The EMBO Journal vol.13 no.15 pp.3618–3630, 1994

The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons

Stefan Grünert and Richard J.Jackson[1]

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK
[1]Corresponding author

Communicated by R.J.Jackson

Nucleotide substitutions were introduced into the initiation site of an influenza virus NS cDNA derivative at the +4, +5 and +6 positions (where the A of the AUG codon is defined as +1), in the background of either AUG or CUG as the initiation codon. Capped transcripts of these constructs were translated in rabbit reticulocyte lysate under conditions where the selection of initiation sites conformed to the scanning ribosome model. With CUG as the initiation codon, the efficiency of initiation was as strongly influenced by the nature of the residue in the +5 position as at +4, whilst the influence of the +6 position was smaller. The residues favourable to initiation were as follows: at +4, only G was stimulatory; at +5, A was strongly stimulatory and C fairly beneficial; and at +6, only U exerted any positive influence. The positive influence of the favourable residues (or the negative influence of unfavourable residues) at each position appeared to be additive. With AUG as the initiation codon, the pattern of response to mutations in the +4 and +5 positions was qualitatively similar, but the quantitative effects were smaller. Thus the optimum downstream context for initiation is A/CUGGAU.

Key words: initiation codon context/non-AUG initiation codons/scanning ribosome model/translation initiation

Introduction

The efficiency with which a scanning eukaryotic ribosome recognizes an AUG codon as an initiation site is dependent on the local sequence context of the codon (Kozak, 1986, 1989a), and is particularly responsive to the residues at the –3 and +4 positions (numbered with respect to the A of the AUG as the +1 position). If there are purines in these two positions, ideally ...ANN**AUG**G..., the efficiency of recognition is considered to be virtually 100%. If there are pyrimidines in both these positions, a substantial fraction of the scanning ribosomes appear to by-pass the initiation site and continue scanning to the next AUG codon (Kozak, 1986, 1989a).

This type of influence of context was first suggested by a statistical analysis of the sequences of the initiation sites of eukaryotic mRNAs (Kozak, 1984), which revealed a very high frequency of occurrence of A in the –3 position, and a somewhat lower frequency of G in the +4. A subsequent analysis (Kozak, 1987a) covering 699 vertebrate mRNA sequences suggested the following consensus: GCCGCCACC**AUG**G. In respect of the upstream context features, the overriding importance of A in the –3 position was confirmed by studying the expression in transfected COS cells of a cloned preproinsulin gene with mutations in this site (Kozak, 1986). The favourable influence of the upstream GCC repeats, and the C residues in the –1 and –2 positions, was demonstrated in the same system, but these context features seem subsidiary in that their influence was significant only when the –3 and +4 positions were occupied by unfavourable pyrimidines (Kozak, 1986, 1987b). As for the downstream context, mutation of the preproinsulin gene in the +4 position showed that a G residue in this site promoted initiation site recognition, whilst a U was unfavourable, but no data on the effect of other base substitutions at this site were presented (Kozak, 1986). Although it seems to be quite widely assumed that an A in the +4 position is also beneficial, any such assumption is based solely on the fact that in the survey of 699 vertebrate mRNAs A occurred significantly more frequently in this position than either U or C (Kozak, 1987a), but there is no direct experimental evidence to support the assumption. In addition, there has been no direct test of whether the efficiency of recognition of the initiation codon is influenced by the nature of residues downstream of the +4 position.

In a few cases initiation occurs at a codon other than AUG, most frequently at CUG, but also at ACG and less frequently at GUG (Becerra *et al.*, 1985; Peabody, 1987, 1989; Curran and Kolakofsky, 1988; Gupta and Patwardhan, 1988; Hann *et al.*, 1988; Patwardhan and Gupta, 1988; Florkiewicz and Sommer, 1989; Prats *et al.*, 1989a,b; Acland *et al.*, 1990; Boeck *et al.*, 1992). This is a very rare event, and by no means every CUG or ACG in the 5′-region of an mRNA functions as an initiation codon (Dasso and Jackson, 1989b). Whilst most of the CUG and ACG codons that are functional initiation codons have optimal context features in respect of the –3 and +4 positions, this alone does not seem sufficient to explain which will be used and which will not. It therefore seems likely that subsidiary context features play an especially important role in determining whether a given non-AUG codon is likely to be used as a functional initiation codon.

In a previous report we showed surprisingly efficient initiation *in vitro* at a CUG codon introduced into the influenza virus NS (non-structural protein) mRNA in place of the normal AUG codon (Dasso and Jackson, 1989b). Indirect evidence suggested that sequences downstream of the initiation site might be important determinants of this high efficiency of initiation, since a small deletion which changed this sequence from A/CUGGAUCCA to A/CUGAGCUUU drastically reduced the utilization of the initiation codon. We have therefore carried out a systematic test of the influence of single base substitutions in the +4, +5 and +6 positions on the efficiency of

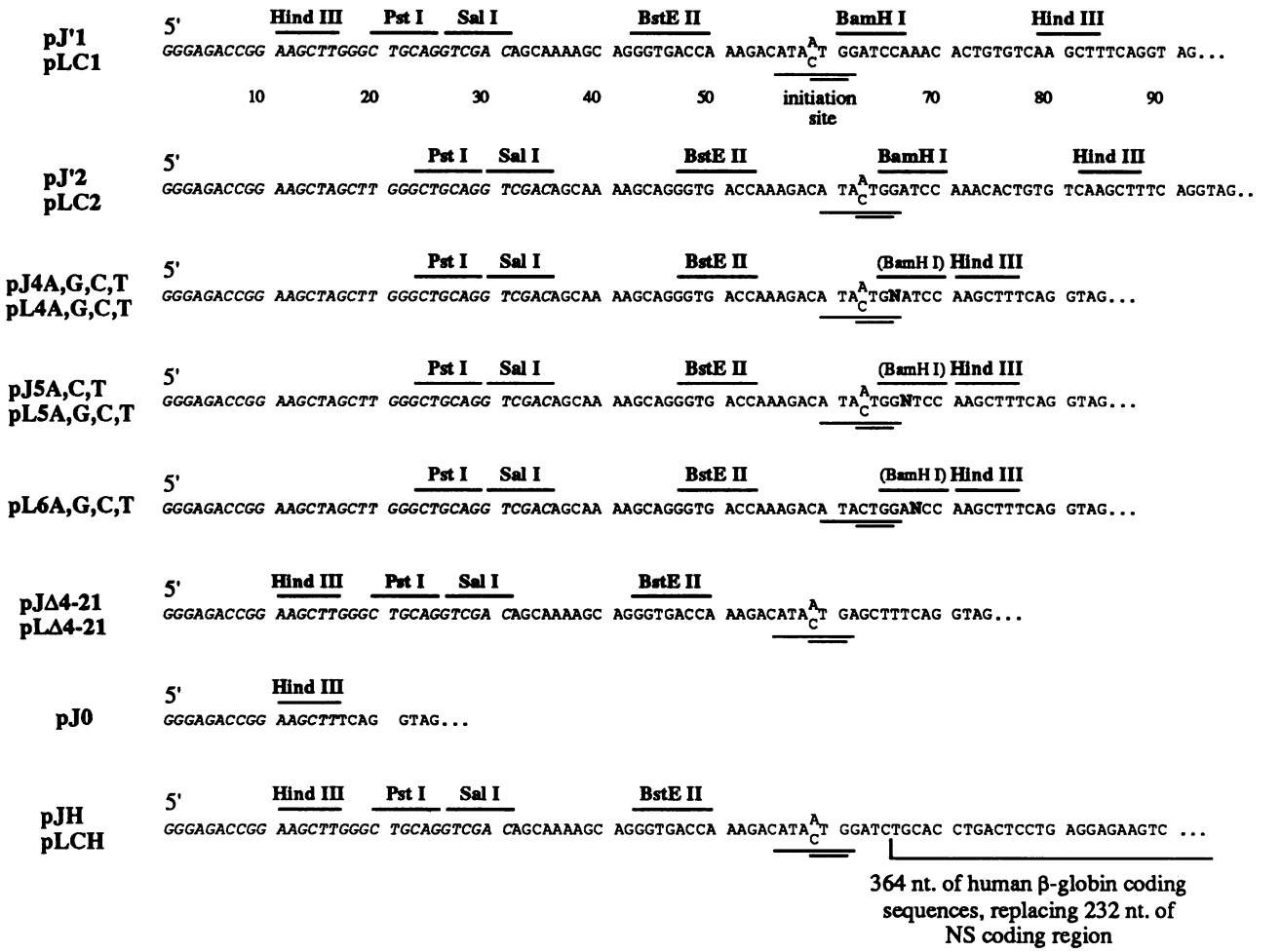

Fig. 1. Nucleotide sequences of the 5′-regions of the influenza virus NS cDNA constructs. The sequences written 5′–3′ correspond to the mRNA sequences, starting with the first transcribed nucleotide and with sequences derived from the pGEM-2 polylinker shown in italics. All constructs with a pJ or pJ′ designation have an ATG at the site of the initiation codon (except pJ0 which carries a deletion covering this site), and all constructs with a CTG in this site are designated pLC or pL. The initiation site from the –3 to +4 position is underlined, and the ATG (CTG) initiation codon doubly underlined. Restriction enzyme sites mentioned in the text are shown.

recognition of the CUG codon as an initiation site, with comparative studies of the efficiency of utilization of an AUG initiation codon. The results show that (i) the nature of the residue in the +5 position is at least as important as the +4, while the +6 position is less important; (ii) in the +4 position only G exerts a positive influence; (iii) in the +5 position, A is strongly beneficial and C exerts a moderately positive influence; and (iv) at +6 only U is stimulatory to initiation site recognition. Thus we conclude that the optimum downstream codon for initiation site utilization is GAU, which is precisely the codon found in this position in the influenza virus NS sequence, and is strikingly in complete agreement with the results of Böck and Kolakofsky (1994) concerning three entirely different mRNA species, as reported in the accompanying paper.

