Efficient initiation of mammalian mRNA translation at a CUG codon

Mary C.Dasso+ and Richard J.Jackson*

Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK

Received June 23, 1989; Revised and Accepted July 21, 1989

ABSTRACT

Nucleotide substitutions were made at the initiation codon of an influenza virus NS cDNA clone in a vector carrying the bacteriophage T7 promoter. When capped mRNA transcripts of these constructs were translated in the rabbit reticulocyte lysate, a change in the initiation codon from ...AUAAUGG... to ...AUACUGG... reduced the *in vitro* translational efficiency by only 50−60%, and resulted in only a small increase in the yield of short products presumed to be initiated at downstream sites. Synthesis of the full-length product was initiated exclusively at the mutated codon, with negligible use either of in-frame upstream CUG or GUG codons, or of an in-frame downstream GUG codon. We conclude that CUG has the potential to function as an efficient initiation codon in mammalian systems, at least in certain contexts.

INTRODUCTION

Initiation of eukaryotic mRNA translation has generally been thought to occur exclusively at AUG codons, by a process in which the 40 S ribosomal subunit scans from the 5′ end of the mRNA and initiates at the most 5′-proximal AUG codon in a favourable sequence context (1,2). If the context of the 5′-proximal AUG codon is poor, the 40 S subunit may by-pass that codon and continue scanning until it contacts an AUG in a better context (2,3). As part of our investigations into factors that may influence scanning and the fidelity of initiation *in vitro*, we constructed an influenza virus NS cDNA clone in which the ATG corresponding to the initiation codon was replaced by a CTG, in order to obtain what we imagined would be a non-functional mRNA for use in competition studies. Genetic studies had shown that CUG was effectively non-functional in yeast (4,5), while the failure of a CUG codon to act as a barrier to the recognition of the downstream AUG initiation codon of preproinsulin mRNA (2) implied that it was also non-functional in mammalian cells. However, ACG had been shown to be the initiation codon used for the synthesis of one of the three capsid proteins encoded by adeno-associated virus RNA (6), and more recently was found to be the initiation codon for synthesis of the C′ protein from the Sendai virus P/C mRNA (7,8). Whilst our work was in progress, two reports touched on the question of whether CUG could be used as an initiation codon, but left the issue in doubt. When the AUG codon corresponding to the initiation site of the P protein of Sendai virus was mutated to CUG, the synthesis of this protein was entirely abolished (9), but when the first two AUG codons were deleted a new product was seen in low yield that was tentatively ascribed to initiation at a downstream CUG codon (8). It has also been shown that of the two proteins translated from c-*myc* mRNA, the shorter is initiated at an AUG codon, but the longer at a non-AUG codon that was thought more likely to be a CUG

than the upstream contiguous ACG (10). However, the evidence on this issue is not unambiguous: mutation of the ACG almost entirely abolished synthesis of the longer protein (10), a result which could be interpreted either (i) as showing that the ACG was the predominant initiation site and the CUG a very minor one, or (ii) as demonstrating a context effect of the ACG sequence on initiation at the CUG codon. These doubts about the status of CUG as a viable initiation codon are now resolved by the combination of a very recently reported study of mutations in mouse dihydrofolate reductase mRNA (11), and the results described here concerning an entirely unrelated mRNA. In addition, our work highlights the fact that efficient initiation at non-AUG codons is rather an unusual event, confined to a few sites which do not appear to be unambiguously distinguishable by the context rules governing the efficiency of initiation at AUG codons (2).

MATERIALS AND METHODS

Plasmid constructs

The starting material for the constructs was a full-length cDNA clone of the influenza virus (strain A/PR/8/34) NS gene, generously provided by Drs. S.C. Inglis and V.C. Blok (Department of Pathology, Cambridge), and derived by them from pAPR801 described by Young *et al.* (12). This cDNA was inserted into the EcoR I site of pGEM-2 (Promega), and a clone, pJ12, with the correct orientation was selected. Plasmid pJ12 DNA was digested with BamH I and Sal I, which removes 32 nucleotides of polylinker and the whole 5′-untranslated region of the NS gene. The large fragment was purified and ligated with synthetic oligonucleotides to give pJ′1, in which the complete 5′-untranslated segment of the NS gene is restored unchanged except for an extra C residue inserted to create a restriction site for BstE II (Fig. 1). Digestion of pJ′1 DNA with BamH I and BstE II, followed by ligation of the large fragment with appropriate oligonucleotides generated pLC1, in which the ATG initiation codon was replaced by CTG (Fig. 1). pL1.1, in which the context of the ATG initiation codon is changed from ATAATGG in pJ′1 to TTTATGG in pL1.1, was constructed in the same way.

Plasmids were propagated by standard methods (13), without chloramphenicol amplification, using *E. coli* TG1. Plasmid DNA for use in transcription reactions was purified by caesium chloride gradient centrifugation in the presence of 0.3 mg/ml ethidium bromide. Ethidium bromide was removed from the plasmid DNA by extraction with butanol, followed by dialysis.

The sequence of the region immediately downstream of the T7 promoter of each construct (corresponding to the 5′-proximal ca. 150 nucleotides of the mRNA) was determined essentially as described by Hattori and Sakaki (14), using a primer complementary to the T7 promoter (Promega). In all cases the sequence obtained was as predicted.

