The EMBO Journal vol.13 no.7 pp.1673-1681, 1994

# Translation of encephalomyocarditis virus RNA: parameters influencing the selection of the internal initiation site

**Ann Kaminski, Graham J.Belsham[1] and Richard J.Jackson**

Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, and [1]AFRC Institute for Animal Health, Pirbright, Woking, Surrey GU24 0NF, UK

Communicated by R.J.Jackson

**The initiation of encephalomyocarditis virus RNA translation is by internal ribosome entry almost exclusively at the 11th AUG codon from the 5′-end, which is the central of the three AUG codons in the sequence..ACGAUGAUAAUAUGGCCACAACCAUG.., and is located some 25 nt downstream from an oligopyrimidine tract conserved amongst related viruses. As the sequences between the oligopyrimidine tract and AUG-10/11 are poorly conserved and thus possibly serve only as a spacer, the influence of this spacer length on initiation frequency at the three AUG codons was examined *in vitro* and *in vivo*. Deletion of 11 residues resulted in initiation almost exclusively at AUG-12 but at significantly reduced overall efficiency. Insertion of eight residues caused a 15-fold increase in initiation frequency at AUG-10 and a decrease at AUG-11. Longer insertions reduced overall efficiency without changing the initiation site preferences. With the wild-type spacing, complete substitution of the oligopyrimidine tract by purines caused a 30–35% decrease in initiation efficiency, and partial substitution only a 10–15% decrease. Thus the internal initiation mechanism selects the initiation site partly on the basis of its distance from upstream elements, of which the oligopyrimidine tract is not the most critical, but for reasons not yet understood a preference for AUG-11 is superimposed on this selection.**

*Key words:* encephalomyocarditis virus/internal ribosome entry/oligopyrimidine tract/translation initiation

## Introduction

It is now well established that the initiation of translation of picornavirus RNAs, or of any mRNAs bearing a picornaviral 5′-untranslated region (5′-UTR) takes place by a mechanism of internal ribosome entry (Jackson *et al.*, 1990; Meerovitch and Sonenberg, 1993), in contrast to the mechanism of ribosome scanning from the 5′-end of the mRNA which is believed to be relevant to the vast majority of cellular and viral mRNAs (Kozak, 1989, 1992). This internal entry mechanism requires a defined segment (~450 nt) of the picornaviral 5′-UTR, which is generally known as the IRES element (for 'internal ribosome entry segment'). On the basis of primary sequence homologies of the IRES element, the picornavirus family can be divided into three groups: (i) entero- and rhinoviruses; (ii) cardio- and aphthoviruses; and (iii) hepatitis A virus. There is quite strong conservation of primary sequence of the IRES element within each of these groups, but almost no conservation between groups apart from a pyrimidine-rich tract located 25 nt from the 3′-end of the IRES and an AUG triplet at the very 3′-end (Jackson *et al.*, 1990; Agol, 1991).

The actual ribosome entry site in all picornaviruses is thought to be at or very close to this AUG triplet at the 3′-end of the IRES (Jackson *et al.*, 1990; Meerovitch and Sonenberg, 1993), but there are differences between the various species as to the events which follow initial ribosome entry at this site. In the cardioviruses, virtually all the entering ribosomes initiate translation at this AUG codon, which is the authentic initiation site for viral polyprotein synthesis (Kaminski *et al.*, 1990). In the aphthoviruses (foot and mouth disease virus) a minority of the entering ribosomes initiate translation at this AUG codon, whilst the majority are transferred, most probably by a scanning mechanism, to the next AUG codon downstream and initiate translation at this site (Belsham, 1992). And in the entero-/rhinoviruses the AUG trinucleotide motif at the 3′-end of the IRES does not seem to be used at all as an initiation site, but appears to be a determinant of the internal ribosome entry site from which the ribosomes are transferred, likewise probably by a scanning mechanism, to the authentic initiation site which is invariably the next AUG codon located ~40 nt or ~160 nt further downstream in rhinoviruses and enteroviruses respectively (Agol, 1991; Pilipenko *et al.*, 1992). However, there are viable deletion mutants of poliovirus in which ~160 nt have been deleted such as to retain just the 5′-proximal portion of the pyrimidine-rich tract and to place the authentic initiation codon ~25 nt downstream from this vestige of the pyrimidine-rich tract, similar to the situation in cardioviruses (Kuge and Nomoto, 1987; Iizuka *et al.*, 1989; Haller and Semler, 1992; Pilipenko *et al.*, 1992).

It follows that cardioviruses offer the best system for studying the accuracy of internal ribosome entry, and with encephalomyocarditis virus (EMCV), strain R, it has been shown that the position of internal ribosome entry is selected very precisely, in that the authentic initiation codon (AUG-11) at nt 834 is utilized at high efficiency, whereas a nearby upstream AUG codon (AUG-10) at nt 826 is not recognized as an initiation site despite having a local sequence context that would make it efficiently used by a scanning ribosome (Kaminski *et al.*, 1990). Whilst internal ribosome entry is clearly dependent on the whole ~450 nt IRES element upstream of AUG-11, the extreme precision of the entry site selection suggests that there might be local determinants allowing such fine discrimination between two closely spaced AUG codons. However, the sequences of the IRES elements between the oligopyrimidine tract and the AUG initiation codon are not highly conserved between different species of cardio-/aphthoviruses, nor even between various strains of any one species (Sangar *et al.*, 1987; Pritchard *et al.*, 1992). On the other hand the length of this segment is quite strongly conserved, and it is invariably

G-poor. These features suggest that the segment may serve as an unstructured spacer element rather than the repository of primary sequence determinants directing internal initiation to AUG-11. This, in turn, raises the possibility that selection of the internal ribosome entry/initiation site and discrimination between AUG-10 and AUG-11 in EMCV strain R is based on the distance of the AUG codon from upstream motifs including the pyrimidine-rich tract. The original aim of the work described here was to test this hypothesis by examining the effect of changing this distance on the discrimination between AUG-11 and AUG-10, and on the overall efficiency of internal initiation.

In the course of these experiments it was observed that the efficiency of internal initiation was only marginally influenced by the number of pyrimidine residues retained in the oligopyrimidine tract of the different mutants. This prompted the construction of mutants in which the pyrimidines were substituted by purines (mainly by A residues). In view of previous claims that at least the 5′-proximal part of the pyrimidine-rich tract is essential for IRES function in several different picornavirus species (Iizuka *et al.*, 1989; Jang and Wimmer, 1990; Kuhn *et al.*, 1990; Meerovitch *et al.*, 1991; Nicholson *et al.*, 1991; Pestova *et al.*, 1991; Pilipenko *et al.*, 1992), it was surprising to find that initiation efficiency was decreased by only 10–15% when all but two of the pyrimidines were substituted by purines, and by not more than 30–35% when the whole pyrimidine-rich tract was replaced in this way.