Results

Constructs and strategy

The parent construct for this work was pJ′1 described by Dasso and Jackson (1989b). This consists of the entire influenza virus (strain A/PR/8/34) NS cDNA (with a C residue inserted in the 5′-UTR in order to create a unique *Bst*EII site) placed downstream of 31 nucleotides of polylinker sequences (Figure 1) designed to mimic the cellular mRNA sequences found at the 5′-end of viral mRNAs in cells infected with influenza virus. pLC1 is identical apart from the substitution of a CUG for the AUG initiation codon, and is translated in the reticulocyte lysate system at ~50% of the efficiency of pJ′1 mRNA, where efficiency is defined as the maximum incorporation observed when assayed over a range of RNA concentrations. For consistency, all values for relative efficiency of translation of different mRNAs in this dose–response assay will be expressed relative to the efficiency of pJ′1 mRNA translation, and all constructs with an AUG initiation codon carry the pJ prefix, whilst those with a CUG are designated pL.

For the purposes of this current work it was useful to in-fill the *Hin*dIII site near the 5′-end of the polylinker sequences, creating pJ′2 and pLC2 (Figure 1), which showed translation efficiencies of 80% and 45% respectively. This small drop in efficiency on in-filling this site may be due to the increase in secondary structure base-pairing near the 5′-end of the resulting RNA. Mutations in the +4, + 5 and +6 positions were then made by

insertion of oligonucleotides (with redundancy in the appropriate position) between the *Bst*EII and *Hin*dIII sites of pJ'2 or pLC2 (Figure 1). These constructs are designated by the position of the mutation and the nature of the residue in that position, i.e. pJ4A has an AUG initiation codon and an A in the +4 position, while pL5G has a CUG initiation codon and a G in the +5 position. It should be noted that mutations in the +6 position were made only in the background of a CUG initiation codon, and that in addition we were unable to obtain pJ5G. The oligonucleotide insertions deleted a small segment of the NS coding region, namely the 12 residues immediately upstream of the *Hin*dIII site in pJ'2/pLC2. This was found to result in a small decrease in translation efficiency as defined above: the translation efficiency of pJ4G was 76% as opposed to 80% for pJ'2, and the efficiency of pL5G was 32% as opposed to ~45% for pLC2. The reasons for this drop in efficiency have not been examined in detail.

Of the other constructs used, pJ0 is a deletion between the two *Hin*dIII sites of pJ'1, which removes all of the influenza virus NS 5'-untranslated region, the initiation codon and a short length of downstream coding sequence, but leaves most of the open reading frame intact. pLΔ4-21 and pJΔ4-21 are effectively deletions between the in-filled *Hin*dIII site in the coding region of pLC1 or pJ'1, respectively, and the *Bam*HI site from which the overhang had been removed (Figure 1). The numbering of these and other deletion mutants refers to the positions deleted, with the A/C of the AUG/CUG initiation codon considered as the +1 position.

All constructs were cloned in pGEM-2 as vector, thus allowing capped transcripts to be prepared *in vitro* using bacteriophage T7 RNA polymerase. Care was taken to remove all excess nucleotides from the transcript preparations, particularly the m^7GpppG cap analogue which can otherwise cause inhibition of translation at high RNA concentrations. It should be noted that all these mutations in the +4, +5 and +6 positions, with the exception of pL6C, will change the identity of the amino acid residue following the initiating methionine residue, which in turn may influence whether this N-terminal methionine is cleaved off the translation product (Sherman *et al.*, 1985, 1993). In the case of the wild-type NS product, with the N-terminal sequence MetAspPro, the methionine residue would be expected to be retained and to be acetylated (Sherman *et al.*, 1993), and indeed the translation product proved to be refractory to sequencing by Edman degradation (Dasso and Jackson, 1989b). Retention of the initiating methionine residue would also be expected for all the mutants studied here except pL5G (Gly), pJ/pL5T (Val) and pJ/pL5C (Ala), though the neighbouring proline residue may inhibit complete cleavage of the initiating methionine residue from the translation product even of these mRNAs (Sherman *et al.*, 1993). The wild-type NS protein has a total of nine methionine residues including the N-terminal methionine. Thus if these mutations in the +5 position of the pL series were absolutely neutral to translation efficiency but resulted in complete removal of the initiating methionine residue, the incorporation of [^{35}S]methionine would decrease to 28% (relative to the pJ'1 standard) as opposed to a value of 32% for the parent pL4G/pL5A with a GAU codon. For the pJ series of constructs the equivalent figures are that the relative

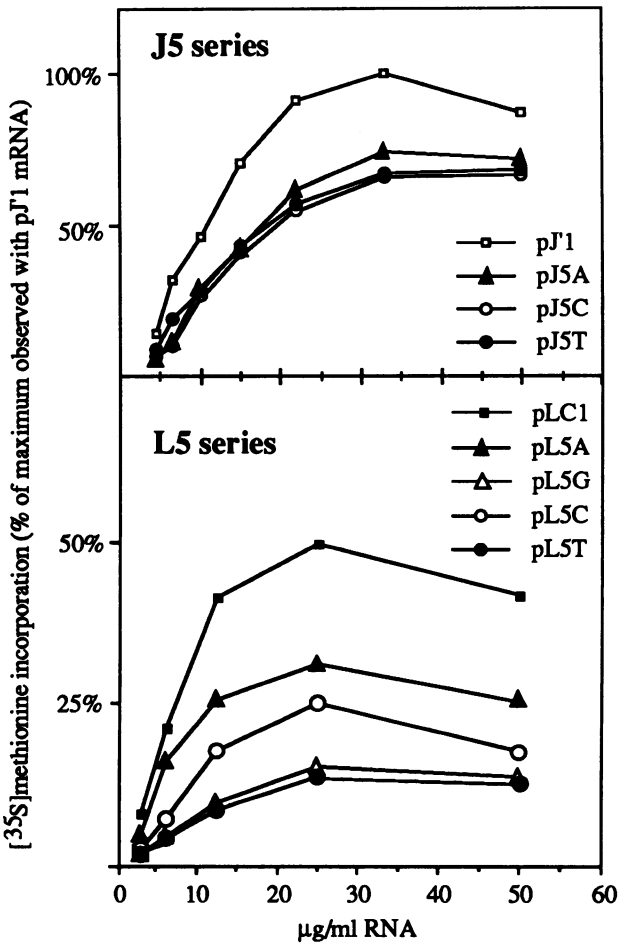

Fig. 2. Representative results from the [^{35}S]methionine incorporation assays. Capped RNA transcripts of the designated constructs were assayed over a range of RNA concentrations (given as μg/ml in the final assay mix), with 0.45 mM added Mg^{2+}. After an incubation period of 60 min, the incorporation of [^{35}S]methionine into acid-precipitable material was determined, and the results are expressed relative to the maximum incorporation observed with pJ'1 mRNA set at 100%. The upper panel shows results with pJ constructs (AUG initiation codon), and the lower panel pLC or pL constructs (CUG initiation codon), with the symbols designating the results with individual constructs as shown in the figure.

incorporation would decrease to 68% for pJ5T/C, as opposed to 76% for the pJ4G/pJ5A parent. In evaluating our data in terms of effects of mutations on the efficiency of translation initiation we have attributed significance only to changes very much larger than these, and thus the fate of the initiating methionine residue will not introduce ambiguity into the interpretation.