Transcription reactions

Unless otherwise stated, plasmid DNAs were linearised by digestion with EcoR I. Transcription reactions (usually 0.05 ml) for the production of capped transcripts were as follows: 40 mM Tris-HCl, pH 8.0; 15 mM $MgCl_2$; 1 mM ATP, CTP, UTP; 40 μCi/ml [α-^{32}P]UTP; 0.1 mM GTP; 0.5 mM m^7GpppG (New England Biolabs); 5 mM dithiothreitol; 10 μg/ml nuclease-free bovine serum albumin (BRL); 2,000 units/ml RNasin (Boehringer); 150 μg/ml linearised template DNA; 50 μl/ml T7 RNA polymerase (gift of S. Mackie; activity ~ 40 units/μl). The reaction was incubated at 37° for 30 min, then made 1 mM in GTP, and incubation continued for a further 60 min. As described previously

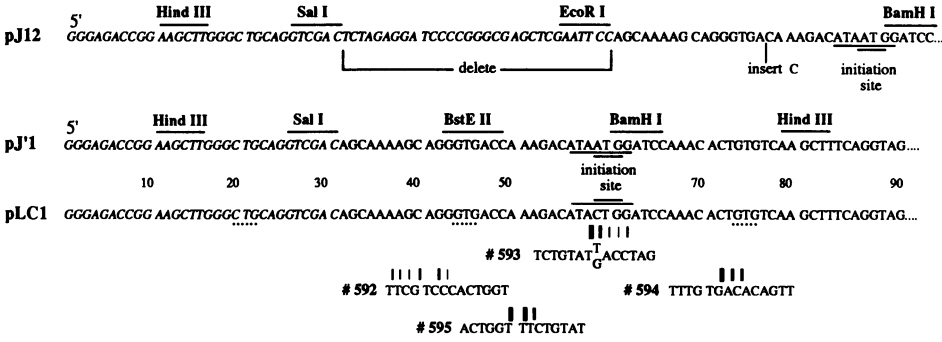

Figure 1. Nucleotide sequences of the 5'-regions of the influenza virus NS cDNA constructs. The sequences written 5'−3' correspond to the mRNA sequences, starting with the first transcribed nucleotide, with sequences derived from the pGEM-2 polylinker shown in italics. The ATG initiation codon of pJ'1 is doubly underlined, and the CTG of pLC1 doubly overlined. The other in-frame GTG and CTG codons are shown by dotted underlining. The oligodeoxynucleotides used for site-specific scission are designated by their production numbers (592−595), and their sequences are written 3'−5' to show their complementarity to the mRNA sequence. Oligonucleotide no. 593 was made with two-fold redundancy in one position so as to be complementary to both pJ'1 and pLC1 mRNA. The sites of RN'aseH scission are shown by vertical lines: the thickness of these lines is approximately proportional to the relative frequency of cleavage at each site. Note that although reasons of clarity required that the sites of scission should be shown on the oligodeoxynucleotide sequences, it is the complementary (RNA) strand that is actually cleaved.

(15), approximately two-thirds of the transcripts produced under these conditions are capped. For the production of uncapped transcripts, the m^7GpppG cap analogue was omitted, and 1 mM GTP was present from the start of the reaction.

The extent of RNA synthesis was calculated from the percentage of radioactivity incorporated from [α-^{32}P]UTP as described previously (15). RNA was isolated from the bulk of the reaction mix by extraction first with phenol, and then with chloroform. It was precipitated three times with ethanol in the presence of 2 M ammonium acetate to remove unincorporated nucleotides. The concentration of the final RNA solution was determined by counting Cerenkov radiation.

Translation assays and analysis of translation products

Rabbit reticulocyte lysates were prepared and treated with micrococcal nuclease as described previously (16). The translation assays were based on previously described methods (16) with the following differences. Nuclease-treated lysate constituted 72% by vol. of the final assay mix, and the final concentration of the added components was: 90 mM KCl, 0.45 mM $MgCl_2$, 10 mM creatine phosphate, 100 μM each unlabelled amino acid (minus methionine), 45 μg/ml calf liver tRNA (Boehringer), 50 μg/ml creatine phosphokinase, 0.45 mCi/ml [^{35}S]methionine (Amersham International; 1200−1500 Ci/mmol). Total incorporation of labelled methionine into protein was assayed as described previously (16). To analyse the products by gel electrophoresis, one volume of translation assay was mixed with an equal volume of 100 μg/ml ribonuclease A, 10 mM EDTA (pH 7.2), incubated for 10 min at room temperature, and then diluted with 18 vol. of the sample buffer described previously (17). Polyacrylamide slab gels were as described previously (17), except that the gel composition was 20% (w/v) acrylamide, 0.066% (w/v) methylenebisacrylamide. The dried gels were exposed to Hyperfilm β-max (Amersham International), or were

