The EMBO Journal vol.11 no.3 pp.1105-1110, 1992

Dual initiation sites of protein synthesis on foot-and-mouth disease virus RNA are selected following internal entry and scanning of ribosomes *in vivo*

G.J.Belsham

AFRC Institute for Animal Health, Pirbright Laboratory, Pirbright, Woking, Surrey GU24 0NF, UK

Communicated by R.J.Jackson

The initiation of protein synthesis on foot-and-mouth disease virus RNA occurs at two sites separated by 84 nucleotides. Immediately upstream from the first of these sites is the internal ribosome entry site (IRES), which directs the translation of this RNA to be cap-independent. The utilization of these two initiation sites has been examined using artificial fusion genes *in vivo* under a variety of conditions. Additional in-frame AUG codons have been introduced between these two authentic start sites to determine the mechanism by which ribosomes recognize the second start site. The results indicate that following internal entry of ribosomes on the 5′ side of the first initiation codon, many fail to initiate protein synthesis at this position and scan along the RNA to the second initiation site. In the presence or absence of the IRES both initiation sites are efficiently used but the utilization of the two sites is slightly biased towards the second initiation site by the IRES. Furthermore, in the presence of the IRES, protein synthesis initiates at both sites independently of the activity of the cap-binding complex.

Key words: internal ribosome entry site (IRES)/picornavirus/ribosome scanning/translation

Introduction

The initiation of protein synthesis in eukaryotic cells usually occurs at the first AUG codon of a mRNA following ribosome binding in the vicinity of the 5′ cap structure and 'scanning' (Kozak, 1989). It is now established that the initiation of protein synthesis on picornavirus RNA occurs by a separate mechanism [see reviews by Sonenberg and Pelletier (1989); Jackson *et al.* (1990)]. The fact that the viral RNAs lack the cap structure suggested that this was probable. Furthermore, the 5′ non-coding regions (NCR) of these RNAs are long (650–1300 bases) and contain multiple AUG codons which are poorly conserved between related viruses and are unlikely to have a coding function. Indeed, in poliovirus none of these upstream AUGs are required for viability although the modification of AUG 7 does produce a modified phenotype (Pelletier *et al.*, 1988). Regions of ~450 bases from the 5′ NCRs of foot-and-mouth disease virus (FMDV) (Belsham and Brangwyn, 1990; Kuhn *et al.*, 1990) and encephalomyocarditis virus (EMCV) (Jang and Wimmer, 1990) which are adjacent to the initiation codons direct internal initiation of protein synthesis within artificial bicistronic mRNAs *in vivo*. This region is referred to as the internal ribosome entry site (IRES). The presence of the IRES immediately adjacent to an open reading frame (ORF) also directs the translation of the ORF when cap-dependent translation is inhibited (e.g. when the p220 component of the cap-binding complex is cleaved) (Belsham and Brangwyn, 1990). Secondary structure predictions for the IRES from EMCV and FMDV are extremely similar (Pilipenko *et al.*, 1989) although the sequence identity is only ~50%. It is probable that the two structures work by a very similar mechanism.

Studies on the selection of the initiation sites in EMCV (strain R1) in rabbit reticulocyte lysate have shown that the IRES very precisely directs the ribosomes to AUG 11, while AUG 10 only 8 bases upstream, is not used (Kaminski *et al.*, 1990). Both AUG 10 and 11 have sequence contexts which are satisfactory for efficient initiation according to the consensus sequence (A/GXXAUGA/G) derived by Kozak (1989); there is extremely limited scope for ribosome scanning following entry between the two AUG triplets. In FMDV initiation of protein synthesis occurs at two distinct sites, separated by 84 bases, both *in vivo* and *in vitro* (Sangar *et al.*, 1987) so that two forms of the leader (L) protein are produced, namely Lab and Lb. The sequence context of the first initation site is generally poor while that of the second is generally good in different strains of FMDV. In certain strains of FMDV an upstream AUG is also present in a position analagous to that of AUG 10 in EMCV (R1) although again, in contrast to the EMCV situation, the context is poor compared with the Kozak consensus sequence.

A section of the poliovirus 5′ NCR has also been shown to direct internal initiation of protein synthesis (Pelletier and Sonenberg, 1988; Trono *et al.*, 1988), in this case the element terminates ~100 bases upstream from the initiation codon (nucleotide 743). The mechanism by which ribosomes reach the initiation codon has not been established. The introduction of a hairpin structure at nucleotide 631 inhibits translation *in vitro* (Pelletier and Sonenberg, 1988) and insertion of a 72 base sequence containing an AUG codon into this region of the virus produces a small plaque size phenotype (Kuge *et al.*, 1989). However, both of these modifications create major changes in the sequence and their mode of action is not clear. There is little apparent similarity in sequence or secondary structure between the IRES of EMCV/FMDV and that in poliovirus.