Assay of translation efficiency over a range of RNA concentrations

Two types of assay were used to evaluate the relative efficiency of initiation at the 5'-proximal site of the various RNA derivatives: (i) assay of [^{35}S]methionine incorporation over a range of mRNA concentrations at a fixed concentration of added Mg^{2+} (0.45 mM); and (ii) the effect of varying Mg^{2+} concentration (at a fixed RNA concentration) on the relative yield of full-length product as opposed to products initiated at internal sites by ribosomes that have scanned past the authentic initiation site. Representative results from the first type of assay are

Table I. Summary of results comparing the translation characteristics of the different mutants

AUG-initiated constructs	Relative max. incorporation	Z/F ratio	CUG-initiated constructs	Relative max. incorporation	Z/F ratio
pJ′1	100	0.29	pLC1	50	0.61
pJ′2	80	0.31	pLC2	45	0.59
pJH	100	n.d.	pLCH	60	n.d.
pJΔ4-21	40	0.71	pLΔ4-21	10	5.82
pJ4A	74	0.41	pL4A	14	3.45
pJ4G	76	0.29	pL4G	32	0.97
pJ4C	74	0.49	pL4C	17	3.08
pJ4T	73	0.44	pL4T	14	3.26
pJ5A	75	0.29	pL5A	31	0.91
			pL5G	16	3.82
pJ5C	73	0.41	pL5C	25	1.57
pJ5T	72	0.61	pL5T	14	3.83
			Assayed in a different batch of lysate		
			pL6A	21	0.82
			pL6G	19	0.88
			pL6C	21	0.86
			pL6T	31	0.65

The relative maximum incorporation was determined in RNA dose–response assays and is expressed relative to the maximum incorporation observed using pJ′1. The Z/F ratio is defined in the text. n.d., not determined because the Z/F ratios for pJH and pLCH cannot be directly compared with the values obtained for the other mRNA species. Note also that the assays of the pL6 series were carried out in a different batch of lysate. The relative maximum incorporation values for these constructs can be legitimately compared with the others, but the Z/F ratios cannot be directly compared in this way.

shown in Figure 2. With all mRNAs that had been purified from putative contamination by carry-over of m^7GpppG cap analogue, incorporation reached a plateau level at a concentration between 25 and 50 μg/ml (Figure 2), similar to our previous results (Dasso and Jackson, 1989b; Dasso *et al.*, 1990). The maximum incorporation at saturating mRNA concentration differed between different RNA species, and, in general, parallel differences were seen at sub-saturating RNA concentration (Figure 2). Table I summarizes the maximum incorporation for the different mRNAs relative to pJ′1 mRNA as the standard.

The results in Table I show that amongst the constructs with a CUG initiation codon, pL4G, pL5A and pL6T were translated with the highest efficiency (31–32% relative to pJ′1), and were roughly equivalent as would be expected since although they are individual constructs they have the same (wild-type) sequence ...**CUG**GAU. pL5C (25% relative incorporation) was considered significantly more efficient than any of the other mutants in the +4 or +5 position (14–17% efficiency), none of which seemed significantly more (or less) efficient than any other, although they do seem significantly worse than pL6A, pL6C and pL6G (all with ~20% relative efficiency). This probably reflects the fact that unfavourable residues in the +6 position have a smaller negative influence than unfavourable or 'neutral' residues at +4 or +5. However, it is notable that by far the most debilitated mutant of the pL series was pLΔ4-21, which deviates from the wild-type sequence in all three positions (AGC as opposed to GAU).

Amongst the pJ series, the differences were much smaller. It is true that pJ4G amongst the pJ4 series, and pJ5A amongst the pJ5 series, were usually slightly more efficiently translated than any others, but the difference was small and not absolutely reproducible in all assays, which is probably a reflection of the influence of variable RNA quality. Notably, the only mutant in the pJ series to show a significant reduction in efficiency (40% as opposed to ~75% for all others) was pJΔ4-21, the counterpart of the very inefficient pLΔ4-21.

It should be noted that while the results summarized in Table I can be taken as an indicator of the rank order, or hierarchy, of the different mRNAs in respect of the efficiency of utilization of the 5′-proximal initiation site, they cannot be interpreted as an absolute quantitative measure of relative efficiency since, as shown below, this is influenced by the Mg^{2+} concentration. A concentration of 0.45 mM added Mg^{2+} was chosen as a compromise for these dose–response assays, since (i) RNAs of the pJ series are translated fairly efficiently under this condition even though, as shown below, it is above the optimum Mg^{2+} concentration for these mRNAs, and (ii) translation of RNAs of the pL series at 0.45 mM Mg^{2+} gives mainly, but not exclusively, the full-length translation product rather than products initiated at downstream sites within the open reading frame as a result of leaky scanning. The compromise is unavoidable, for although a lower Mg^{2+} concentration would promote more efficient translation of all the mRNAs which have an AUG codon, it would also result in a significant increase in the yield of incomplete products from the pL mRNAs (Figures 3 and 4), which would completely invalidate the use of [^{35}S]methionine incorporation as a measure of the utilization of the 5′-proximal site. However, even at 0.45 mM added Mg^{2+}, the synthesis of short products from the least efficient of the pL series of mRNAs makes a contribution to overall [^{35}S]methionine incorporation that cannot be regarded as negligible, as shown below. These considerations mean that the data on [^{35}S]methionine incorporation in Table I will overestimate the efficiency of utilization of the CUG in the pL mRNAs; and that the magnitude of this over-estimate will become proportionately greater in the case

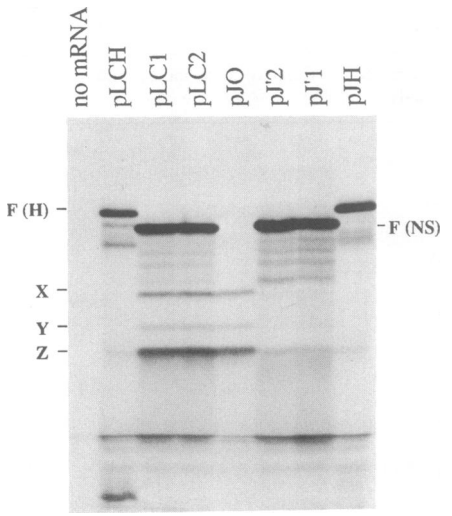

Fig. 3. The small products translated from pLC and pL mRNAs result from initiation at internal sites. Capped transcripts of the designated constructs were translated at a final concentration of 15 μg/ml, and the products analysed by gel electrophoresis, followed by autoradiography of the dried gel. F (NS) designates the full-length translation product (the influenza virus NS polypeptide) from the 5′-proximal initiation site of pJ′1, pJ′2, pLC1 and pLC2 mRNAs; the origin of the translation products designated X, Y and Z, whose synthesis is initiated at downstream sites within the NS open reading frame, is discussed in the text. F(H) designates the full-length translation product from the chimeric NS/β-globin constructs pJH and pLCH. Note that the translation assay was carried out with 0.5 mM added EDTA in order to increase the yield of short translation products, and the autoradiograph was deliberately over-exposed to reveal them.

of the least efficient mRNA species, most notably pLΔ4-21 (Table I).

The effect of varying Mg^{2+} concentration on initiation site utilization

The reasons for the original choice of the influenza virus NS cDNA for our studies of initiation site selection *in vitro*, as reported previously (Dasso and Jackson, 1989b; Dasso *et al.*, 1990) and continued here, were: (i) this mRNA is translated with exceptionally high efficiency in our lysates, even when a CUG is substituted for the AUG initiation codon, and (ii) it gives far fewer incomplete products than many other mRNAs (Dasso and Jackson, 1989a,b; Dasso *et al.*, 1990). With the NS mRNA, two different patterns of incomplete products are seen under two quite different sets of circumstances. In translations of the parent pJ′1 mRNA, or any derivative with a strong AUG initiation codon, a heterogeneous group of incomplete products, most of them quite large, is seen particularly at high mRNA concentrations (Figure 3). Such products are likely to be the result of premature termination of translation, and their synthesis is a reflection of the general breakdown in fidelity of translation at super-saturating mRNA concentrations that we have reported previously (Dasso and Jackson, 1989a).

The other scenario leading to the synthesis of incomplete products, but a quite different set of smaller proteins, is when the initiation codon is used inefficiently, for example a CUG in a mediocre context. In this case the synthesis of the smaller products is fairly independent of mRNA concentration. Since the same products are synthesized when the 5′-proximal initiation codon has been entirely deleted, as in the case of pJ0 (Figure 3), and yet their synthesis is abrogated by insertion of a hairpin loop at the *Pst*I site near the 5′-end of the mRNA (M.C.Dasso and R.J.Jackson, unpublished observations), they must arise as a result of initiation by ribosomes that have scanned past the CUG initiation codon and have selected downstream sites for initiation, entirely in agreement with the postulates of the scanning ribosome model (Kozak, 1989a). There are three such products routinely observed, designated X, Y and Z (Figure 3). The size of Z suggests that it arises from initiation at one or both of two closely spaced AUG codons in the sequence .....AAA**AUG**A-CC**AUG**GGC, the fifth and sixth AUG codons from the 5′-end of the pJ′ series of constructs. We consider it likely that initiation occurs at both of these AUG codons, since deletion of just the downstream of these two (AUG-6) reduced the yield of Z by some 40% (L.Johnson and R.J.Jackson, unpublished observations). Moreover, the deletion of the upstream AUG codon (AUG-5) and the downgrading of the context of AUG-6 to UUC**AUG**GGC, as occurs when the 5′-proximal sequences of the NS open reading frame are replaced by β-globin coding sequences in pJH and pLCH, results in a marked reduction in the relative yield of Z (Figure 3). The second, third and fourth AUG codons all lie in a different reading frame, and all have a poor context by the usual criteria: two have G at −3 but a pyrimidine at +4, and the other has a C at −3 coupled with A at +4. None of the three reading frames has an AUG-initiated open reading frame long enough to account for the synthesis of X and Y, and thus these products probably originate from initiation at in-frame non-AUG codons, a supposition consistent with the fact that the synthesis of both of them is inhibited by low Mg^{2+}, in distinct contrast to the synthesis of Z, as discussed below. On the basis of product size, X almost certainly stems from initiation at an in-frame CUG codon (....GGU**CUG**GAC...), and Y from a GUG codon (....AUA**GUG**GAG...), both of which have a good context by the usual criteria of a purine at –3 and a G at the +4 position (Kozak, 1986, 1989a), and in addition both have an A at +5, which, as shown in this paper, exerts an equally positive influence as the G at +4.