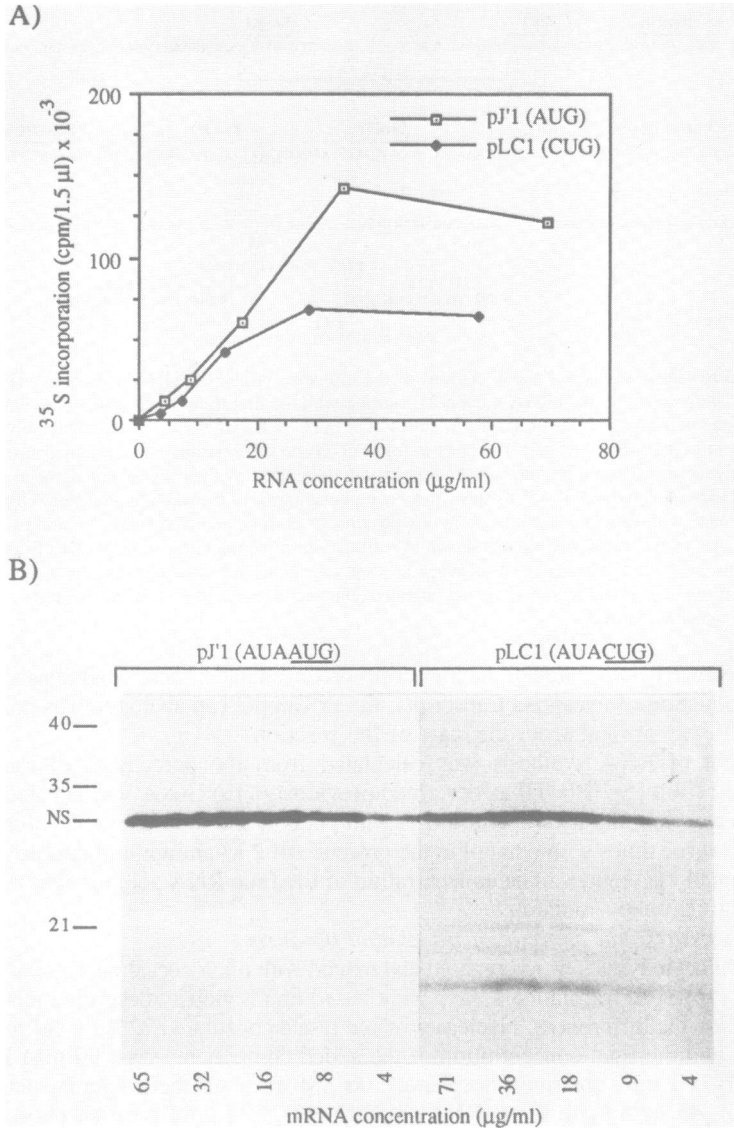

Figure 2. Comparison of the efficiency of translation of pJ'1 (AUG) and pLC1 (CUG) mRNAs. Capped pJ'1 and pLC1 mRNAs were each translated for 35 min at the final RNA concentrations shown, in two separate experiments: (A) for determination of [^{35}S]methionine incorporation, shown as c.p.m. per 1.5 µl sample for assays of pJ'1 mRNA (□———□) and pLC1 mRNA (◆———◆); (B) for analysis of translation products by gel electrophoresis, and autoradiography of the dried gel.

fluorographed using Amplify (Amersham International) and exposed to pre-flashed Fuji RX film. Where appropriate, the developed films were scanned using a Transidyne 2955 Scanning Densitometer.

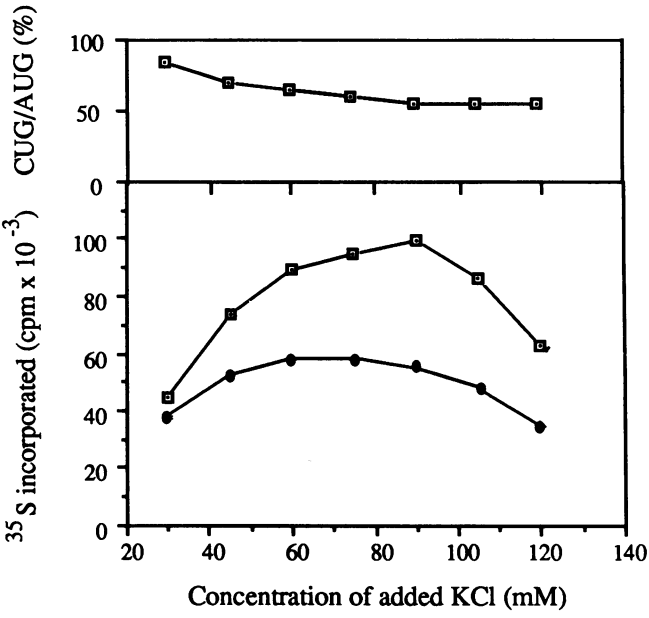

Figure 3. Potassium chloride optima for pJ'1 and pLC1 mRNA translation. Capped mRNAs were translated at 20 μg/ml under standard conditions except that the concentration of added KCl was varied. After 30 min incubation, 1.5 μl samples were taken for determination of [^{35}S]methionine incorporation, which is plotted in the lower panel for pJ'1 mRNA (□———□) and pLC1 mRNA (●———●) against the final concentration of added KCl. The upper panel plots pLC1 mRNA translation as a percentage of pJ'1 mRNA translation at each KCl concentration.