## Results

### *Constructs and strategy*

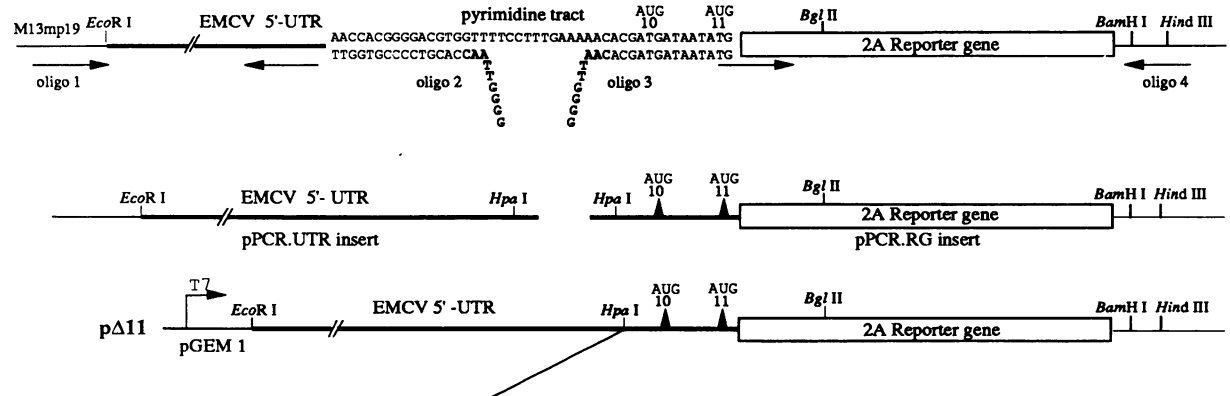
Monocistronic constructs with the complete EMCV 5′-UTR downstream of the poly(C) tract were used throughout these investigations, since our previous work has demonstrated that scanning from the 5′-end of such RNA transcripts makes such a negligible contribution to initiation at AUG-10 and AUG-11 that internal initiation can legitimately be studied with monocistronic mRNAs (Kaminski *et al.*, 1990). The starting point for this work was the construct pSG1, in which the insert is basically the same as in pS32A3 of Jang *et al.* (1988) and pAΔ258 described previously (Kaminski *et al.*, 1990), except that it has been transferred into pGEM-1. This insert consists of a segment of the EMCV genome from nt 259 to 848 (the authentic initiation codon for viral polyprotein synthesis is the 11th AUG codon from the 5′-end and is at nt 834) linked to the poliovirus 2A cistron as the reporter, which is followed by two tandem termination codons (Jang *et al.*, 1988). The sequence at the junction is ...AUGAUAAUAUGGCCACAACCAUG*UAUG*... (Duke *et al.*, 1992), where AUG-11, the authentic initiation codon for viral polyprotein synthesis, is underlined, and the sequences of the poliovirus 2A reporter cistron are in italics. Thus there are four AUG codons within a 27 nt segment. The central pair of AUG codons (AUG-11 and AUG-12) are in-frame with the reporter cistron (and in the viral genome are in-frame with the polyprotein coding sequences), while AUG-10 and AUG-13 are both in the same alternative reading frame. In order to allow initiation at AUG-10 and AUG-13 to be detected without perturbing the sequences in this critical segment with the four AUG codons, a frame-shift is introduced by in-filling a unique *Bgl*II site located some 60 nt downstream from this region (Kaminski *et al.*, 1990). Thus by using a pair of constructs, one with the frame-shift in-fill and the other without, it is possible to assay the yield of translation product from each of these four AUG codons, with the caveat that the products arising from initiation at AUG-11 and AUG-12 migrate so close together on polyacrylamide gels that it is not possible to make quantitative estimates of the relative yield of each product, though it is possible to distinguish which site was used in cases where initiation occurred at only one of the two sites.

In order to investigate how the spacing between the start of the pyrimidine-rich tract and AUG-11 influences the overall efficiency of internal initiation, and the discrimination between these four AUG codons as potential initiation sites, a deletion of 11 residues was first made in the 5′-proximal half of this segment to produce pΔ11. This deletion was constructed by fusing two PCR products at a unique *Hpa*I site in a way that generated an exact 11 nt deletion (Figure 1). The PCR products were sequenced in their entirety to ensure that no unplanned mutations had been introduced.

Oligonucleotides of various lengths were then introduced at this *Hpa*I site of pΔ11 in order to increase the distance between the pyrimidine-rich tract and AUG-11. These constructs are designated according to the number of nucleotides inserted, i.e. pΔ11/19 signifies an insertion of 19 nt into pΔ11, which effectively increases the distance between the oligopyrimidine tract and AUG-10 by 8 nt over the distance in the wild-type viral genome, and thus places AUG-10 in the same position relative to upstream elements as the authentic initiation site (AUG-11) occupies in the wild-type (Figure 1). As a fortuitous by-product of this reconstruction resulting from an unplanned cloning artefact, a construct was obtained with 15 nt inserted in the *Hpa*I site (pΔ11/15), i.e. a 4 nt increment over the spacing in the wild-type genome. The 19 nt insert includes a unique *Bcl*I site, so that by in-filling this site the insert at the *Hpa*I site of the deleted construct could be expanded to 23 nt, or 12 nt over the wild-type spacing (Figure 1).

The oligonucleotides for insertion at the blunt-ended *Hpa*I site were designed so that in one orientation they would regenerate the complete pyrimidine-rich stretch found in the EMCV genome; this orientation was designated the 'sense' orientation, denoted by the suffix 's' in the nomenclature of the plasmid constructs, while the reverse orientation is designated by the suffix 'a' signifying antisense (Figure 1). In addition to the constructs generated by oligonucleotide insertion at the *Hpa*I site, a number of additional constructs were generated by site-directed mutagenesis. Figure 1 shows the sequences between the oligopyrimidine tract and AUG-11 for all of the various mutants studied. All constructs were also subjected to the frame-shifting in-fill at the *Bgl*II site, so that the utilization of AUG-10 and AUG-13 as initiation sites could be monitored. Such constructs are designated by the additional suffix '(fs)', for frame-shift.

All these constructs were made in the same background of the pGEM-1 vector as the parent pSG1, thus permitting *in vitro* transcripts to be prepared using T7 RNA polymerase. These transcripts were then translated in the rabbit reticulocyte lysate system over a range of RNA concentrations, and the labelled translation products were analysed by gel electrophoresis followed by autoradiography of the dried gel and densitometry of the autoradiogram to obtain quantitative data on the yield of translation products resulting from initiation at the four AUG codons. Although the