As Z is produced by those ribosomes which have scanned past the authentic initiation codon, the ratio of the yield of Z relative to the yield of full-length product (F) provides an internally controlled measure of the efficiency with which the 5′-proximal initiation codon is recognized. This analysis was found to be especially informative if it is carried out over a range of Mg^{2+} concentrations, from a maximum concentration of 1.5 mM added Mg^{2+} down to 1.2 mM Mg^{2+} 'subtracted' (i.e. 1.2 mM EDTA added). At the highest Mg^{2+} concentration used, there was almost exclusive synthesis of the full-length product, though a low yield of some incomplete products was seen in translations of some mRNAs with a CUG initiation codon (Figure 4). As the Mg^{2+} concentration was lowered progressively, the first change noted with all mRNAs was an increase both in total [^{35}S]methionine incorporation and in the yield of full-length product. Further reduction in Mg^{2+} concentration resulted in a decrease in the yield of full-length product and an increase in the yield of the incomplete products, especially of Z.

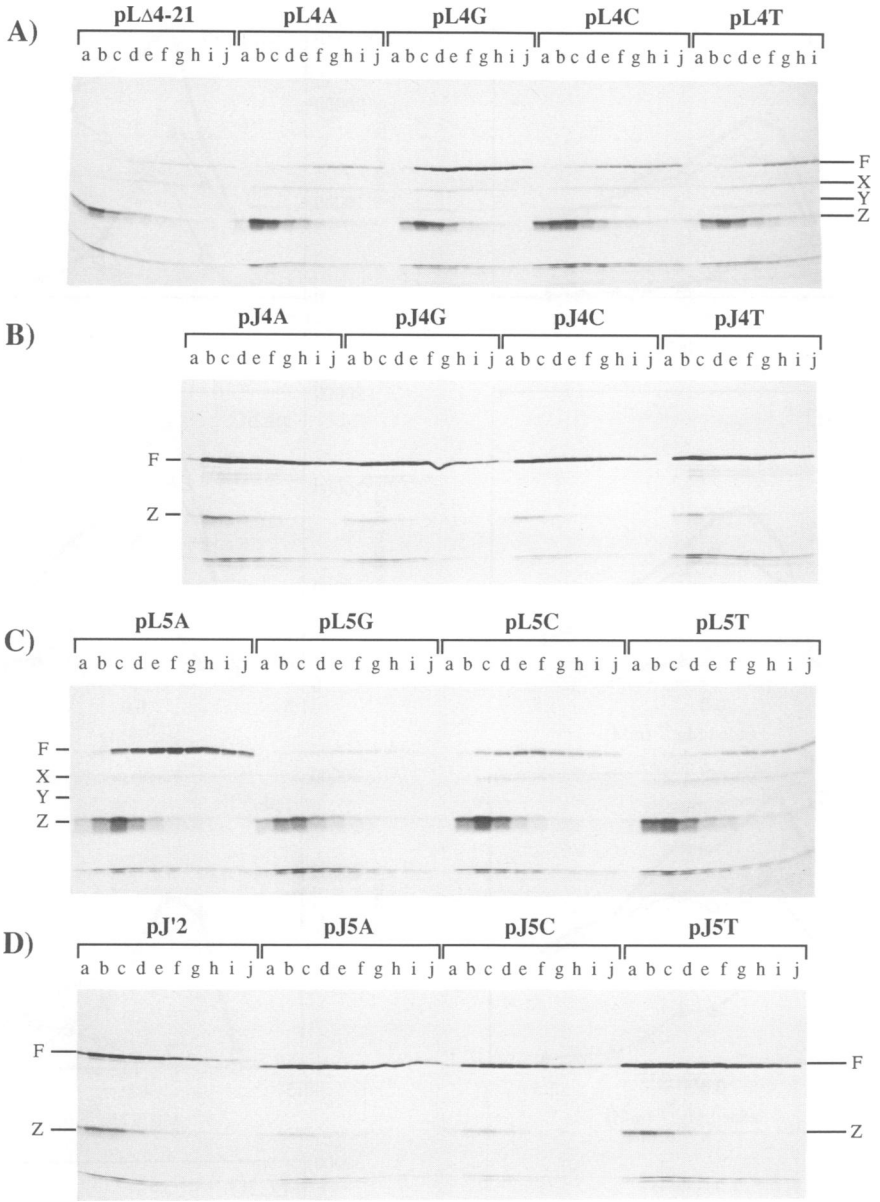

Fig. 4. The effect of Mg^{2+} concentration on the pattern of translation products derived from the (**A**) pL4, (**B**) pJ4, (**C**) pL5 and (**D**) pJ5 series of mRNAs. Capped transcripts were translated at a final concentration of 13 μg/ml under standard conditions with: (a) 1.2 mM EDTA added; (b) 0.9 mM EDTA; (c) 0.6 mM EDTA; (d) 0.3 mM EDTA; (e) no added EDTA or $MgCl_2$; (f) 0.3 mM $MgCl_2$ added; (g) 0.6 mM $MgCl_2$; (h) 0.9 mM $MgCl_2$; (i) 1.2 mM $MgCl_2$; and (j) 1.5 mM $MgCl_2$. The full-length product is designated as F, and the incomplete products discussed in the text (X, Y and Z) are shown.

This transition from synthesis of predominantly the full-length product to an increasing frequency of initiation at downstream sites occurred at significantly different Mg^{2+} levels according to the mRNA species (Figure 4).

With the pJ series of constructs, synthesis of full-length product was particularly resistant to the inhibitory effects of low [Mg^{2+}], but nevertheless some synthesis of Z occurred at the lower end of the Mg^{2+} concentration range (Figure 4). With the pL4 series of constructs, the transition from synthesis of mainly full-length product to predominant synthesis of Z clearly occurred at a lower Mg^{2+} level with pL4G and pL5A, than in the case of pL4A, pL4C, pL4T, pL5G or pL5T, whilst pL5C was intermediate between these two extremes. This correlates with the fact that the absolute yield of full-length product from pL4G and pL5A mRNA was higher than that from pL5C which in turn was higher than that from any other mRNA of the pL4 and pL5 series (Figure 4).

In order to place these data on a quantitative basis, scanning densitometry of different exposures of the autoradiographs was used to estimate the yields of the full-length product and Z. The products X and Y were ignored since they make only a minor contribution to the overall translational yield. Representative plots of these scanning densitometry data are given in Figure 5, which clearly shows the transition from synthesis of mainly the full-length product to increasing yield of Z as the Mg^{2+} concentration was reduced. To compare the different mRNAs, the ratio of the maximum yield of Z to the maximum yield of the full-length product (F) was calcu-