RESULTS

Mutation of the AUG initiation codon to CUG reduces translation efficiency by only 50%. The constructs used in these experiments are shown in Fig. 1. The control construct, pJ'1, consists of the entire influenza NS cDNA placed downstream of 31 nucleotides of polylinker sequence, whilst pLC1 is identical except for the substitution of a CUG codon for the AUG initiation codon. When capped transcripts of these two plasmids were translated in the rabbit reticulocyte lysate system, we were surprised to find that pLC1 (CUG) mRNA was translated at almost 50% of the efficiency of pJ'1 (AUG) mRNA throughout the range of RNA concentrations tested (Fig. 2). Although both RNA species yield one main translation product, some fainter bands corresponding to labelled proteins of lower molecular weight are visible on the autoradiogram (Fig. 2). These are relatively more abundant in the translation of pLC1 (CUG) RNA than in the translation of pJ'1 (AUG) RNA. The same bands were of intermediate intensity in the translation of pL1.1 RNA in which the local context of the AUG codon is changed from ...AUAAUGG... to ...UUUAUGG..., whilst they were the major products with an mRNA construct in which the AUG (CUG) initiation codon had been deleted (data not shown). They are therefore most probably products of initiation at downstream sites by ribosomes that have scanned past the principal initiation codon. An increase in the yield of such products when the AUG codon is replaced by a CUG would indeed be predicted by the scanning ribosome model (2). In quantitative

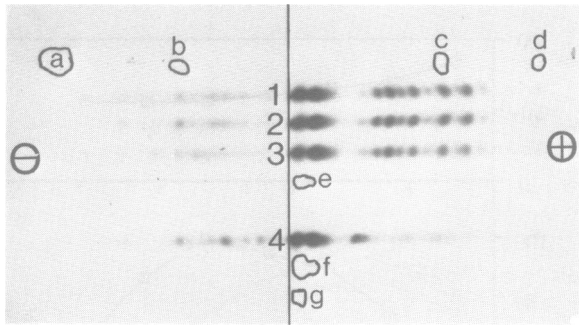

Figure 4. Labelled peptides synthesised in the presence of elongation inhibitors. Unlabelled capped mRNAs were translated at a concentration of approx. 30 μg/ml under standard conditions except for the presence of 0.1 mM sparsomycin and 0.4 mM anisomycin. After 4 min incubation, the reactions were diluted with ice-cold low salt buffer and centrifuged on sucrose gradients for 2 h at 50,000 rev./min (25). RNA was extracted with phenol from the fractions of each gradient that contained the monomeric 80S ribosomes, then incubated with 1% trimethylamine for 1 h at 37° to hydrolyse the peptidyl-tRNA linkage (18), followed by incubation with 2 M H_2O_2 for 1 h to oxidise the methionine residues. The labelled peptides were separated by electrophoresis at pH 6.5 on thin-layer cellulose plates, together with unlabelled markers. The positions of the markers were detected using fluorescamine staining: (a) lysine, (b) histidine, (c) met-glu, (d) DNP-asp, (e) met-thr, (f) met-leu, and (g) methionine. The labelled peptides were detected by autoradiography. The tracks correspond to the following mRNAs assayed: (1) pL1.1 mRNA, (2) pJ'1 mRNA, (3) pLC1 mRNA, and (4) capped human β-globin mRNA.

terms, however, the model would predict a much higher yield of these shorter products, and a far lower yield of the full-length product.

The size of the most abundant of these small products is consistent with initiation at either the 5th or 6th AUG codons, which are only 9 nucleotide residues apart. As both of these have a good local sequence context, ...AAAAUGA... and ACCAUGG... respectively, ribosomes that scan as far as these sites would be expected to use them for initiation with high efficiency. The 2nd, 3rd and 4th AUG codons are in a different reading frame and their use would give only very short products. Since all three have a rather poor context, most scanning ribosomes would be expected to by-pass them, and continue scanning to AUG-5 or AUG-6. Moreover, the few ribosomes that may initiate at these out-of-frame sites would, according to the current scanning ribosome model (2), have the potential to reach AUG-5 and AUG-6 by resuming scanning after translation of the short open-reading frames. Therefore the scanning ribosome model readily accounts for the identity of the most abundant of the short products. However, when the 5'-proximal AUG codon of pJ'1 is mutated to CUG, or deleted entirely, it is clear that the decrease in yield of the full-length product was not matched by a quantitatively equivalent increase in the yield of the short products. Thus, not all the ribosomes that by-pass the principal initiation site actually reach AUG-5/6. Our results suggest that about one third of the ribosomes which scan past the authentic initiation site succeed in initiating at downstream in-frame sites.

The assays shown in Fig. 2 were carried out at a final concentration of 90 mM added KCl, which is the optimum for translation of the control (pJ'1) mRNA. However, the mutant mRNA with the CUG substitution (pLC1) shows a lower salt optimum of approximately 60 mM added KCl (Fig. 3). At salt concentrations below 90 mM the translational efficiency of pLC1 mRNA is therefore closer to that of the control RNA.

In contrast, raising the salt concentration above 90 mM did not reduce the relative translation efficiency of pLC1 mRNA below the level of about 50% observed at 90 mM KCl (Fig. 3).