| | | Relative yield of product initiated at: | | | |
|---|---|---|---|---|---|
| | | AUG-10 | AUG-11/12 | AUG-13 | Total |
| pΔ11 | GUU AACACGAUGAUAAU**AUG**GCCACAACCAUG*UAUG* | 0 | 15* | 1 | 16 |
| pΔ11/11s (same as wild type) | GUU UUCCUUUGAAA AACACGAUGAUAAU**AUG** | 1 | 100 | 1 | 102 |
| pΔ11/11a | GUU UUUCAAAGGAA AACACGAUGAUAAU**AUG** | 1 | 93 | 1 | 95 |
| pΔ11/15s | GUU UCCUUUGAUCAUACU AACACGAUGAUAAU**AUG** | 6 | 74 | 1 | 81 |
| pΔ11/19s | GUU UUCCUUUGAUCAUACUAAA AACACGAUGAUAAU**AUG** | 15 | 76 | <1 | 91 |
| pΔ11/19a | GUU UUUAGUAUGAUCAAAGGAA AACACGAUGAUAAU**AUG** | 15 | 72 | <1 | 87 |
| pΔ11/23s | GUU UUCCUUUGAUCGAUCAUACUAAA AACACGAUGAUAAU**AUG** | 10 | 45 | <1 | 55 |
| pΔ11/23a | GUU UUUAGUAUGAUCGAUCAAAGGAA AACACGAUGAUAAU**AUG** | 7 | 36 | <1 | 43 |
| pAUG-10M | GUU UUCCUUUGAAA AACACGAUGGUAAU**AUG** | 1 | 81 | 1 | 83 |
| pΔ11/19s AUG-10M | GUU UUCCUUUGAUCAUACUAAA AACACGAUGGUAAU**AUG** | 16 | 65 | 1 | 82 |
| pΔ11/11PYR | GUU UUCCUUUUUUU AACACGAUGAUAAU**AUG** | 1 | 72 | 1 | 74 |
| pΔ11/11PUR | GUU AAAAAAAGGAA AACACGAUGAUAAU**AUG** | 1 | 89 | 1 | 91 |
| pΔ11/11PURMAA | GAA AAAAAAAGGAA AACACGAUGAUAAU**AUG** | n.d. | 66 | n.d | >66 |
| pΔ11/11PURMAAc | UUC..GAA AAAAAAAGGAA AACACGAUGAUAAU**AUG** | n.d. | 68 | n.d | >68 |

**Fig. 1.** Construction and sequences of mutants used to study initiation site selection, and summary of results. The upper part of the figure (not to scale) depicts the construction of pΔ11, starting with a single-stranded sense orientation of the parent pSG1 insert cloned into M13mp19. The sequence of the two dedicated oligonucleotides (#2 and #3) used for the PCR reactions is shown; oligonucleotides #1 and #4 are sense and anti-sense versions, respectively, of M13mp19 sequences flanking the multiple cloning site. After subcloning each PCR product individually into pGEM-1, pΔ11 was constructed (also in pGEM-1) from the two subclones as described in Materials and methods. The lower part of the figure shows the (RNA) sequences between the start of the oligopyrimidine tract and the authentic initiation codon (AUG-11) which were generated by insertion of various oligonucleotides into the *HpaI* site of pΔ11 (the inserted sequences are flanked by gaps), and, in the case of some constructs, by subsequent site-directed mutagenesis. The authentic initiation codon (AUG-11) is shown in bold, and AUG-10 is underlined. The sequence given for pΔ11 shows the sequences immediately downstream of AUG-11 up to and including AUG-13, a sequence which was common to all constructs. The sequence from AUG-11 to AUG-12 inclusive is EMCV sequence coding for the N-terminus of the viral polyprotein, and the sequence in italics represents the start of the poliovirus 2A reporter cistron. The yield of labelled translation products initiated at each AUG codon in each construct determined by scanning densitometry of autoradiographs is given as the percentage of the yield of product initiated at AUG-11/12 when the parent construct pSG1 (identical to pΔ11/11s) was translated. For each construct, the relative yields were calculated separately for each RNA concentration tested (25, 12.5 and 6.3 μg/ml) in a representative set of assays, and the average of the three values is shown. The relative yields of products initiated at AUG-10 and AUG-13 were determined from the translation assays of transcripts of the constructs bearing the frame-shift in-fill of the *BglII* site. An asterisk denotes that in this particular case all initiation was at AUG-12 and none at AUG-11 (Figure 2); n.d. signifies that the construct to evaluate initiation at AUG-10 and AUG-13 was not made.

concentration of the RNA in the assay system influenced the overall yield of translation products, it was found to have no significant influence on the relative frequency of initiation at the four AUG codons under scrutiny.

### *Selection of the translation initiation codon is partly dependent on its distance from the oligopyrimidine tract*

A representative set of autoradiographs from the gel electrophoretic separations of the translation products of all the different transcripts is shown in Figures 2 and 3, and a summary of the quantitative results obtained by scanning densitometry of such autoradiograms is given in Figure 1. It may be noted that besides the products from initiation at the four AUG codons, a number of smaller products were synthesized (Figures 2 and 3). For the majority, the yield increased with increasing RNA concentration and with increasing efficiency of the main initiation site(s), which are characteristics of products generated by premature termination of translation rather than by initiation at interior sites in the reporter cistron (Dasso and Jackson, 1989a,b; Dasso *et al.*, 1990). Moreover, our previous results with this same system demonstrate that none of these smaller products can originate from initiation at AUG-8 or AUG-9 (Kaminski *et al.*, 1990), and their size excludes the possibility that they are products of initiation at AUG codons further upstream in the IRES.

The translation assays of transcripts of the parent construct pSG1 and its frame-shift derivative pSG1(fs) show that the initiation frequency at AUG-10 and AUG-13 is ~1% of the