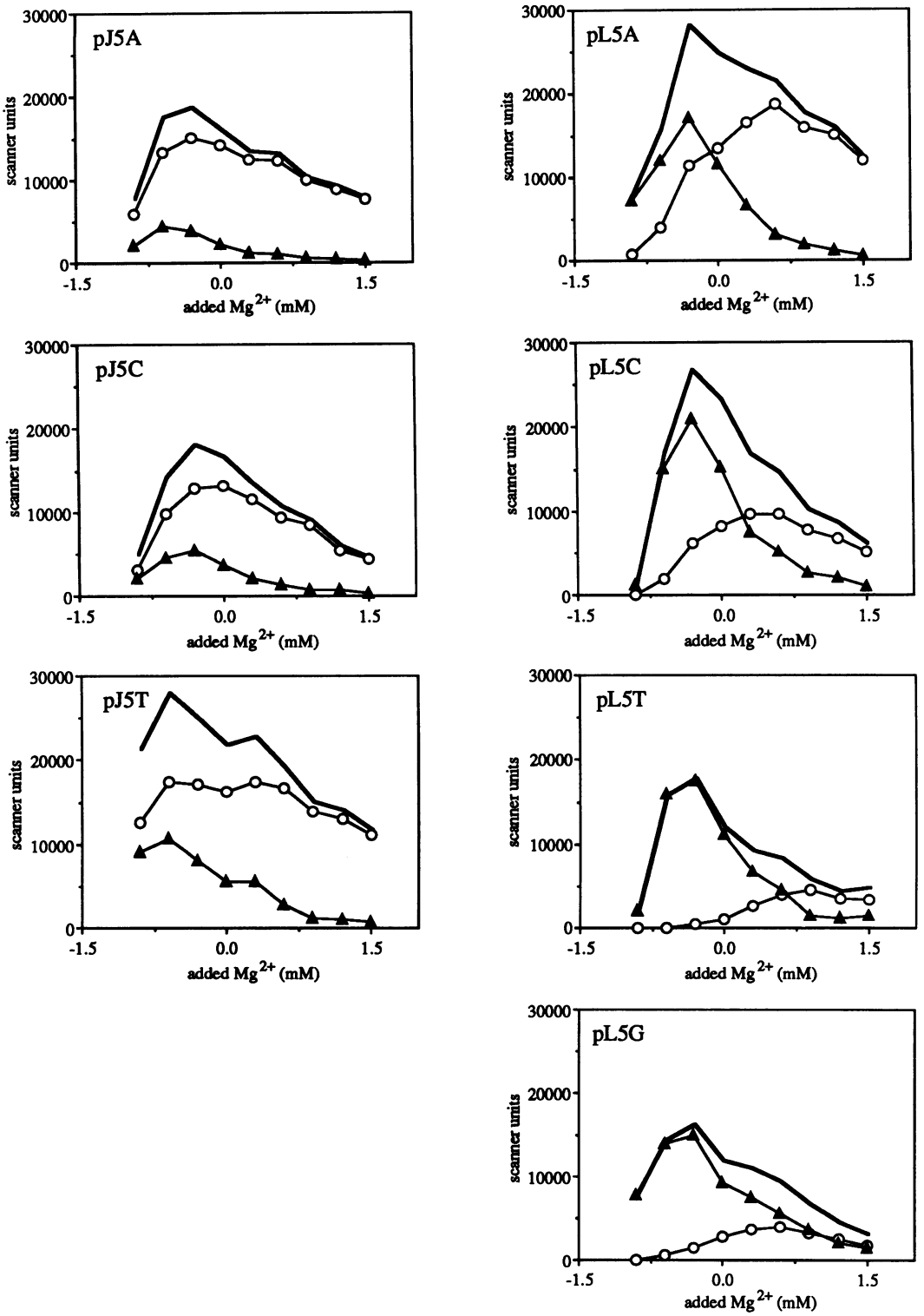

Fig. 5. A representative analysis of the effect of Mg^{2+} concentration on the yield of the full-length translation product (F) and the shorter product (Z) obtained on translation of mRNAs of the pJ5 and pL5 series. The yields of the different products were determined by scanning densitometry of different exposures of autoradiographs similar to Figure 4. The yields of the full-length product, F (○) and of Z (▲) are plotted against the concentration of Mg^{2+} added or 'subtracted' (i.e. the concentration of EDTA added). The thickened line graph without symbols represents the sum of the yields of F and Z, as determined by scanning densitometry.

lated for each mRNA species, giving the results shown in Table I. A low ratio of maximum yields of Z/F is indicative of high efficiency of utilization of the AUG or CUG initiation codon, and a high Z/F ratio indicates low efficiency. The advantage of this method of assessing initiation efficiency is that it is internally controlled and is not influenced by uncontrolled variables such as differences in the quality of the RNA preparation, or slight errors in RNA concentration. The reproducibility of the method is shown by the fact that pL4G and pL5A,

which, though they were individual constructs, are actually identical to each other and to the wild-type sequence, gave very similar values for the ratio of Z/F: 0.97 and 0.91 respectively. However, different batches of reticulocyte lysate seem to differ in the extent to which they are permissive to scanning penetrating far into the open reading frame, and this explains why pL6T, though nominally identical to pL4G and pL5A, consistently gave a lower Z/F ratio when assayed in a different batch of lysate (Table I). Thus the values of Z/F for the pL6 series of constructs cannot be compared directly with the values presented for the other constructs in Table I, though the data for [^{35}S]methionine incorporation do seem to be directly comparable between different batches of lysate, if normalized to the pJ′1 standard (Table I).

For the mRNAs with a CUG initiation codon, the results shown in Table I for the Z/F yield ratio are in general agreement with the translational efficiency data taken from the RNA dose–response assays, in that they suggest the same hierarchy of the constructs. For the +4 position, only a G in this position has a beneficial effect, and for the others the Z/F ratio varies from 3.08 for pL4C up to 3.45 for pL4A, a difference which is probably not significant. For the pL5 series, pL5A clearly gives almost the same result as the identical pL4G (Z/F ratio of 0.91), whilst pL5C (1.57) is better than pL5G or pL5T which gave a Z/F ratio of 3.8. The higher value of the Z/F ratio for pL5G/T than for pL4A/C/T could be explained by the fact that the initiating methionine residue is likely to be removed from the full-length translation product of pL5G/T, but not pL4A/C/T. These considerations lead to the conclusion that the influence of the +5 position is certainly at least as significant as the +4, but probably not more significant. On the other hand, the +6 position clearly exerts a smaller effect since the maximum Z/F ratio observed, albeit in a different batch of lysate which gave a lower yield of Z/F with all mutants tested, was only ~0.85. This is consistent with the fact that in the [^{35}S]methionine incorporation assay pL6A, pL6C and pL6G gave a relative incorporation (with pJ′1 as the standard) of ~20%, whereas with the unfavourable residues at the +4 and +5 positions the relative incorporation was 14–15% (Table I). Nevertheless the most striking fact is that by far the highest ratio of Z/F (5.8) was seen with pLΔ4-21, which was also very significantly less efficient in the [^{35}S]methionine incorporation assay than any other construct of the pL series.

Although the incorporation assay could not reveal any absolutely consistent significant differences amongst the pJ4 and pJ5 series, the more sensitive assay of determination of the Z/F ratio clearly reveals that pJ4G and the identical pJ5A (Z/F = 0.29 in both cases) are distinguishable from all other pJ4 and pJ5 mRNAs (Table I), consistent with the results obtained studying the pL series. Notably, the only construct in the pJ series to show a markedly elevated Z/F ratio (0.71) was pJΔ4-21, which also showed the lowest activity in the [^{35}S]methionine incorporation assay (Table I).

The contributions of the individual downstream positions are additive

Since the initiation codon in the NS mRNA is followed by a GAU codon, which has the optimal residue in all three positions, mutation of any one position still leaves the optimum residue in the other two positions of this codon. This raises the possibility that the influence of each position might be additive, thus providing an explanation for the very high efficiency of utilization of the CUG initiation codon in the original pLC1 mutant, and its derivatives. One observation consistent with this idea is that by the criteria of both the dose–response assay and the ratio of the yields of Z/F, the CUG of pLΔ4-21 is recognized very much less efficiently than the CUG of any of the pL4, pL5 or pL6 series of constructs, including pL5G and pL4A (Figure 4 and Table I). The main difference between these lies just downstream of the initiation site. pLΔ4-21 has the sequence **CUG** AGC UUU CAG, whereas pL4A has the sequence **CUG** AAU CCA AGC UUU CAG, and pL5G has **CUG** GGU CCA AGC UUU CAG. Thus the negative contributions of the unfavourable A in the +4 position and G in the +5 of pLΔ4-21 seem to be additive to a certain degree.

This is supported by the fact that the equivalent mutant with an AUG initiation codon, pJΔ4-21, was translated with surprisingly low efficiency and a high Z/F ratio, and indeed was the only construct of the pJ series to show such a severe decrease in utilization of the 5′-proximal initiation codon (Table I). In contrast, a deletion of only slightly greater length (pJΔ8-28 by the same nomenclature), which retains the AUGGAU sequence of the wild-type, showed the high [^{35}S]methionine incorporation of ~75% (relative to pJ′1) and a low Z/F ratio of 0.33 (data not shown), typical of all pJ constructs apart from pJΔ4-21 (Table I).

We consider it unlikely that the primary sequence beyond the +6 or +7 position exerts an influence on initiation codon recognition. In the first place, the influence of mutations at +6 was much smaller than those at +4 or +5 (Table I). An additional reasoning is based on the properties of constructs in which the A/CUGGAUC of pJ′1 or pLC1 was retained, but the following 232 nt of NS coding sequences were replaced by a 364 nt segment of coding sequences from human β-globin cDNA to generate pJH and pLCH, respectively (Figure 1). As this substitution makes no change in the number of in-frame methionine codons, [^{35}S]methionine incorporation can be used to compare the translation efficiency of pJH and pJ′1 or pLCH and pLC1 mRNAs, but the Z/F ratio cannot be used because the substitution deletes the upstream of the two AUG codons from which Z is synthesized, while downgrading the context of the other. pJH mRNA was translated with an efficiency very similar to that of pJ′1 mRNA, whilst pLCH was, if anything, slightly more efficient than pLC1 (Figure 3 and Table I). Since the substitution of the globin sequences radically altered the primary sequence downstream of the +7 position, this result argues that the influence of coding sequences on the efficiency of recognition of an initiation codon is confined to at most the first four downstream positions. Although the introduced β-globin cDNA sequence starts from the +5 position of the globin open reading frame, and thus might include any downstream determinants for recognition of the initiation site of β-globin mRNA, it was inserted into pJH/pLCH one codon further downstream from the initiation codon than its position in globin mRNA. Thus if any primary sequence motif in the globin