We also tested the effect of varying Mg^{2+} concentration, as it has been reported that the synthesis of the Sendai virus C′ protein (initiated at an ACG codon), if not actually stimulated by raising the Mg^{2+} concentration, is certainly more resistant to the inhibitory effects of supra-optimal Mg^{2+} than is the synthesis of the P and C proteins from AUG codons on the same mRNA (18). In addition, initiation at an ACG codon of a laboratory-constructed mutant of mouse dihydrofolate reductase was said to be promoted by Mg^{2+} concentrations higher than the optimum for initiation on the parent (AUG) mRNA (19). In our hands, the optimum Mg^{2+} concentrations for translation of pLC1 mRNA and pJ′1 mRNA were not significantly or reproducibly different. On the other hand, the results did show an increase in the relative efficiency of pLC1 mRNA translation with increasing Mg^{2+} concentration, because supra-optimal concentrations were more inhibitory to the translation of pJ′1 mRNA than pLC1 mRNA, whilst sub-optimal levels inhibited pLC1 mRNA translation more than pJ′1 mRNA (F.P. Carlotti, M.T. Howell and R.J.J., unpublished observations). In addition, sub-optimal Mg^{2+} promoted the synthesis of the shorter products which are thought to arise from initiation at downstream sites by ribosomes that have scanned past the normal initiation site. This effect was seen with both pJ′1 mRNA and pLC1 mRNA, but was more accentuated in the latter case.

Whilst these experiments show that the efficiency of utilisation of the CUG codon in pLC1 mRNA, relative to initiation at the AUG of pJ′1 mRNA, can be subtly influenced by the actual assay conditions, it would require a very drastic departure from the conditions optimum for translation of the control mRNA in order to reduce the relative efficiency of CUG utilisation by as much as 2-fold, i.e. to 25% of the control (pJ′1 mRNA). It is important to emphasise that the control (pJ′1) mRNA is among the more efficiently translated mRNAs we have encountered, being comparable with tobacco mosaic virus RNA and encephalomyocarditis virus RNA in this respect. It follows that even if our assay conditions may have slightly overestimated the efficiency of translation of the mutant pLC1 mRNA (CUG), it must still be considered a fairly efficient mRNA, at least as good as many mRNAs with AUG initiation codons.

Translation of pLC1 mRNA is initiated at the new CUG codon

Although pLC1 mRNA was translated to give a product of apparently the same size as pJ′1, it was important to establish that initiation occurred at the newly introduced CUG codon, rather than at any other non-AUG potential initiation codon situated nearby (Fig. 1). Starting with the most 5′-proximal alternative, there is a CUG codon in a moderately favourable local (Kozak) context (...GGGCUGC...) which would code for a protein 13 amino acids longer than the product translated from the first AUG codon of pJ′1, but the resolution of our gel system is sufficient to eliminate the possibility that this was used as an initiation site. There is also a GUG codon in a good local context (...AGGGUGA...) encoding a protein with 5 extra amino acids, and a GUG in a fairly good (...ACUGUGU...) local context encoding a product that is shorter by 5 amino acid residues (Fig. 1). It is not clear that our gel system would be able to distinguish products initiated at either of these GUG codons.

Attempts to determine directly the sequence of the products of pJ′1 and pLC1 mRNA translation were not successful, presumably because the *N*-terminus of the protein was blocked. Examination of the short peptides produced by incubation of the lysate with mRNA

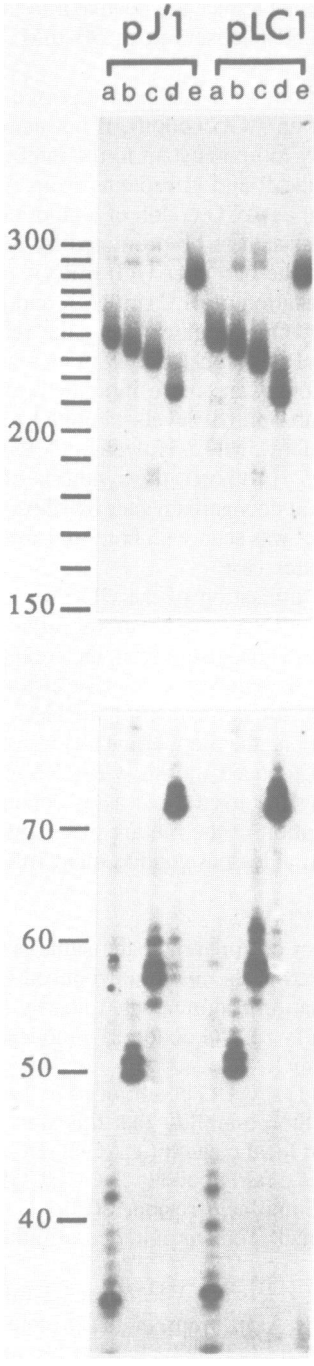

Figure 5. Site-specific cleavage of pJ'1 and pLC1 mRNA derivatives by RN'ase H and oligonucleotides. Plasmid DNAs were linearised by digestion with Hae III, and transcribed to give short (approx. 300 nucleotides) ^{32}P-labelled uncapped RNAs, which were cleaved in the presence of (a) oligonucleotide no. 592, (b) no. 595, (c) no. 593, (d) no. 594 (see Fig. 1), or (e) minus oligonucleotide, by incubation for 30 min at 32°C under the following conditions: 200 μg/ml RNA, 33 μg/ml oligonucleotide, 100 units/ml RN'ase H, 1 mM dithiothreitol, 100 mM KCl, 50 μg/ml bovine serum albumin (nuclease-free), 1 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5. Aliquots (2 μl) of each reaction were analysed on a 6% polyacrylamide sequencing gel, with a DNA sequencing reaction loaded on the same gel to provide size standards. The two relevant portions of the autoradiogram are shown (different exposures), with the positions of marker DNA fragments indicated in the left margin.