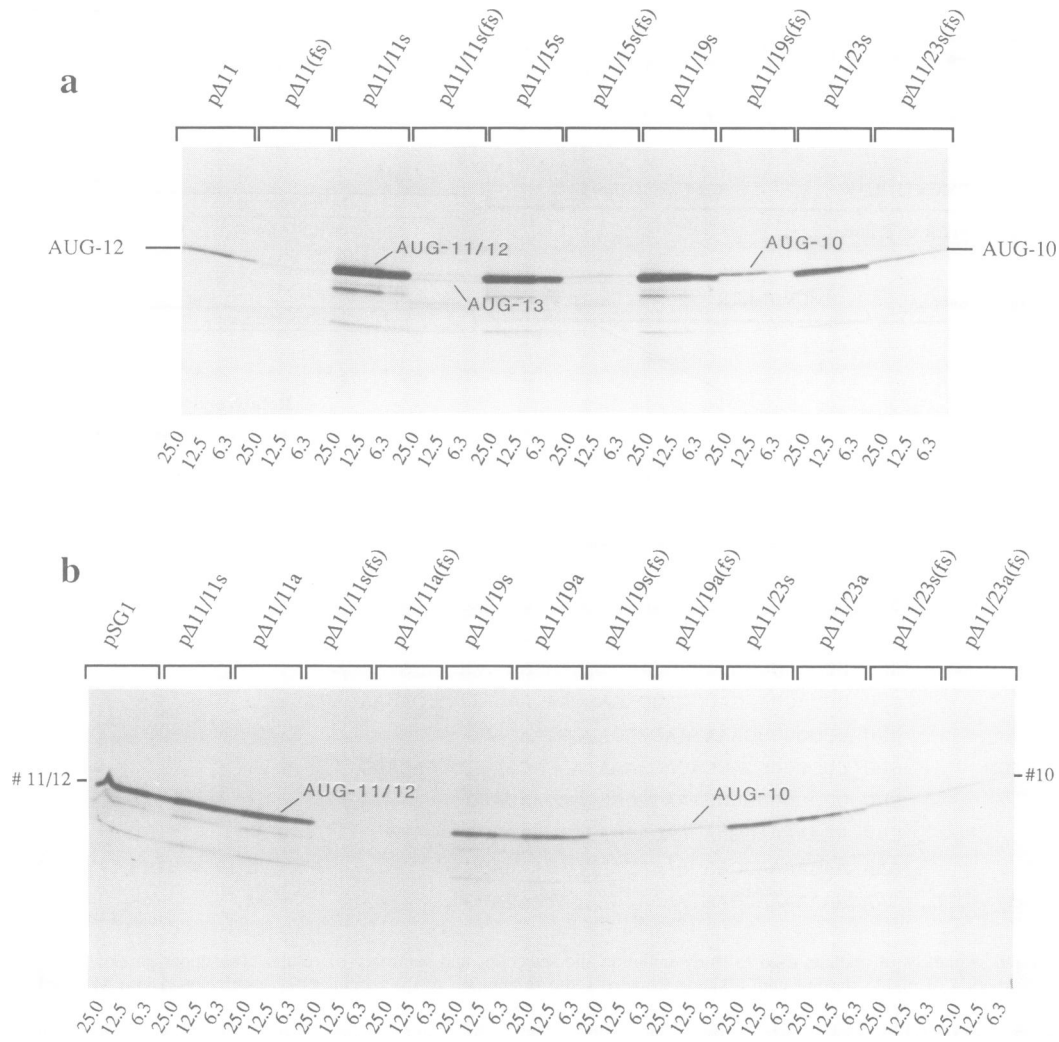



**Fig. 2.** Products of *in vitro* translation of transcripts of deletion and insertion mutants of the EMCV IRES/poliovirus 2A constructs showing: (**a**) the effect of an 11 nt deletion and of (re)insertions of different lengths of sequence at the site of this deletion; and (**b**) the influence of the orientation of the inserted sequences. (a) and (b) depict the autoradiographs of the gel electrophoretic analysis of the $^{35}$S-labelled translation products from two separate translation assays of transcripts of the designated plasmids assayed at final concentrations of 25, 12.5 or 6.3 μg/ml as shown (see Figure 1 for details of the sequence changes at the site of deletion/insertion, and for the quantitative evaluation of these results determined by scanning densitometry of autoradiographs). For orientation purposes the products in the extreme left and right tracks and in some central tracks are identified by the AUG codon from which synthesis was initiated.

frequency of initiation at AUG-11/12 (data not shown), in agreement with our previous results from experiments using a slightly different vector (Kaminski *et al.*, 1990). The same result was obtained with transcripts of the reconstructed version (pΔ11/11s) of this wild-type sequence obtained by (re)insertion of the 11 nt motif into what is effectively an 11 nt deletion version generated by fusion of the two PCR products (Figures 1 and 2). This supports the conclusion drawn from sequencing that there are no unplanned mutations in the PCR fusion construct pΔ11.

The transcripts from the 11 nt deletion construct (pΔ11) were translated at only ~16% of the efficiency of the parent pSG1 (Figure 1). Significantly, there was almost no initiation at AUG-11, and the majority of initiation events on the transcript of pΔ11 were at AUG-12 (Figures 1 and 2a). With transcripts of the frame-shift version of this construct, pΔ11/11s(fs), there was no detectable initiation whatsoever at AUG-10 (in contrast to the wild-type version, pSG1), and a low frequency at AUG-13 (Figures 1 and 2a).

When the distance between the oligopyrimidine tract and AUG-11 was increased over that which occurs in the wild-type, there was an increase in the utilization of AUG-10 and a decrease in the initiation frequency at AUG-11/12, as is strikingly obvious from visual inspection of the autoradiograph shown in Figure 2a. This increase in initiation at AUG-10 was already apparent with pΔ11/15s (the fortuitous cloning artefact), and was more significant in pΔ11/19s (Figures 1 and 2a). Thus with the wild-type distance, the ratio of initiation frequency at AUG-10 relative to that at AUG-11/12 is 1:100; in pΔ11/15s it is ~1:12; and in pΔ11/19s it is ~1:5 (Figure 1). It may also be noted that the increase in distance reduced utilization of AUG-13 to the point where the translation product was virtually undetectable (Figures 1 and 2a). These changes in the frequency of utilization of the different sites with changing spacing were accompanied by only a slight decrease in the overall efficiency of translation of the mRNA (Figure 1).

However, when the distance between the pyrimidine-rich

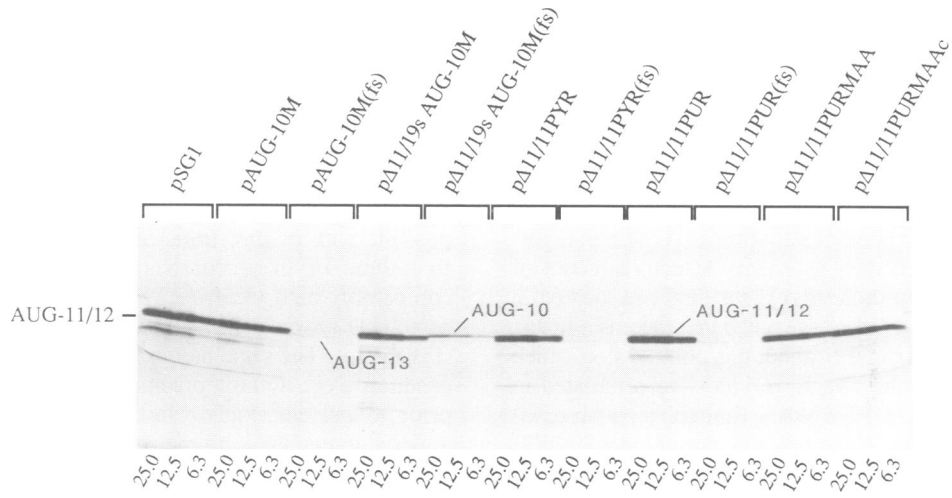



**Fig. 3.** Products of *in vitro* translation of transcripts of deletion, insertion and point mutants of the EMCV IRES/poliovirus 2A constructs showing the effect of changing the context of AUG-10, and of substituting purine-rich sequences in place of the oligopyrimidine tract. The figure depicts the autoradiograph of the gel electrophoretic analysis of the $^{35}$S-labelled translation products from translation assays of transcripts of the designated plasmids assayed at final concentrations of 25, 12.5 or 6.3 μg/ml as shown (see Figure 1 for details of the sequence changes at the site of deletion/insertion or point mutations, and for the quantitative evaluation of these results determined by scanning densitometry of autoradiographs). For orientation purposes the products in the extreme left and right tracks and in some central tracks are identified by the AUG codon from which synthesis was initiated.

tract and AUG-11 was increased by an additional four residues, as in the pΔ11/23 set of constructs, overall efficiency of translation of the mRNA decreased quite markedly (Figures 1 and 2). This decrease was not accompanied by a significant shift in initiation site preference towards utilization of AUG-10 (Figures 1 and 2a). The ratio of initiation frequency at AUG-10 to that at AUG-11/12 remained at ~1:5, the same as for the pΔ11/19 set of constructs (Figure 1). This pattern of results was obtained when even greater lengths of spacer were inserted: overall translation efficiency decreased further but without any striking increase in the utilization of AUG-10 relative to AUG-11/12 (data not shown).