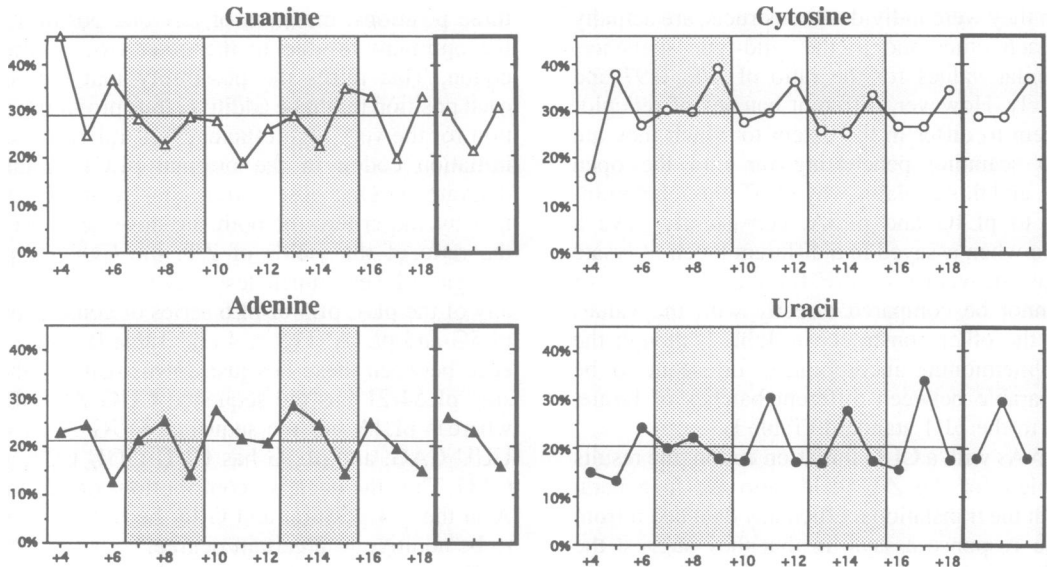

Fig. 6. Analysis of the frequency of occurrence of different nucleotides in sequences downstream of the initiation codon in the ~700 vertebrate mRNAs analysed for upstream context features by Kozak (1987a). The frequency of occurrence of each nucleotide in the first five codons downstream of the initiation codon (i.e. positions +4 to +18) is shown as a per cent occupancy of each position. The codon triplets are highlighted by alternate shading, and the horizontal line shows the average of the frequency of each type of residue over all 15 positions of all the compiled sequences. For each type of residue, the extreme right hand panel highlighted by thickened outlining shows the frequency of occurrence in each position of the average reading frame codon, taken from the data of Maruyama *et al.* (1986) weighted as described in the text.

mRNA tract were exerting the same type of influence on initiation site recognition in pJH/pLCH mRNAs as in β-globin mRNA, such effects could not be dependent on exact positioning relative to the initiation codon, whereas all primary sequence context effects described so far are critically dependent on position (Kozak, 1986, 1987b, 1989a). Therefore it is most unlikely that the globin mRNA sequences could have contributed to the primary sequence context determinants of initiation site recognition. On the other hand this does not eliminate the type of contribution by secondary structure motifs appropriately located in the coding region which Kozak (1989b, 1990) has shown can influence the recognition of 'weak' initiation sites.

Analysis of sequences downstream of the initiation sites of vertebrate mRNAs

The consensus upstream context of vertebrate initiation sites was derived by Kozak (1987a) from an analysis of 699 such sequences, and has since been reassessed with a much larger database by Cavener and Ray (1991). The analysis by Kozak included the +4 position (46% G), but did not extend further downstream, whilst Cavener and Ray (1991) present data for the +5 position. In order to see whether any patterns might emerge indicative of an extended downstream consensus, we have analysed the sequences of the same 699 mRNAs from the +4 to +18 positions (codons 2–6 counting the AUG as the first codon). The results of this analysis are depicted in graphical form in Figure 6, which shows the frequency of occurrence of each type of nucleotide residue in each position of these five downstream codons.

In order to be able to attribute any significance to these patterns, it was necessary to compare the data with that of the average codon found in vertebrate coding sequences. This was obtained from the codon usage patterns given in Table II of Maruyama *et al.* (1986), but because codon usage patterns seem to differ slightly between different vertebrates, a weighted average codon usage was calculated from their compiled data. The weighting was made according to the breakdown of different species in the list of sequences compiled by Kozak (1987a), used by her to determine the upstream consensus sequence and by us to search for a downstream consensus: 39.9% human, 2% hamster, 15.6% mouse, 19% rat, 5.7% bovine and other ruminants, 2% rabbit, 10.4% chick, 1.7% fish and 3.7% frog. The frequency of each type of nucleotide in each position of this weighted average codon is shown in the extreme right hand panels in Figure 6.

It is clear from Figure 6 that codons 4–6 (the last three codons in the analysis) conform rather closely to the weighted average codon, whilst the third codon is slightly deviant. However, what is most striking is that the second codon, the codon immediately following the initiation codon, deviates strongly from the weighted average codon: U is unusually rare in the first two positions of this codon, and is rather more frequent in the wobble position than expected; C occurs at very high frequency in the second position, but is atypically low in the wobble position and in the first position; the distribution of A is somewhat over-represented in the +5 position and under-represented at +4; the distribution of G is qualitatively normal, but its frequency in the +4 position is unusually high, while it is slightly higher than average at the +6 position. Our finding of a bias in the second codon is at variance with the conclusions of Cavener and Ray (1991) who claimed no bias in positions +5 and +6. However, this conclusion seems to be based on a comparison of the frequency of occupancy of positions +5 and +6 relative to positions −10 to −20, rather than with the coding sequences, as in our analysis.

As Cigan and Donahue (1987) pointed out in an earlier analysis of this type restricted to yeast mRNAs, the

Table II. Frequency of occurrence of synonymous codons in the position next to the initiating methionine codon relative to the average codon usage frequency within the reading frame

	Over-represented at the 2nd codon		Under-represented at the 2nd codon	
Arg	AGG (23)	2.42	CGC (3)	3.16
	CGU (5)	1.24	AGA (3)	3.10
			CGA (3)	1.32
			CGG (6)	1.11
Leu	UUG (6)	1.36	CUA (1)	2.35
	CUG (20)	1.12	UUA (1)	1.90
	CUU (5)	1.08	CUC (7)	1.29
Ser	UCG (10)	2.23	UCA (4)	2.26
	UCU (26)	1.75	UCC (14)	1.48
	AGU (10)	1.07	AGC (16)	1.36
Ala	GCG (23)	1.99	GCC (45)	1.36
	GCU (45)	1.08		
	GCA (27)	1.07		
Lys	AAG (40)	1.19	AAA (12)	1.53
Gly	GGG (26)	1.72	GGA (12)	1.44
			GGU (9)	1.38
			GGC (26)	1.09
Pro	CCG (9)	2.15	CCA (5)	2.00
	CCC (16)	1.09	CCU (10)	1.13
Val	GUG (23)	1.44	GUC (3)	2.94
	GUU (6)	1.12	GUA (1)	2.87
Thr	ACG (5)	1.77	ACC (8)	1.56
	ACA (9)	1.28		
	ACU (7)	1.07		
Glu	GAG (32)	1.33	GAA (7)	2.13
Asp	GAU (18)	1.28	GAC (16)	1.25

The relative frequency of occurrence of each synonymous codon was calculated (i) when that amino acid occurs next to the initiating methionine residue, and (ii) for the weighted average codon within the reading frame. The larger number was divided by the smaller to give the index of over- or under-representation of each codon when it occurs next to the AUG codon. Numbers in parentheses give the actual number of occurrences next to the AUG codon.

distribution of nucleotides in the +4 to +6 positions may reflect not so much any context effect for initiation site recognition, but the requirement for a particular type of amino acid residue at the N-terminus of the protein. The nature of the amino acid residue next to the initiating methionine is the determinant of whether the methionine is removed by amino-peptidase activity (Sherman *et al.*, 1985, 1993). In addition, the nature of the N-terminal amino acid is an important determinant of the rate of turnover of the protein (Bachmair *et al.*, 1986). The high frequency of G in the +4 position and C in the +5 is largely a reflection of the fact the codons found most frequently at this site are alanine codons, but we do not know whether this pattern has evolved because there are advantages in having alanine as the N-terminal amino acid or for reasons of maintaining a context moderately favourable to the recognition of the neighbouring initiation codon.

However, even if the bias in nucleotide distributions in the +4 and +5 positions is mainly due to the dictates of the requirement for a particular species of amino acid, for many amino acids this constraint is irrelevant to the +6, or wobble, position. Nevertheless, the nucleotide distribution in this +6 position seems atypical, with U and G at a higher frequency, and C and A at a lower frequency than in the wobble position of downstream codons. To examine this in more detail, we compared the frequency of usage of synonymous codons for each type of amino acid residue when that residue is adjacent to the initiating methionine residue, and the frequency in the weighted average codon usage. This allowed us to determine which codons were over-represented in the codon 2 position and which were under-represented relative to the typical codon usage pattern. The analysis, presented in Table II, was limited to those amino acid residues which occur adjacent to the N-terminal methionine in more than 28 of the mRNAs in the list compiled by Kozak (1987a). With one exception (CGG), all codons with G in the wobble position occur at higher frequency at the second codon than elsewhere, and this includes some codons with strong over-representation (AGG, UCG, GCG and CCG). Of the 10 codons in Table II with U in the wobble position, eight are over-represented at the second codon, though by only a small margin in most cases except UCU. Codons with A or C in the wobble position are generally under-represented at the second codon, some of them strongly under-represented, but with three exceptions (GCA, CCC and ACA) which are slightly over-represented. Thus the preference for the wobble base in the second codon would appear to be G > U > A = C.