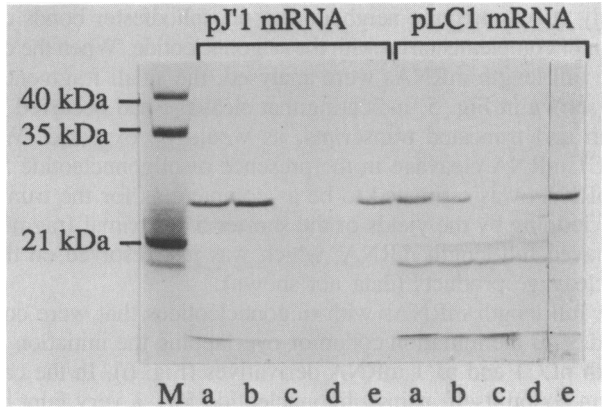

Figure 6. Products of translation of pJ'1 and pLC1 mRNAs subjected to cleavage at specific sites. Full-length uncapped pJ'1 and pLC1 mRNAs were incubated as in Fig. 5 with RN'ase H and (a) oligonucleotide no. 592, (b) no. 595, (c) no. 593, (d) no. 594, or (e) no oligonucleotide, then translated without further purification, using final concentrations of 20 μg/ml for pJ'1 mRNA and 30 μg/ml for pLC1 mRNA derivatives. After 35 min incubation, samples were taken for product analysis by gel electrophoresis and fluorography of the dried gel.

and [^{35}S]methionine in the presence of anisomycin and sparsomycin to limit elongation (20) showed that pJ'1 and pLC1 mRNAs produced exactly the same pattern of labelled peptides, which were quite different from those obtained using β-globin mRNA (Fig. 4). The peptides synthesised in the presence of the influenza NS mRNAs were acidic (Fig. 4), consistent with the predicted *N*-terminal sequence (met-asp-pro-asn ...) for initiation at the authentic initiation site, whereas initiation at either of the two GUG codons could not produce a set of predominantly acidic peptides. The identity of the pattern of labelled peptides produced from the two mRNA species shows that the translation of pLC1 mRNA was initiated at the same site as pJ'1 mRNA, and further implies initiation of pLC1 mRNA translation using a methionine residue. It should be noted that pL1.1 mRNA, in which the local sequence context of the initiation site has been mutated to ...UUUAUGG..., gave precisely the same peptide pattern as the parent pJ'1 (...AUAAUGG...) and the mutant pLC1 mRNA (Fig. 4). Thus the pattern of peptides was uninfluenced by the presence or absence of the AUA codon immediately upstream of the AUG (CUG) codon. This implies that very little, if any, initiation occurred at this AUA codon of the influenza NS mRNA, although AUA has been shown to function as a moderately efficient initiation codon in a mutated form of mouse dihydrofolate reductase mRNA (11).

As an alternative approach to identify the initiation site, the influenza NS mRNA derivatives were cut at a specific site (prior to translation) by incubation with appropriate complementary oligonucleotides and ribonuclease H (21). Uncapped transcripts were used for these experiments as oligonucleotide/RNase H scission produces uncapped RNA fragments. In addition to transcripts of the entire NS cDNA, scission was also carried out with short (300 nt.) transcripts, so that the efficiency of cleavage could be monitored using sequencing gels to detect the yield of both the large and small fragments. Fig. 5 shows that scission of these short transcripts was complete in all cases, and also allows the sites of cleavage to be determined from the sizes of the small fragments. The deduced cleavage sites are shown in Fig. 1: except with oligonucleotide 592, scission occurred

almost exclusively at two or three neighbouring phosphodiester bonds close to the mid-point of the region of complementarity with the oligonucleotide. When the cleavage products derived from the full-length mRNAs were analysed, the small fragments were precisely the same size as shown in Fig. 5, indicating that cleavage had occurred at the same sites on the full-length and truncated transcripts, as would be expected. With the possible exception of pLC1 mRNA cleavage in the presence of oligonucleotide 594, cleavage of the full-length mRNAs was estimated to be as complete as for the truncated transcripts shown in Fig. 5, judging by the yields of the shorter 5′-proximal fragments and the lack of residual uncleaved full-length mRNA, which was just resolved on the gels from the long (5′-distal) cleavage products (data not shown).

Scission of the full-length mRNAs with oligonucleotides that were complementary to sequences located 3′ to the initiation codon or overlapping the initiation codon abolished translation of both pLC1 and pJ′1 mRNA derivatives (Fig. 6). In the case of the RNAs cut at the initiation codon itself, using oligonucleotide 593, a very faint band was visible on the original autoradiogram, corresponding to a protein slightly shorter than the full-length product, which may possibly have arisen by initiation at the downstream GUG codon. In contrast, scission at sites located upstream of the AUG(CUG) codon had no significant effect on the efficiency of translation or the nature of the products synthesised (Fig. 6). The identical behaviour of pJ′1 and pLC1 mRNAs in these experiments again shows that the translation of both mRNAs was initiated at the same site, which is clearly located downstream of the 5′-proximal CUG and GUG codons and can be pin-pointed to the region complementary to oligonucleotide 593, i.e. nt. 52−65, where the AUG (CUG) is situated. Essentially the same result was obtained when the truncated (~300 nt) transcripts were examined using a similar experimental protocol, except that in this case the short length of the transcripts limited the analysis to assay of [^{35}S]methionine incorporation and precluded examination of the translation products. The results obtained with the transcripts of pJ′1 and pLC1 were indistinguishable: in both cases scission using oligonucleotides 593 and 594 almost completely inhibited incorporation, but oligonucleotides 592 and 595 did not (data not shown).