In the pΔ11/19 set of constructs, AUG-10 is at the same distance from the pyrimidine-rich tract as is AUG-11 in the wild-type genome (Figure 1). Although this insertion of 8 nt increased the frequency of initiation at AUG-10 by ~15-fold, it did not result in a complete switch of initiation site preference from AUG-11/12 to AUG-10. One possible explanation for the relative inefficiency of AUG-10 utilization even when it is at what would seem to be the appropriate distance (as in pΔ11/19s), may be that the context of AUG-10, which is ACG**AUG**A, is not the ideal context for a translation initiation codon. Accordingly, this context was changed to ACG**AUG**G, both in the wild-type background and in pΔ11/19s; these mutants are designated pAUG-10M and pΔ11/19s AUG-10M, respectively. In the wild-type background this mutation decreased overall translation efficiency slightly, but had no significant effect on the relative initiation frequency at AUG-10 (Figures 1 and 3). With transcripts of pΔ11/19s AUG-10M there was again a slight decrease in overall translation efficiency, and this was accompanied by a marginal increase in the relative frequency of initiation at AUG-10 as compared with the construct lacking the context mutation (pΔ11/19s). The ratio of initiation frequency at AUG-10 to that at AUG-11/12 changed from 1:5 (in the case of pΔ11/19s) to 1:4 when this context change was made in pΔ11/19s AUG-10M (Figure 1).

### *The oligopyrimidine tract is not essential for efficient internal initiation at the correct site*

For each particular length of insert, the orientation of the insert had virtually no influence on the initiation site selection preference (Figures 1 and 2b). However, the overall efficiency of translation was somewhat lower with the antisense than with the sense orientation (Figure 1). Nevertheless, with the debatable exception of the pΔ11/23 pair of constructs, this effect of orientation on translation efficiency was quite small, which is somewhat surprising in view of the fact that (i) the antisense orientation generally generated a much shorter oligopyrimidine tract than the sense orientation (Figure 1), and (ii) this pyrimidine-rich tract has been implicated in other work as an important determinant for the internal initiation process (see Discussion).

To investigate this question of the importance of the oligopyrimidine tract in more detail, the 11 nt deleted in the generation of pΔ11 were replaced with oligonucleotides which would restore the wild-type genome spacing by inserting either 11 pyrimidines (in one orientation), or 11 purines. These constructs are designated pΔ11/11PYR and pΔ11/11PUR respectively. When transcripts of each of these two mutants were translated, the frequency of selection of the different AUG codons was very similar to that seen with the wild-type RNA (Figures 1 and 3). The overall efficiency of translation was actually lower for transcripts of pΔ11/11PYR than of pΔ11/11PUR (Figure 1), and the latter was translated only slightly less efficiently than the wild-type, pSG1 or its reconstructed equivalent pΔ11/11s.

The construct with the purine-rich insert described above still retains two U residues (at the *Hpa*I site) from the extreme 5′-end of the original oligopyrimidine tract found in the wild-type genome (Figure 1). Site-directed mutagenesis was used to mutate these last two U residues to As, generating the

construct designated pΔ11/11PURMAA (Figure 1) and thus replacing the whole oligopyrimidine tract found in the wild-type genome by an oligopurine tract. Although the outcome was a decrease in translational efficiency as compared with the RNA which retained the two U residues (pΔ11/11PUR), the decrease was only 30–35% (Figures 1 and 3), much less than would have been expected if the pyrimidine-rich tract is as essential for internal initiation as is often thought.

In the generally accepted secondary structure model of Pilipenko *et al.* (1989) these two U residues are located in a base paired stem, and indeed we noted that there is a strong stop to reverse transcriptase at about this point, a stop which was abrogated when the two U residues were mutated to As to generate pΔ11/11PURMAA (unpublished observations). Thus the lower efficiency of translation of pΔ11/11PURMAA as compared with pΔ11/11PUR might be due to disruption of secondary structure rather than to loss of the two U residues *per se*. Therefore, on the basis of the secondary structure model of Pilipenko *et al.* (1989), we introduced into pΔ11/11PURMAA mutations designed to restore the base-pairing as a result of the replacement of two upstream neighbouring A residues by U residues (Figure 1). However, this compensatory mutation generating pΔ11/11PURMAAc did not improve translation efficiency over that observed with pΔ11/11PURMAA (Figures 1 and 3).

### *The results of in vitro translation assays of the deletion and insertion mutants are also observed in vivo*

All of the constructs depicted in Figure 1 were tested for efficiency of translation and discrimination between different initiation sites *in vivo*, by transfection of BSC40 cells previously infected with recombinant vaccinia virus capable of expressing T7 RNA polymerase, and then incubation in the presence of [$^{35}$S]methionine. When the labelled translation products were separated by gel electrophoresis, only the 2A polypeptide product translated from AUG-11/12 can be seen against the background of vaccinia virus encoded polypeptides (Figure 4a), and is evident in transfection assays with the following constructs: pSG1, pΔ11/19s, pΔ11/23s, pΔ11/11PYR and pΔ11/11PUR. Although this type of assay does not lend itself to the detailed quantitative analysis possible with *in vitro* translation assays, it is clear that the first three of these constructs generated labelled 2A in comparably high yield, pΔ11/11PYR and pΔ11/11PUR in somewhat lower yield, and pΔ11/23s in yet lower amounts (Figure 4a). The same hierarchy was seen when the labelled products were immunoprecipitated with anti-2A antiserum prior to gel electrophoresis (data not shown), and is in general agreement with the result of the *in vitro* translation assays (Figure 1). Although no labelled 2A polypeptide product could be identified with any certainty in the transfection assay with pΔ11 (Figure 4a), this is perhaps not surprising given that transcripts of this construct were translated *in vitro* at <20% of the efficiency of the wild-type (pSG1, or pΔ11/11s) transcripts (Figure 1), and it seems unlikely that such a low level of translation product could be detected reliably against the background of labelled vaccinia virus encoded polypeptides (Figure 4a).

Likewise, on this type of gel electrophoretic analysis of total cell extracts, labelled products arising from initiation at either AUG-10 or AUG-13 cannot be detected with any certainty. However, on immunoprecipitation of the extracts with anti-2A antiserum followed by gel electrophoresis, the product of initiation from AUG-10 could be readily detected in the assays of pΔ11/19s(fs), pΔ11/23s(fs), pΔ11/19s AUG-10M(fs) (Figure 4b). A very low yield of this product from the transfections with the parent pSG1(fs), and pAUG-10M (fs) in which AUG-10 occupies the same position as in pSG1, was also apparent on the original

