer extension assay of initiation complexes

Prior to ribosome binding, each mRNA was annealed with a ^{32}P-labeled deoxyoligonucleotide that would serve to prime the final reverse transcriptase step. The 23 nucleotide primer CTCAAAATGTT-CTTTACGATGCC is complementary to codons 16–23 in the CAT coding domain. The primer was first labeled at the 5′ end by incubation with T4 polynucleotide kinase and [γ-^{32}P]ATP (3000 Ci/mmol). An aliquot of the ^{32}P-labeled primer was then incubated with mRNA (~5 pmol of each) in 11 μl of 50 mM Tris–HCl (pH 7.5) for 2 min at 65°C followed by 10 min at 37°C. The primer–mRNA complexes were transferred to wet ice and held briefly while the reticulocyte reaction mixtures were assembled.

A rabbit reticulocyte lysate was used under the conditions described above except that [^{3}H]leucine was omitted and inhibitors of elongation (sparsomycin at 200 μM and cycloheximide at 90 μg/ml) were added. These inhibitors cause accumulation of initiation complexes in which the 80S ribosome is held at the AUG codon. Aliquots of a common reaction mixture were dispensed to glass tubes which were pre-incubated for 2 min at 30°C before adding the mRNA–primer complexes. After 4 min incubation at 30°C to allow ribosomes to engage the mRNA, the samples were applied to Sepharose CL-4B columns (15×0.7 cm) at 4°C. The column elution buffer contained 50 mM Tris–HCl (pH 8.3), 40 mM KCl, 6 mM $MgCl_2$, 5 mM dithiothreitol (DTT) and cycloheximide at 90 μg/ml. Column purification was omitted when more than six mRNAs were tested at one time.

Sepharose column fractions that contained ^{32}P-labeled ribosome–mRNA complexes were supplemented with 600 μM dATP, dGTP, dCTP and dTTP and with murine leukemia virus reverse transcriptase (Gibco/BRL Superscript II, used at 2 U/μl). Incubation at 37°C for 15 min allowed the primer to be extended up to the position of the bound ribosome. The positions of ribosomes on each mRNA were deduced from the lengths of the primer extension products, as determined by co-electrophoresis with an RNA sequence ladder. Appropriate ladders were generated from dideoxy sequencing reactions carried out at 42°C with avian myeloblastosis virus reverse transcriptase. Denaturing 8% polyacrylamide gels were used for electrophoresis. Autoradiograms of the dried gels, obtained in most cases with Kodak LS film, were quantified by densitometry. When weak start codons were tested (e.g. CUG in Figure 7D), AR film was used with an intensifying screen at –70°C.

A previous study that used this primer extension (toeprinting) assay describes some additional details and controls (Kozak, 1995).

Acknowledgements

Research in my laboratory was supported by grant GM33915 from the National Institute of General Medical Sciences, National Institutes of Health.

References

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.

Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.

Boeck,R. and Kolakofsky,D. (1994) Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. *EMBO J.*, **13**, 3608–3617.

de Groot,R.J., Rümenapf,T., Kuhn,R.J., Strauss,E.G. and Strauss,J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **88**, 8967–8971.

Donahue,T.F., Cigan,A.M., Pabich,E.K. and Valavicius,B.C. (1988) Mutations at a Zn(II) finger motif in the yeast eIF-2β gene alter ribosomal start-site selection during the scanning process. *Cell*, **54**, 621–632.

Dorris,D.R., Erickson,F.L. and Hannig,E.M. (1995) Mutations in *GCD11*, the structural gene for eIF-2γ in yeast, alter translational regulation of *GCN4* and the selection of the start site for protein synthesis. *EMBO J.*, **14**, 2239–2249.

Fajardo,J.E. and Shatkin,A.J. (1990) Translation of bicistronic viral mRNA in transfected cells: regulation at the level of elongation. *Proc. Natl Acad. Sci. USA*, **87**, 328–332.

Flinta,C., Persson,B., Jörnvall,H. and von Heijne,G. (1986) Sequence determinants of cytosolic N-terminal protein processing. *Eur. J. Biochem.*, **154**, 193–196.

Gonda,D.K., Bachmair,A., Wünning,I., Tobias,J.W., Lane,W.S. and Varshavsky,A. (1989) Universality and structure of the N-end rule. *J. Biol. Chem.*, **264**, 16700–16712.

Grünert,S. and Jackson,R.J. (1994) The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. *EMBO J.*, **13**, 3618–3630.