DISCUSSION

Hitherto the status of CUG as an initiation codon in eukaryotic systems has been somewhat controversial (see Introduction), and the only non-AUG codon to have gained general recognition is ACG. This is found at the initiation site for the adeno-associated virus B coat protein, and is utilised *in vivo* at an efficiency of 5−10% relative to initiation of C capsid protein synthesis at an AUG codon located in a favourable context on the same mRNA species (6,22). Synthesis of the C′ protein encoded by the Sendai virus P/C mRNA is initiated at an ACG codon, with an efficiency *in vivo* comparable to that of P protein synthesis (initiated at an AUG codon in unfavourable local sequence context) and about 20% relative to initiation of C protein synthesis, which is at an AUG codon with optimal context (7). When the initiation codon of murine dihydrofolate reductase was mutated to an ACG codon, it was used *in vivo* at an efficiency in the range 5−15% of the control (11,19). These values contrast with the <1% relative efficiency previously considered the norm (4,5).

When these three mRNAs were assayed in cell-free systems, the relative efficiency of initiation at the ACG codon was reported to be about the same as *in vivo* in the case of adeno-associated virus mRNA (6,22), and slightly higher than *in vivo* for the Sendai virus

P/C mRNA (7). Only with the dihydrofolate reductase mutant could it be said that ACG and other non-AUG codons functioned at significantly higher relative efficiency *in vitro* than *in vivo* (11,19). A tendency to see higher efficiency *in vitro* may be related to the particular assay conditions chosen, since our work and that of others (18,19) suggests that the utilisation of non-AUG codons may be more sensitive to these variables than initiation at AUG codons. Whilst these considerations argue that a low efficiency of initiation at non-AUG codons *in vitro* might justifiably be viewed with suspicion as an *in vitro* artefact of questionable physiological relevance, the efficiency of pLC1 mRNA translation is so high (at 50% relative to pJ'1 mRNA which is itself an unusually active message) that it surely cannot be dismissed in this way. Our assays were conducted under conditions optimised for the translation of pJ'1 mRNA, and we were unable to find modified conditions in which initiation at the CUG codon was reduced to negligible levels yet efficient pJ'1 mRNA translation was retained.

What is particularly striking about efficient initiation at non-AUG codons is that it seems such a rare event, confined so far to a few particular sites on a very limited number of mRNA species, and yet in-frame ACG and CUG codons would seem to be not uncommon in the 5'-untranslated segments of eukaryotic mRNAs. Certainly the adeno-associated virus capsid protein mRNA has numerous upstream ACG codons, including three in what appear to be favourable local sequence contexts, but only one of these is actually functional as an initiation codon, and the others were as silent *in vitro* as *in vivo* (6,22). The sequence also has two in-frame CUG codons, both in favourable contexts, but the product analysis clearly shows that these were also silent *in vitro* and *in vivo* (6). Inspection of other randomly chosen mRNA sequences gives further indication that there is not the same degree of selection against the occurrence of ACG or CUG in 5'-untranslated sequences as there is against the presence of AUG codons (23), but it would be important to substantiate this impression with a thorough statistical analysis.

The exclusive use of the mutated initiation codon in pLC1 mRNA, rather than the upstream CUG and GUG codons, reinforces the argument that the high efficiency of pLC1 mRNA translation cannot be ascribed to a general lack of stringency of initiation site selection in the reticulocyte lysate system. As the local (Kozak) context of the authentic initiation site does not seem to be distinctly more favourable than that of the non-functional upstream GUG codon, our results suggest that there must be some wider context that allows such efficient use of the mutant CUG codon. An observation consistent with this notion is that when the local (Kozak) context of pJ'1 mRNA was changed from ...AUAAUGG... to...UUUAUGG... the reduction of translation (to about 80% of control levels) was far less than the decrease expected from studies of preproinsulin mRNA translation, from which it was predicted that translation would be less than 20% of control levels (3).

The laboratory-constructed mutants of murine dihydrofolate reductase mRNA provide an interesting parallel. Changes in the local (Kozak) context influenced the efficiency of utilisation of the normal initiation site *in vitro* only with mRNA with an ACG initiation codon, not with the wild type AUG initiator (19). A plausible explanation is that there are wider context features of the initiation sites of both the NS and dihydrofolate reductase mRNAs which make these sites exceptionally efficient, such that (a) they can still function quite effectively when non-AUG codons are substituted, and (b) changes in local (Kozak) context have only minor effects when the initiation codon is AUG, but can be influential when the 'strength' of the initiation site is reduced by substituting a non-AUG initiation codon.