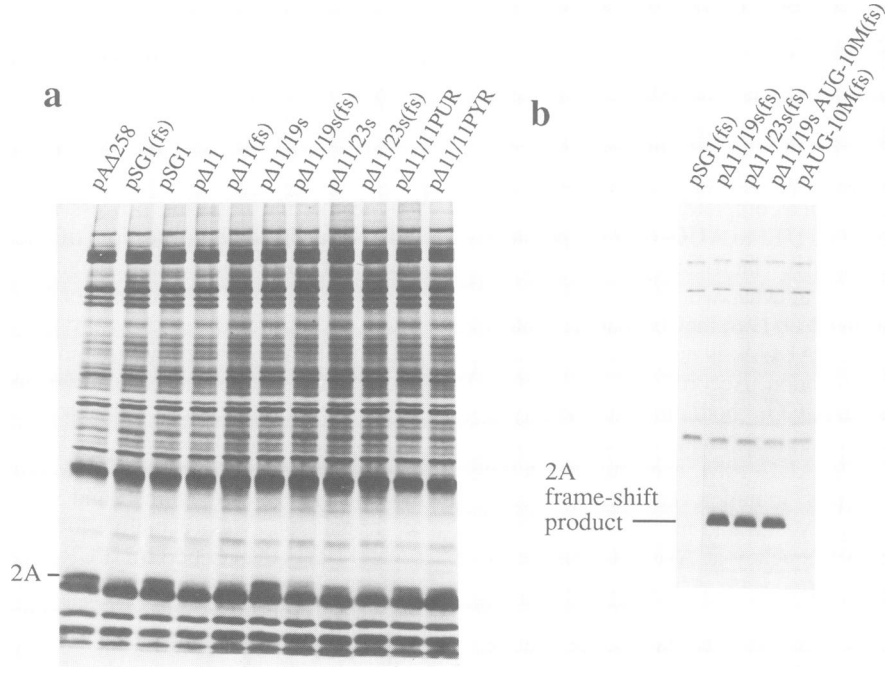



**Fig. 4.** Products expressed from the mutated EMCV IRESes in transfected BSC40 cells infected with recombinant vaccinia virus expressing T7 RNA polymerase. The figure depicts the autoradiograph of the gel electrophoretic separation of (**a**) total cell extracts, and (**b**) immunoprecipitates of total cell extracts using anti-2A antiserum. The cells were labelled by incubation with [$^{35}$S]methionine for 2 h starting ~20 h after transfection with the designated plasmid constructs.

autoradiograph but the faint band has been lost on photographic reproduction.

It should be noted that immunoprecipitation of the labelled products from such transfection assays cannot be used to estimate the relative frequency of initiation at AUG-10 as opposed to AUG-11/12, since the product initiated at AUG-10 of the frame-shift version differs in the N-terminal 25 amino acid residues from that initiated at AUG-11/12 of the construct lacking the frame-shift, a difference which could in principle, and does in practice, influence the efficiency of immunoprecipitation. This means that the only valid comparisons are those between constructs which all carry the frame-shift and thus code for the same polypeptide. Despite this limitation, the results shown in Figure 4b clearly demonstrate that the difference of 8 nt in the spacing between the start of the oligopyrimidine tract and AUG-10/11 in pΔ11/19s as opposed to the parent pSG1 results in an increase in the initiation frequency at AUG-10 *in vivo* at least as great as the 15-fold increase observed in the reticulocyte lysate translation assays (Figure 1).

## Discussion

One of the few features of the translation of picornavirus RNAs shared by all species of this group of viruses is that initiation is by an internal ribosome entry mechanism requiring a contiguous 450 nt segment of the viral 5′-UTR (the IRES element), that directs the ribosome to first contact the RNA in a potentially initiation-competent manner at an AUG codon located at the 3′-end of the IRES, some 25 nt downstream of an oligopyrimidine motif which is the only extended primary sequence motif common to all picornavirus IRESes (Jackson *et al.*, 1990; Pilipenko *et al.*, 1992; Meerovitch and Sonenberg, 1993). In consequence, considerable attention has been given to the role of the oligopyrimidine tract and the sequences between this tract and the AUG triplet at the 3′-end of the IRES.

The oligopyrimidine tract starts with a UUUUC or UUUC motif, but in the foot and mouth disease virus (FMDV) family this is preceded by two Cs (Sangar *et al.*, 1987) which are not usually considered part of the tract. The total number of consecutive pyrimidine residues ranges from seven to nine in the FMDV family (Sangar *et al.*, 1987) and nine in EMCV, to 15−17 (but usually with a single purine in the middle) in most entero-/rhinoviruses (sequences compiled in Pestova *et al.*, 1991). The distance between this tract and the AUG triplet at the 3′-end of the IRES is quite highly conserved: if measured on the basis used by Pilipenko *et al.* (1992) as the number of intervening residues between the end of the (U)UUUC motif and the A of the AUG codon, it is 21−24 nt in all strains of all picornavirus species except the FMDV family where it is 18−20 nt. However, the sequences following this conserved oligopyrimidine tract are unusually variable. Amongst FMDV strains this is one of the most variable regions of the whole IRES, and the variability extends up to the AUG initiation codon (Sangar *et al.*, 1987). Similarly, this region seems to be a hot-spot for strain-dependent variations in Theiler's murine encephalomyelitis virus (TMEV), although there is more conservation in the 4−5 nt just before the AUG codon (Pritchard *et al.*, 1992). The positions just downstream of the oligopyrimidine tract in polioviruses are also very variable even between different isolates of the same serotype (Poyry *et al.*, 1992), and even though the 9 nt immediately before the AUG are strongly (but not absolutely) conserved amongst all human enteroviruses, this conservation does not extend to the closely related bovine enterovirus and human rhinoviruses. Amongst all the species of picornaviridae the most conserved features of this segment are therefore the 5′-proximal oligopyrimidine tract, the length of the segment between this tract and the AUG codon, and the relative lack of G residues throughout this region.

All these features suggest that the segment between the oligopyrimidine tract and the AUG codon might be required only as an unstructured spacer, and that the selection of the AUG codon as the ribosome entry site and (in the case of cardioviruses) the initiation site may be determined by the length of this spacer. However, as the position of the oligopyrimidine tract with respect to upstream IRES sequences is also critical for internal initiation, at least in polioviruses (Pilipenko *et al.*, 1992) the relevant parameter may be the distance of the AUG triplet from upstream IRES elements rather than from the oligopyrimidine tract itself.

In this paper we have examined two interconnected questions: what is the influence of the number of pyrimidine residues in the oligopyrimidine tract near the 3′-end of the EMCV IRES on the efficiency of internal initiation; and what is the influence of the distance between this tract and the authentic initiation codon on the discrimination between this authentic initiation codon and nearby AUG triplets? Given the considerations discussed above, the insertion mutants constructed to examine these questions were designed to preserve the G-poor character of the putative spacer element, and the insertions were made at a site which appears to be unstructured according to the results of biochemical probing assays (unpublished observations).

The results reported here are consistent with the hypothesis that the distance of the AUG codon from the oligopyrimidine tract or other IRES segments located further upstream is an important parameter (but not the sole parameter) which determines the selection of the initiation codon. Thus the 11 nt deletion in pΔ11 abrogated initiation at AUG-10 and AUG-11, which is entirely consistent with the distance hypothesis although this model does not explain the decrease in overall initiation efficiency. The other results reported here do not support the view that the whole of this decrease in overall efficiency could be due to the extreme truncation of the oligopyrimidine tract caused by this deletion. One possible explanation is that the context of AUG-12 is less than ideal for internal ribosome entry/initiation.

When the distance of AUG-11 from the upstream motifs was enlarged over the wild-type distance, there was a 15-fold increase in utilization of AUG-10 at the expense of initiation at AUG-11/12, whilst use of AUG-13 was abrogated, again consistent with the distance hypothesis. However, the extreme prediction that all initiation might be switched from AUG-11/12 to AUG-10 in pΔ11/19 was not fulfilled. Attempts to force this switch by increasing the distance even further were not successful, but instead resulted in a decrease in overall translation efficiency with little further change in initiation site preference, a result which could also be interpreted as showing that there is an optimal distance of the AUG triplet from the oligopyrimidine tract and other IRES motifs for maximum efficiency of internal initiation. Nevertheless it is noteworthy that the distance hypothesis was fulfilled in as much as the maximum absolute frequency of initiation at AUG-10 was observed with pΔ11/19s, in

which AUG-10 occupies the same relative position as AUG-11 in the wild-type genome.