This study is directed towards determining the mechanism by which initiation of protein synthesis occurs at the two distinct sites of FMDV within cells. In particular the question of whether initiation at the second start site follows direct binding of the ribosomes immediately upstream of this site or follows 'leaky' scanning past the first initiation codon is addressed.

Results

To facilitate the examination of the initiation at the two start sites of FMDV under a variety of conditions, it was necessary to construct synthetic fusion genes since the intact L protein inhibits cap-dependent protein synthesis. This effect probably results from its activity to initiate the cleavage (and inactivation) of the p220 cap-binding complex component (eIF-4F) (Belsham and Brangwyn, 1990; Devaney *et al.*, 1988). An *Eco*RI restriction site closely follows the second initiation codon in the cDNA of certain strains of FMDV and had been constructed into the CA103 cassette (O1K cDNA cassette) (see Figure 1 and Belsham *et al.*, 1990). Ligation of the FMDV sequence up to this site to the cleaved *Eco*RI site within the chloramphenicol acetyl transferase (CAT) gene results in the production of a fusion gene which encodes L−CAT fusion proteins which can be immunoprecipitated with anti-CAT antibodies (see Figure 2). The product made by initiation at the first (Lab) start site was termed LabCAT and the smaller product generated from the Lb start site LbCAT.

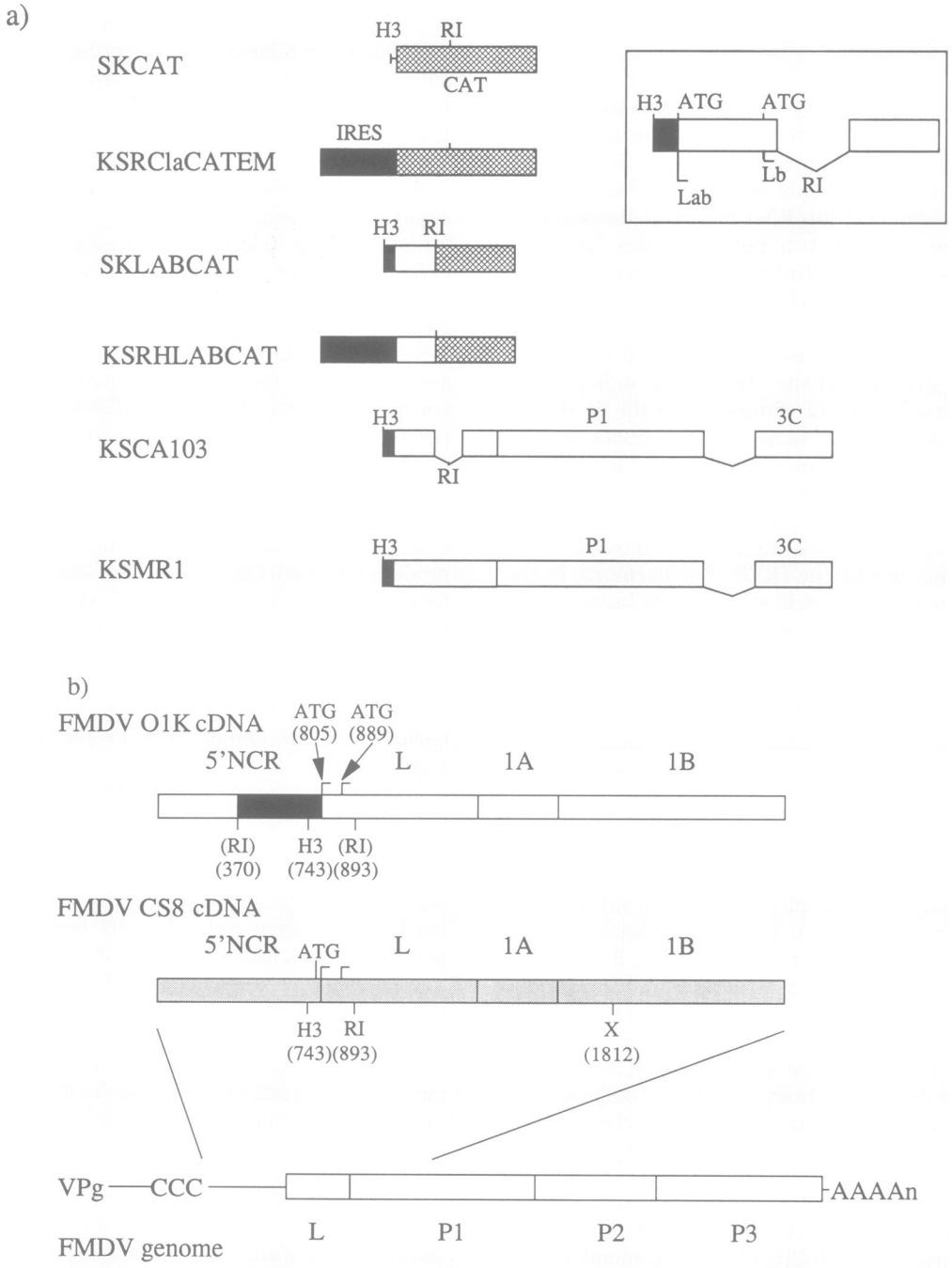


Fig. 1. A. Construction of L−CAT fusion genes. Each construct is preceded by the T7 promoter from within pKS+ or pSK+ as indicated. Sequences from the CAT gene are indicated in cross hatching and the IRES sequence from FMDV O1K cDNA is indicated by the solid fill. Coding sequences from O1K FMDV cDNA for the P1 structural protein precursor and the 3C protease are indicated by open blocks. The inset indicates the positions of the two initiation codons in the L coding sequence (type O1K). **B**. Map of FMDV and the regions of the O1K and CS8 cDNA used in this study. Restriction sites used in the constructions are indicated in abbreviated form; H3, *Hin*dIII; RI, *Eco*RI; X, *Xba*I. Both *Eco*RI sites indicated in the O1K cDNA, (in brackets) were introduced by site-directed mutagenesis (see Belsham and Brangwyn, 1990). CS8 cDNA is indicated by the speckled fill. Nucleotide numbers, based on the O1K numbering system (from the polycytosine tract) (Forss *et al.*, 1984), are indicated in brackets.