As for what the putative wider context features may be, the most striking feature of the portion of the untranslated leader sequence derived from the NS mRNA is its high purine content, especially adenine (Fig. 1). This sequence has little homology with the 'wider' context (...$(GCC)_nGCCACCAUGG$...) recently proposed by Kozak (24). The possibility that sequences 3′ to the initiation codon are important deserves serious consideration, particularly as the RNase H cleavage experiments suggested that the CUG codon was still used quite efficiently even with only a short residual 5′-untranslated segment (Fig. 6). This notion is supported by recent experiments in which 18 nucleotides between the BamH I site and the downstream Hind III site (Fig. 1) were deleted from both pJ′1 and pLC1 in two different ways: after cutting with one of these enzymes and removing the 5′-protrusion with mung bean nuclease, the plasmid was cut with the other enzyme and the end in-filled before ligation. When it was the BamH I site that was in-filled, generating the sequence A/CUGGAUUUUCAG..., the relative efficiency of initiation at the CUG codon was the same as with the undeleted parent mRNAs. However, the other type of deletion with an in-filled Hind III site, which generated the sequence A/CUGAGCUUUCAG..., reduced the efficiency of initiation at the CUG codon to about 10% relative to utilisation of the AUG in the otherwise identical construct (F.P.Carlotti, M.T.Howell and R.J.J., unpublished observations). The low yield of full-length product initiated at the CUG of this mRNA was coupled with an increase in the yield of the shorter products discussed previously.

The significant point is that it is no longer valid to interpret mRNA sequences with the assumption that only AUG codons can serve as potential initiators. In addition to ACG, CUG must now be listed amongst the permitted mammalian initiation codons. This suggests that all codons related to AUG by a single base substitution may have this potential, although the recent studies of mouse dihydrofolate reductase mRNA mutants indicated that ACG and CUG were by far the most effective, whilst AGG and AAG were completely silent (11). Our results, and the other evidence cited here, show that ACG or CUG codons located at particular sites can be at least as efficient initiation codons as the average AUG codon situated in a mediocre or unfavourable local (Kozak) context. Clearly, the identification of the special context features that allow such unexpectedly efficient initiation at non-AUG codons at selected sites is now an intriguing and urgent problem.

ACKNOWLEDGEMENTS

We thank Drs. S.C. Inglis and V.C. Blok for the gift of the NS cDNA clone; Mike Howell, Tim Hunt and Jon Pines for advice; and Ann Kaminski for carrying out the peptide mapping. This work was supported by a grant from the Medical Research Council, and a grant from the Wellcome Foundation for the establishment of the Departments of Biochemistry and Pathology oligonucleotide production facility. M.C.D. gratefully acknowledges the support of a scholarship from the Marshall Aid Commemoration Commission.

*To whom correspondence should be addressed

+Present address: Department of Biology, University of California at San Diego, La Jolla, CA 92093, USA

REFERENCES

1. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
2. Kozak, M. (1989) *J. Cell Biol.* 108, 229–241.
3. Kozak, M. (1986) *Cell* 44, 283–292.

4. Donahue, T.F. and Cigan, A.M. (1988) *Mol. Cell. Biol.* **8**, 2995–2963.
5. Clements, J.M., Laz, T.M. and Sherman, F. (1988) *Mol. Cell. Biol.* 8, 4533–4536.
6. Becerra, S.P., Rose, J.A., Hardy, M., Baroudy, B.M. and Anderson, C.W. (1985) *Proc. Natl. Acad. Sci. USA,* 82, 7919–7923.
7. Curran, J. and Kolakofsky, D. (1988) *EMBO J.* **7**, 245–252.
8. Gupta, K.C. and Patwardhan, S. (1988) *J. Biol. Chem.* 263, 8553–8556.
9. Patwardhan, S. and Gupta, K.C. (1988) *J. Biol. Chem.* 263, 4907–4913.
10. Hann, S.R., King, M.W., Bentley, D.L., Anderson, C.W. and Eisenman, R.E. (1988) *Cell* **52**, 185–195.
11. Peabody, D.S. (1989) *J. Biol. Chem.* **264**, 5031–5035.
12. Young, J.F., Desselberger, U., Graves, P., Palese, P., Shatzman, A. and Rosenberg, M. (1983) in The Origin of Pandemic Influenza Viruses, ed. Laver, W.G. New York, Elsevier Science Publishing Co. pp 129–138.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1986). Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor Laboratory.
14. Hattori, M. and Sakaki, Y. (1986) *Anal. Biochem.* 152, 232–238.
15. Dasso, M.C. and Jackson, R.J. (1989) *Nucleic Acids Res.* 17, 3129–3144.
16. Jackson R.J. and Hunt, T. (1983) *Methods in Enzymology* **96**, 50–74.
17. Jackson, R.J. (1986)*Virology* 149, 114–127.
18. Gupta, K.C. and Kingsbury, D.W. (1985) *Biochem. Biophys. Res. Commun.* 131,91–97.
19. Peabody, D.S. (1987) *J. Biol. Chem.* **262**, 11847–11851.
20. Jackson, R.J. and Hunter, A.R. (1970) *Nature* 227, 672–676.
21. Minshull, J. and Hunt, T. (1987) *Nucleic Acids Res.* **14**, 6433–6451.
22. Becerra, S.P., Koczot, F., Fabisch, P. and Rose, J.A. (1988) *J. Virol.* **62**, 2745–2754.
23. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
24. Kozak, M. (1987) *J. Mol. Biol.* 196, 947–950.
25. Darnbrough, C.H., Legon, S., Hunt, T. and Jackson, R.J. (1973) *J.Mol.Biol.* **76**, 379–403.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.