In so far as 46% of the ~700 sequences analysed have G at +4, and a total of 62% have A or C at +5 (24% A and 38% C), it is clear that in a large number of mRNAs the second codon has at least one of the positive context features which we have identified here. However, the greater frequency of G in the wobble position rather than U would not have been predicted from our observations, and may perhaps reflect some other constraints. One feature of decoding that is unique to the translation of the second codon, is that it is carried out by the first tRNA to enter the A-site, circumstances in which the E-site (exit site) is almost certainly unoccupied (Rodnina and Wintermeyer, 1992).

Discussion

It is clear that over the lower part of the range of Mg^{2+} concentrations used in this work, the results conform closely with the predictions of the scanning ribosome model in so far as any change which reduced the efficiency of utilization of the 5′-proximal initiation site, for example the substitution of a CUG codon and/or a reduction in the Mg^{2+} concentration, was accompanied by an increased yield of Z synthesized by ribosomes that had scanned past the 5′-proximal initiation site. It is interesting that this relationship did not hold at the top end of the [Mg^{2+}] range studied. Although the translation of any pL mRNA under these conditions gave a lower yield of F than obtained with any pJ mRNA, this deficit in full-length product was not matched by a corresponding increase in the synthesis of X, Y or Z. Likewise, the yield of full-length product with pL4G at high Mg^{2+} concentrations was greater than with any other mRNA in the pL4 series, yet the yield of Z was very similar in all four cases. At high Mg^{2+} it appears that the majority of the ribosomes which by-pass the CUG codon do not reach the downstream initiation sites. This could be the result of high Mg^{2+} inhibiting the actual scanning migration activity of

the ribosomes, or the consequence of Mg^{2+} stabilizing secondary structure to the point where it blocks scanning (Kozak, 1989a). A third type of explanation could be invoked if (i) high Mg^{2+} promotes initiation at AUG codons with a poor context such as the out-of-frame AUG-2, AUG-3 and AUG-4 of the NS mRNA, and (ii) translation of the short open reading frames following these AUGs somehow interferes with recognition of, and initiation at, the downstream AUG-5 and AUG-6, the sites at which the synthesis of Z is initiated.

Our data are entirely consistent with the results of Kozak in relation to her observation that reducing the Mg^{2+} concentration in wheat germ extracts was more inhibitory to the recognition of weak initiation sites (either non-AUG codons in a favourable context, or an AUG codon in an unfavourable context) than strong (Kozak, 1989b). Reduced Mg^{2+} therefore increased the discrimination between 'weak' and 'strong' sites, as in our hands. In her experiments, this increased discrimination was achieved at the price of a lower incorporation efficiency at the lower Mg^{2+} concentration, a reduction in efficiency that could be largely compensated by adding 0.34 mM spermidine (Kozak, 1989b). Using the reticulocyte lysate system at low Mg^{2+} (0.5 mM added EDTA), we have tested whether the addition of spermidine affected overall translation efficiency or the relative utilization of different initiation sites in the pL4 mRNAs. We found no effect on the efficiency of recognition of CUG as an initiation codon, and a slight inhibition of overall incorporation (data not shown). An important procedural difference between the two cell-free systems is that the wheat-germ system is normally gel-filtered prior to use, whereas the reticulocyte lysate is not and already contains, on average, 0.45 mM spermidine, which is close to the optimum (Jackson *et al.*, 1983) and thus probably explains the lack of response to additional spermidine.

Given that our translation assays behave in a way that conforms well to the established principles of the scanning ribosome mechanism, it follows that the ratio of the yield of full-length product to the yield of small products initiated at downstream sites provides a good indicator of the relative efficiency of initiation at the 5′-proximal site, internally controlled against variabilities such as the precise concentration of the mRNA or its quality. This is similar to the approach previously used by Kozak (1989b, 1990, 1991) except that in her constructs the two in-frame initiation codons were much closer together. Moreover, the conclusions drawn from this type of assay are entirely consistent with the results of the less sensitive RNA dose–response assay. These experiments have revealed several novel features concerning the influence of the downstream context on the efficiency of recognition of the initiation codon. (i) The identity of the nucleotide in the +5 position is at least as important as the +4 position, whilst the nature of the residue in the +6 position has a lesser, though not negligible influence. (ii) With regard to the +4 position, only G exerts a positive effect on the frequency of initiation, and the fact that A occurs significantly more frequently (25%) in this position in vertebrate mRNAs than either U or C (Kozak, 1987a; Cavener and Ray, 1991) is unlikely to have any functional implications for initiation codon recognition. (iii) In the +5 position, A exerts a strong positive influence on recognition of the initiation codon, and C a somewhat smaller beneficial effect. (iv) In the +6 position, only U is beneficial to initiation codon recognition. (v) There appears to be an additive effect of the positive influence of favourable residues at each position from +4 to +6 (or, there is an additive effect of the negative influence of unfavourable residues at each of these positions).

Although the influence of the nucleotide residues in the +4 to +6 positions was more pronounced when the initiation codon was CUG rather than AUG, we believe that our results have a general significance and are not particular to cases where initiation is at a CUG codon. This supposition is based on the fact that utilization of an AUG initiation codon was slightly favoured if the +4 position was occupied by a G residue and the +5 position by an A, and by the fact that pJΔ4-21, in which the +4 to +6 positions were all unfavourable (AGC), was translated with unusually low efficiency for any type of pJ mRNA derivative. The reason why the influence of single site mutations had a smaller influence on the recognition of the AUG in the pJ series than the CUG in the pL series is almost certainly because the upstream context of the AUG is highly favourable to recognition of this codon by scanning ribosomes (Dasso *et al.*, 1990). If an AUG initiation codon is in a sufficiently favourable context to be recognized by >90% of the scanning ribosomes, then no matter how many additional favourable context features are introduced, the efficiency of translation of the mRNA cannot be increased by >10%. For this reason, the influence of additional context features on initiation frequency is most easily detected either if the initiation codon is in a weak context background, with pyrimidines at both the –3 and +4 positions as exploited by Kozak (1987b) to show the positive influence of appropriately phased upstream GCC motifs, or if the initiation codon is itself rather inefficiently recognized because a non-AUG codon has been substituted.

Although Kozak (1990) has shown that downstream secondary structure elements of sufficient stability and in the appropriate location can promote initiation at non-AUG codons or at AUG codons in a weak context background, we do not believe that this plays a major role in the mRNAs studied here. Computer predictions and biochemical probing experiments indicate that the coding region of the NS mRNA, as in pJ′1/pLC1 and pJ′2/pLC2 mRNAs, has a very stable hairpin loop starting at +39 (on a numbering system where the A/C of the A/CUG initiation codon is +1). However, this is much greater than the optimal distance for a positive influence of downstream secondary structure as determined by Kozak (1990): the strongest influence was when the spacing between the initiation codon and the base of the stem was 14 nt, a modest effect was seen at 8 nt, and with a spacing of either 2 or 32 nt the hairpin exerted essentially no effect. In our constructs the distance between the initiation codon and the base of the predicted hairpin loop was reduced to 27 nt in the pJ4, pL4, pJ5, pL5 and pL6 series of constructs with no more than a modest decrease in initiation efficiency if the GAU codon immediately downstream of the initiation codon was preserved. In pJΔ4-21 and pJΔ8-28, the separation between the initiation codon and the predicted hairpin loop was reduced further to 21 or 18 nt, respectively, with very little effect on the

efficiency of translation of pJΔ8-28 mRNA which retains a GAU codon following the initiation codon, yet a severe reduction in the case of pJΔ4-21, which has an AGC codon in this position. In view of the lack of any consistent position effect or any consistent correlation between initiation efficiency and the size of the deletions in the coding region, contrasted with the consistently high efficiency whenever the GAU codon was retained, we conclude that the downstream secondary structure motif in the NS mRNA coding region can exert no more than a very minor influence on the efficiency of recognition of the 5′-proximal initiation site in the mutants we have studied. The major determinant is clearly the identity of the downstream codon, which is optimally GAU, with GCU as a somewhat less effective alternative.

This conclusion is strikingly vindicated by the accompanying paper of Böck and Kolakofsky (1994), who have independently shown that a downstream GAU codon, but not a GUA codon, allows efficient initiation at a GUG codon in human parainfluenza virus-1 P/C mRNA, at an ACG codon in Sendai virus P/C mRNA, and at a CUG codon in c-*myc* mRNA. This provides overwhelming support to our conclusion that a GAU codon downstream of any functional initiation codon exerts a very strong positive influence on the recognition of the initiation codon, irrespective of other primary sequence and secondary structure features around the initiation codon, and irrespective of the actual identity of the initiation codon. Moreover, the fact that the ACG codon in Sendai virus P/C mRNA is utilized with only slightly less efficiency in the wild-type mRNA, where it is followed by GCU, than when a GAU codon is substituted, provides a vindication for our conclusion that C in the +5 position exerts a positive influence on initiation site recognition, albeit less than the influence of an A. Thus the most favourable context for initiation codon recognition by scanning ribosomes can now be redefined as

$$\text{gccgcc}\ \underset{G}{\mathbf{A}}\ \text{cc(AUG)}\mathbf{G}\ \underset{C}{\mathbf{A}}\ \text{u}$$

where the residues in bold upper case exert the strongest positive influence, those in upper case italics a somewhat smaller beneficial effect, while the residues in lower case have a yet smaller but nevertheless still positive influence. The significance of this redefinition of the optimal context is that it should allow a more reliable prediction of those instances in which non-AUG codons are used as functional initiation codons, or 5′-proximal AUG codons with unfavourable pyrimidine residues at the –3 and/or +4 positions are used with much higher efficiency than would have been predicted according to the previous criteria.