Hartz,D., McPheeters,D.S., Traut,R. and Gold,L. (1988) Extension inhibition analysis of translation initiation complexes. *Methods Enzymol.*, **164**, 419–425.

Hatfield,D., Matthews,C.R. and Rice,M. (1979) Aminoacyl-tRNA populations in mammalian cells. Chromatographic profiles and patterns of codon recognition. *Biochim. Biophys. Acta*, **564**, 414–423.

Hatfield,D., Varricchio,F., Rice,M. and Forget,B.G. (1982) The aminoacyl-tRNA population of human reticulocytes. *J. Biol. Chem.*, **257**, 3183–3188.

Irwin,B., Heck,J.D. and Hatfield,G.W. (1995) Codon pair utilization biases influence translational elongation step times. *J. Biol. Chem.*, **270**, 22801–22806.

Kendall,R.L. and Bradshaw,R.A. (1992) Isolation and characterization of the methionine aminopeptidase from porcine liver responsible for the co-translational processing of proteins. *J. Biol. Chem.*, **267**, 20667–20673.

Kozak,M. (1986a) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.

Kozak,M. (1986b) Bifunctional messenger RNAs in eukaryotes. *Cell*, **47**, 481–483.

Kozak,M. (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**, 8125–8148.

Kozak,M. (1989a) The scanning model for translation: an update. *J. Cell Biol.*, **108**, 229–241.

Kozak,M. (1989b) Context effects and (inefficient) initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol. Cell. Biol.*, **9**, 5073–5080.

Kozak,M. (1990a) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl Acad. Sci. USA*, **87**, 8301–8305.

Kozak,M. (1990b) Evaluation of the fidelity of initiation of translation in reticulocyte lysates from commercial sources. *Nucleic Acids Res.*, **18**, 2828.

Kozak,M. (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell. Biol.*, **115**, 887–903.

Kozak,M. (1992) A consideration of alternative models for the initiation of translation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.*, **27**, 385–402.

Kozak,M. (1994) Features in the 5′ noncoding sequences of rabbit α and β-globin mRNAs that affect translational efficiency. *J. Mol. Biol.*, **235**, 95–110.

Kozak,M. (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc. Natl Acad. Sci. USA*, **92**, 2662–2666 (and p. 7134).

Kozak,M. (1996) Interpreting cDNA sequences: some insights from studies on translation. *Mammalian Genome*, **7**, 563–574.

Kozak,M. and Shatkin,A.J. (1977) Sequences and properties of two ribosome binding sites from the small size class of reovirus messenger RNA. *J. Biol. Chem.*, **252**, 6895–6908.

Lin,F.-T., MacDougald,O.A., Diehl,A.M. and Lane,M.D. (1993) A 30-kDa alternative translation product of the CCAAT/enhancer binding protein α message: transcriptional activator lacking antimitotic activity. *Proc. Natl Acad. Sci. USA*, **90**, 9606–9610.

Moerschell,R.P., Hosokawa,Y., Tsunasawa,S. and Sherman,F. (1990) The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine *in vivo*. *J. Biol. Chem.*, **265**, 19638–19643.

Ossipow,V., Descombes,P. and Schibler,U. (1993) CCAAT/enhancer binding protein mRNA is translated into multiple proteins with different transcription activation potentials. *Proc. Natl Acad. Sci. USA*, **90**, 8219–8223.

Peabody,D.S. (1987) Translation initiation at an ACG triplet in mammalian cells. *J. Biol. Chem.*, **262**, 11847–11851.

Portis,J.L., Spangrude,G.J. and McAtee,F.J. (1994) Identification of a sequence in the unique 5′ ORF of the gene encoding glycosylated gag which influences the incubation period of neurodegenerative disease induced by a murine retrovirus. *J. Virol.*, **68**, 3879–3887.

Stewart,A.E., Arfin,S.M. and Bradshaw,R.A. (1994) Protein NH_2-terminal asparagine deamidase. *J. Biol. Chem.*, **269**, 23509–23517.

Varshavsky,A. (1995) The N-end rule. *Cold Spring Harbor Symp. Quant. Biol.*, **60**, 461–478.

Weyens,G., Charlier,D., Roovers,M., Piérard,A. and Glansdorff,N. (1988) On the role of the Shine–Dalgarno sequence in determining the efficiency of translation initiation at a weak start codon in the *car* operon of *Escherichia coli* K12. *J. Mol. Biol.*, **204**, 1045–1048.

Zhang,S., Zubay,G. and Goldman,E. (1991) Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene*, **105**, 61–72.

Received on November 7, 1996; revised on January 13, 1997