It appears as though there must be features which cause AUG-11 to be favoured over AUG-10 even when AUG-11 is at an inappropriately large distance from the upstream motifs. Although all these constructs retained the first five codons of viral polyprotein coding sequence, we consider it unlikely that it is these coding sequences which cause AUG-11 to be the favoured initiation site, since we have found that the efficiency of internal initiation promoted by the wild-type EMCV and TMEV IRESes is not compromised if the viral coding sequences are substituted by other sequences of non-picornaviral origin, unless G-rich codons are inserted immediately after the AUG initiation codon (Hunt *et al.*, 1993).

Distance from the oligopyrimidine tract also appears to play a role in the selection of the initiation site in hepatitis A virus, in which there are two AUG codons separated by 17 and 23 intervening residues from the first C residue of the oligopyrimidine tract. The downstream of these two AUG codons is used preferentially *in vivo* and *in vitro* when there is an intact IRES, but when most of the IRES is deleted to convert the RNA from one translated by internal initiation to an mRNA translated by the scanning ribosome mechanism, the two potential initiation sites are used with approximately equal frequency *in vitro* (Tesar *et al.*, 1992).

The spacing between the oligopyrimidine tract in the poliovirus IRES and the AUG triplet located ~25 nt downstream (at nt 586 in poliovirus type-1) also seems critical for efficient internal initiation and for infectivity, although as this AUG is not used as a functional initiation site but is presumed to be part of the ribosome entry site, the poliovirus system gives less direct information concerning the precision of internal ribosome entry than the EMCV system with its closely spaced AUG codons all of which can potentially serve as translation initiation sites. *In vitro* translation and infectivity assays of a large number of poliovirus type-1 mutants and phenotypic revertants of poliovirus type I showed that the optimum spacing was 22 intervening residues between the UUUC motif and the AUG triplet (the same spacing as in EMCV), and that virtually no translation was detected if the spacing was 14 nt or less (Pilipenko *et al.*, 1992).

The high degree of conservation of the pyrimidine-rich tract at the 5′-end of this segment has suggested that it must play an essential role in the internal initiation mechanism. However, there are viable, if slightly debilitated, deletion mutants of poliovirus and coxsackie B1 virus in which all of this tract has been lost except for the 5′-proximal UUUC motif (Iizuka *et al.*, 1989, 1991; Pilipenko *et al.*, 1992). In addition, site-directed mutagenesis of the 15 nt pyrimidine-rich tract of poliovirus type-2 showed that only the extreme 5′-proximal UUUC motif was required for efficient IRES function in transfected COS cells (Nicholson *et al.*, 1991). Mutations introduced further downstream in the pyrimidine-rich tract had little effect on IRES function *in vivo* (Nicholson *et al.*, 1991), nor in two investigations of *in vitro* translation with the poliovirus type-1 or coxsackie B1 IRES in HeLa cell extracts (Iizuka *et al.*, 1991; Pestova *et al.*, 1991), although they were deleterious in a third such study with poliovirus type-2 (Meerovitch *et al.*, 1991). Even mutation of the 5′-proximal conserved UUUC was without effect when poliovirus type-1 IRES function was assayed in a cell-free system from Krebs II ascites cells (Pestova *et al.*, 1991). Moreover, it should be noted that in many of these experiments the pyrimidines were replaced by G residues, sometimes in tandem array, which could have made the spacer less unstructured, possibly generating new base-pairings, and certainly introducing stronger base-stacking interactions.

Point mutations in the pyrimidine-rich tract of the FMDV IRES decreased the efficiency of internal initiation in the rabbit reticulocyte lysate system, but again these mutations involved introducing G residues (Kuhn *et al.*, 1990). Similar qualifications apply to the finding that a linker replacement of the 5′-part of the EMCV IRES oligopyrimidine tract abrogated internal initiation in the reticulocyte lysate, for in this case the inserted linker had the potential to base-pair with a sequence in the spacer immediately downstream (Jang and Wimmer, 1990). Thus one cannot be certain whether the loss of IRES function was due to the destruction of the pyrimidine-rich tract *per se*, or to the introduction of base-paired secondary structure into the spacer.

Taken as a whole, all these previous data suggest that at most it is only the extreme 5′-proximal part of the tract that is important for internal initiation in the assay systems used, and there is a distinct possibility that even this motif may be dispensable in some systems. In view of this, our finding that internal initiation efficiency *in vivo* or *in vitro* was only slightly compromised when the entire oligopyrimidine tract was replaced by purines (mainly A residues rather than Gs) is perhaps less surprising than one might at first suppose.

Clearly, internal initiation is dependent on sequence motifs located throughout the ~450 nt IRES, irrespective of any role of the pyrimidine-rich tract. It seems likely that some of these motifs may be less critical than others even to the extent of being virtually redundant. It is interesting that a deletion which removed most of the oligopyrimidine tract of poliovirus type-1 (except for the 5′-proximal UUUCC sequence) compromised infectivity only slightly when it was in a Mahoney background, but severely when in the background of the Sabin strain (Iizuka *et al.*, 1989). Since the main difference between the attenuated Sabin strain and the neurovirulent Mahoney strain is thought to be ascribable to a point mutation in the 5′-UTR (i.e. in the IRES element) which affects the efficiency of translation initiation under certain circumstances (Svitkin *et al.*, 1985), this suggests that a longer oligopyrimidine tract may be required if the sequence/structure of the upstream IRES element is itself less than the ideal for maximum efficiency of internal initiation. This in turn suggests the possibility that there might be IRESes, perhaps that of EMCV, in which the upstream sequence motifs are so optimized that even the 5′-proximal vestige of the oligopyrimidine tract is not essential for internal initiation and plays only a rather modest stimulatory role. Thus the oligopyrimidine tract should perhaps be regarded as only of minor importance as a primary determinant of internal ribosome entry, and is probably better regarded as the 5′-proximal part of the unstructured spacer at the 3′-end of the IRES.

## Materials and methods

### *Construction of plasmids*

Plasmid pAΔ258 (Kaminski *et al.*, 1990) was digested with *Eco*RI and *Ssp*I to excise a fragment consisting of all the EMCV sequences (nt 259–848

in the virion RNA numbering system) and the fused poliovirus 2A reporter cistron plus 193 bp of downstream vector sequences. This *Eco*RI–*Ssp*I fragment was subcloned into *Eco*RI–*Sma*I digested pGEM-1 (Promega) to create the parent construct used in this work, pSG1, and was also subcloned between the *Eco*RI and *Sma*I sites of M13mp19 RF for the generation of the two PCR products according to the scheme shown in Figure 1. The PCR derived fragments were end-repaired using both the Klenow fragment of DNA polymerase I and T4 DNA polymerase. The PCR product from amplification of the EMCV 5′-UTR was digested with *Eco*RI and subcloned between the *Eco*RI and *Sma*I sites of pGEM-1 to generate pPCR.UTR, and the other PCR product was digested with *Hin*dIII and inserted between the *Sma*I and *Hin*dIII sites of pGEM-1 to generate pPCR.RG (Figure 1). The inserts of both pPCR.UTR and pPCR.RG were sequenced in their entirety, and then the smaller *Eco*RI–*Hpa*I fragment excised from pPCR.UTR was inserted between the *Eco*RI and *Hpa*I sites of pPCR.RG to generate pΔ11 (Figure 1).