Materials and methods

Plasmid constructs

The starting constructs used in this work were pJ′1 and pLC1 described previously (Dasso and Jackson, 1989b). To destroy the 5′-proximal *Hin*dIII site, pGEM-2 was digested with *Hin*dIII, the overhanging ends in-filled using the Klenow fragment of DNA polymerase I, and religated to give pGEM-2H. pGEM-2H, pJ′1 and pLC1 were each digested with both *Pst*I and *Eco*RI, and the fragments gel-purified. The large fragment obtained from pGEM-2H was ligated with the small fragment from pJ′1 or pLC1, to give pJ′2 and pLC2 respectively (Figure 1). pLC2 was digested with *Bst*EII and *Hin*dIII, then dephosphorylated and the large fragment gel-purified before ligation with appropriate oligonucleotides to create the pJ4, pL4, pJ5, pL5 and pL6 series of plasmids (Figure 1). The oligonucleotides were synthesized with four-fold redundancy in the relevant position.

pLΔ4-21 was made from pLC1 (and pJΔ4-21 from pJ′1) by digestion with *Bam*HI followed by treatment with mung bean nuclease to remove the 5′-overhang. Then the DNA was cut at the unique *Nco*I site in the coding sequence, and the large fragment gel-purified. Another sample of pLC1 DNA was digested with *Hin*dIII and the overhanging ends in-filled using the Klenow fragment of DNA polymerase I, before digestion with *Nco*I and purification of the small (218 bp) fragment. This small fragment was ligated with the large fragment derived from pLC1 as described above, generating pLΔ4-21 (Figure 1).

pJ0 was made by digestion of pJ′1 with *Hin*dIII and religating the large fragment, thus deleting a tract corresponding to most of the 5′-untranslated region and the first 21 nt of coding sequences, including the initiation codon (Figure 1).

pJH and pLCH were made by digestion of pJ′1 or pLC1, respectively, with *Bam*HI and *Nco*I, and the overhanging ends in-filled before gel purification of the large fragment. This was ligated with a fragment derived from a human β-globin cDNA clone by digestion with *Apa*LI and *Eco*RI, followed by in-filling the overhanging ends, and gel purification of the small (364 bp) fragment.

Plasmids were propagated by standard methods (Sambrook *et al.*, 1989) using *E.coli* TG1. Plasmid DNA was purified by caesium chloride gradient centrifugation, and the sequence of the ~150 nucleotides downstream of the bacteriophage T7 promoter was verified as described previously (Dasso and Jackson, 1989b; Dasso *et al.*, 1990).

Transcription and translation reactions

All plasmids were linearized by digestion with *Eco*RI prior to transcription, which was carried out using the methods for the production of capped RNA described previously (Dasso and Jackson, 1989b; Dasso *et al.*, 1990), with the modification that the RNA solution was passed through a Sephadex G-50 spin column before the final ethanol precipitation step. With this protocol, capping efficiency is estimated to be at least 70% (Dasso and Jackson, 1989b). Translation assays in the micrococcal nuclease treated rabbit reticulocyte lysate, and the analysis of translation products by gel electrophoresis were carried out essentially as described previously (Dasso and Jackson, 1989b; Dasso *et al.*, 1990), except that the final concentration of [^{35}S]methionine (Amersham International; 1200–1500 Ci/mmol) was 0.17 mCi/ml. Determination of the RNA dose–response was carried out using added KCl at 90 mM and added $MgCl_2$ at 0.45 mM, with RNA concentrations up to at least 50 μg/ml. For the Mg^{2+} optimization experiments, the RNA concentration was 13 μg/ml, the added KCl 90 mM, and the Mg^{2+} concentration ranged from 1.5 mM added $MgCl_2$ down to 1.2 mM added EDTA (equivalent to 1.2 mM Mg^{2+} subtracted). In order to guarantee a constant mRNA concentration in all assays, the RNA was added to the master reaction mix, before this was divided into aliquots to which the appropriate amounts of $MgCl_2$ or EDTA were added.

Dried gels were exposed to Hyperfilm β-max (Amersham International), and the developed films were scanned using a Transidyne 2955 Scanning Densitometer. Different exposures of each film were scanned to ensure that data were taken only from exposures within the linear response range of the film.

Acknowledgements

We thank Michael Howell, Tim Hunt, Ann Kaminski and Andy Borman for advice; Michael Howell for the construction of pJH and pLCH; Franco Carlotti for the construction of pJ0, pJΔ4-21 and pLΔ4-21; Tim Hunt for the gift of reticulocyte lysate; and Lucy Johnson for technical assistance. This work was supported by grants from the Medical Research Council and the Wellcome Trust. S.G. was supported by a studentship from the Science and Engineering Research Council and a scholarship from Boehringer Ingelheim Fonds.

References

Acland,P., Dixon,M., Peters,G. and Dickson,C. (1990) *Nature*, **343**, 662–665.

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *Science*, **234**, 179–186.

Becerra,S.P., Rose,J.A., Hardy,M., Baroudy,B.M. and Anderson,C.W. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 7919–7923.

Böck,R. and Kolakofsky,D. (1994) *EMBO J.*, **13**, 3608–3617.
Boeck,R., Curran,J., Matsuoka,Y., Compans,R. and Kolakofsky,D. (1992) *J. Virol.*, **66**, 1765–1768.
Cavener,D.R. and Ray,S.C. (1991) *Nucleic Acids Res.*, **19**, 3185–3192.
Cigan,A.M. and Donahue,T.F. (1987) *Gene*, **59**, 1–18.
Curran,J. and Kolakofsky,D. (1988) *EMBO J.*, **7**, 245–252.
Dasso,M.C. and Jackson,R.J. (1989a) *Nucleic Acids Res.*, **17**, 3129–3144.
Dasso,M.C. and Jackson,R.J. (1989b) *Nucleic Acids Res.*, **17**, 6485–6497.
Dasso,M.C., Milburn,S.C., Hershey,J.W.B. and Jackson,R.J. (1990) *Eur. J. Biochem.*, **187**, 361–371.
Florkiewicz,R.Z. and Sommer,A. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 3978–3981.
Gupta,K.C. and Patwardhan,S. (1988) *J. Biol. Chem.*, **263**, 8553–8556.
Hann,S.R., King,M.W., Bentley,D.L., Anderson,C.W. and Eisenman,R.E. (1988) *Cell*, **52**, 185–195.
Jackson,R.J., Campbell,E.A., Herbert,P. and Hunt,T. (1983) *Eur. J. Biochem.*, **131**, 289–301.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Kozak,M. (1986) *Cell*, **44**, 283–292.
Kozak,M. (1987a) *Nucleic Acids Res.*, **15**, 8125–8148.
Kozak,M. (1987b) *J. Mol. Biol.*, **196**, 947–950.
Kozak,M. (1989a) *J. Cell Biol.*, **108**, 229–241.
Kozak,M. (1989b) *Mol. Cell. Biol.*, **9**, 5073–5080.
Kozak,M. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8301–8305.
Kozak,M. (1991) *Gene Expr.*, **1**, 111–115.
Maruyama,T., Gojobori,T., Aota,S. and Ikemura,T. (1986) *Nucleic Acids Res.*, **14**, r151–r197.
Patwardhan,S. and Gupta,K.C. (1988) *J. Biol. Chem.*, **263**, 4907–4913.
Peabody,D.S. (1987) *J. Biol. Chem.*, **262**, 11847–11851.
Peabody,D.S. (1989) *J. Biol. Chem.*, **264**, 5031–5035.
Prats,A.C., De Billy,G., Wang,P. and Darlix,J.-L. (1989a) *J. Mol. Biol.*, **205**, 363–372.
Prats,H., Kaghad,M., Prats,A.C., Klagsbrun,M., Lelia,J.M., Liauzun,P., Chalon,P., Tauber,J.P., Amalric,F., Smith,J.A. and Caput,D. (1989b) *Proc. Natl Acad. Sci. USA*, **86**, 1836–1840.
Rodnina,M.V. and Wintermeyer,W. (1992) *J. Mol. Biol.*, **228**, 450–459.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sherman,F., Stewart,J.W. and Tsunasawa,S. (1985) *Bioessays*, **3**, 27–31.
Sherman,F., Moerschell,R.P., Tsunasawa,S., Sternglanz,R. and Dumont,M.E. (1993) In Ilan,J. (ed.), *Translational Regulation of Gene Expression 2*. Plenum Press, New York, pp. 117–141.

Received on March 22, 1994; revised on April 26, 1994