Plasmid pΔ11 was digested with *Hpa*I and treated with calf intestinal phosphatase before ligation with various complementary oligodeoxynucleotides (previously phosphorylated at the 5′-end by incubation with bacteriophage T4 polynucleotide kinase and ATP), to generate the majority of the constructs used in this work which are listed in Figure 1. The other constructs were generated by site-directed mutagenesis: pAUG-10M from pSG1, pΔ11/19s AUG-10M from pΔ11/19s, pΔ11/11PURMAA from pΔ11/11PUR, and pΔ11/11PURMAAc from pΔ11/11PURMAA (see Figure 1). Site-directed mutagenesis was carried out by sub-cloning into a phagemid vector, generation of a single-stranded dU-substituted template strand, and the use of mis-matched oligonucleotides as described by Brierley *et al.* (1989), and based upon the method of Kunkel (1985).

Apart from pΔ11/11PURMAA and pΔ11/11PURMAAc, all the plasmid constructs were also digested with *Bgl*II, the ends in-filled using the Klenow fragment of DNA polymerase I, and the plasmid religated to generate a frame-shift within the poliovirus 2A reporter cistron (Kaminski *et al.*, 1990).

### *In vitro transcription and translation assays*

All plasmids were propagated in *Escherichia coli* TG1 using ampicillin selection. All plasmid DNA was linearized downstream from the reporter gene by digestion with *Bam*HI prior to *in vitro* transcription using bacteriophage T7 RNA polymerase to generate uncapped transcripts as previously described (Kaminski *et al.*, 1990). All transcript RNAs were translated in the micrococcal nuclease-treated rabbit reticulocyte lysate at three different RNA concentrations under the conditions previously described (Kaminski *et al.*, 1990), with [$^{35}$S]methionine (Amersham International) as the radiolabelled amino acid. The translation reaction was subjected to gel electrophoresis using 5–30% polyacrylamide gradient gels (Kaminski *et al.*, 1990). The gels were dried and exposed to Hyperfilm β-max (Amersham International) for several different exposure times and the resultant autoradiographs were scanned using a Transidyne 2955 scanning densitometer.

### *In vivo expression assays*

Expression of 2A from each of the aforementioned constructs was assayed in a transient expression system (Fuerst *et al.*, 1986), essentially as described by Belsham and Brangwyn (1990). BSC40 cells (in 35 mm diameter dishes) were infected with the recombinant vaccinia virus vTF7-3, which expresses T7 RNA polymerase, and the plasmids (5 μg DNA) were introduced after 1 h using lipofectin. After ~18 h, the cells were incubated with methionine-free media for 30 min and then 50 μCi [$^{35}$S]methionine was added to each dish. After 2 h the cells were harvested and total cell extracts were separated on 12.5% polyacrylamide gels. From other samples of the total cell extracts, the poliovirus 2A polypeptide was immunoprecipitated and the immunoprecipitation products were also separated on 12.5% polyacrylamide gels as described previously (Belsham and Brangwyn, 1990; Belsham, 1992).

## Acknowledgements

We thank Mike Howell for much valued advice, Tim Hunt for the gift of reticulocyte lysates, Anne McBride for carrying out structure probing of the EMCV IRES, Ian Brierley for advice on site-directed mutagenesis, and Eckard Wimmer for the gift of antiserum against the poliovirus 2A polypeptide under a collaboration supported by a grant from NATO Scientific Affairs Division. This work was supported by a grant from the Wellcome Trust.

## References

Agol,V.I. (1991) *Adv. Virus Res.*, **40**, 103–180.
Belsham,G.J. (1992) *EMBO J.*, **11**, 1105–1110.
Belsham,G.J. and Brangwyn,J.K. (1990) *J. Virol.*, **64**, 5389–5395.
Brierley,I., Digard,P. and Inglis,S.C. (1989) *Cell*, **57**, 537–547.
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.
Duke,G.M., Hoffman,M.A. and Palmenberg,A.C. (1992) *J. Virol.*, **66**, 1602–1609.
Fuerst,T.R., Niles,E.G., Studier,F.W. and Moss,B. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 8122–8126.
Haller,A.A. and Semler,B.L. (1992) *J. Virol.*, **66**, 5075–5086.
Hunt,S.L., Kaminski,A. and Jackson,R.J. (1993) *Virology*, **197**, 801–807.
Iizuka,N., Kohara,M., Hagino-Yamagishi,K., Abe,S., Komatsu,T., Tago,K., Arita,M. and Nomoto,A. (1989) *J. Virol.*, **63**, 5354–5363.
Iizuka,N., Yonekawa,H. and Nomoto,A. (1991) *J. Virol.*, **65**, 4867–4873.
Jackson,R.J., Howell,M.T. and Kaminski,A. (1990) *Trends Biochem. Sci.*, **15**, 477–483.
Jang,S.K. and Wimmer,E. (1990) *Genes Dev.*, **4**, 1560–1572.
Jang,S.K., Krausslich,H.-G., Nicklin,M.J.H., Duke,G.M., Palmenberg,A.C. and Wimmer,E. (1988) *J. Virol.*, **62**, 2636–2643.
Kaminski,A., Howell,M.T. and Jackson,R.J. (1990) *EMBO J.*, **9**, 3753–3759.
Kozak,M. (1989) *J. Cell Biol.*, **108**, 229–241.
Kozak,M. (1992) *Crit. Rev. Biochem. Mol. Biol.*, **27**, 385–402.
Kuge,S. and Nomoto,A. (1987) *J. Virol.*, **61**, 1478–1487.
Kuhn,R., Luz,N. and Beck,E. (1990) *J. Virol.*, **64**, 4625–4631.
Kunkel,T.A. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 488–492.
Meerovitch,K. and Sonenberg,N. (1993) *Semin. Virol.*, **4**, 217–227.
Meerovitch,K., Nicholson,R. and Sonenberg,N. (1991) *J. Virol.*, **65**, 5895–5901.
Nicholson,R., Pelletier,J., Le,S.Y. and Sonenberg,N. (1991) *J. Virol.*, **65**, 5886–5894.
Pestova,T.V., Hellen,C.U.T. and Wimmer,E. (1991) *J. Virol.*, **65**, 6194–6204.
Pilipenko,E.V., Blinov,V.M., Chernov,B.K., Dimitrieva,T.M. and Agol,V.I. (1989) *Nucleic Acids Res.*, **17**, 5701–5711.
Pilipenko,E.V., Gmyl,A.P., Maslova,S.V., Svitkin,Y.V., Sinyakov,A.N. and Agol,V.I. (1992) *Cell*, **68**, 119–131.
Poyry,T., Kinnunen,L. and Hovi,T. (1992) *J. Virol.*, **66**, 5313–5319.
Pritchard,A.E., Calenoff,M.A., Simpson,S., Jensen,K. and Lipton,H.L. (1992) *J. Virol.*, **66**, 1951–1958.
Sangar,D.V., Newton,S.E., Rowlands,D.J. and Clarke,B.E. (1987) *Nucleic Acids Res.*, **15**, 3305–3315.
Svitkin,Y.V., Maslova,S.V. and Agol,V.I. (1985) *Virology*, **147**, 243–252.
Tesar,M., Harmon,S.A., Summers,D.F. and Ehrenfeld,E. (1992) *Virology*, **186**, 609–618.

*Received on October 27, 1993; revised on December 23, 1993*