All cassettes were cloned into vectors under the control of the bacteriophage T7 promoter and were assayed in a transient expression assay using cells infected with the recombinant vaccinia virus vTF7-3 (Fuerst *et al.*, 1986) which expresses the T7 RNA polymerase. Each cassette was cloned in two forms; either alone (SK prefix) or following the intact FMDV IRES (KSRH or KSRCla prefix). The intact CAT gene was used as a control (see Figure 1A). In all constructs containing the FMDV initiation sites, the construct lacking the IRES does contain the sequence from nucleotide 743 (*Hin*dIII site) to the first initiation site at 805 (Figure 1B). The IRES containing constructs include the sequence from the *Eco*RI site introduced at nucleotide 370 (Belsham and Brangwyn, 1990) through to the initiation codons. The plasmids were assayed under three conditions: (i) alone, (ii) co-transfected with KSMR1 (intact L coding sequence) or (iii) co-transfected with KSCA103 (interrupted L coding sequence) (see Figure 1A). Cells were incubated with [^{35}S]methionine and CAT related proteins were immunoprecipitated and analysed by SDS−PAGE (Figure 2). Expression of the CAT protein from the plasmid SKCAT was greatly inhibited in the presence of KSMR1 but was largely unaffected by co-transfection with KSCA103 (lanes 1−3), a pattern indicative of cap-dependent translation. However, expression of CAT continued efficiently from plasmid KSRClaCATEM, which contains the IRES, under each condition (lanes 4−6) as expected from previous studies (Belsham and Brangwyn, 1990); the synthesis of CAT reflected the level of CAT activity in these extracts (not shown). The L−CAT fusion gene plasmid lacking the IRES, SKLABCAT, produced the two proteins, LabCAT and LbCAT which indicates synthesis initiating at the two different sites (lanes 7−9). The production of both proteins was abolished in the presence of KSMR1 but persisted in the presence of KSCA103, again indicative of cap-dependent translation. The partner plasmid containing the IRES, KSRHLABCAT, directed the production of the same proteins and the synthesis was resistant to the action of the L protein (lanes 10−12). A modest bias towards the use of the second start site, indicated by the production of more LbCAT than LabCAT, was evident in the presence of the IRES (compare lanes 7 and 10). In the presence of the IRES the abolition of cap-dependent translation had little further effect on the ratio of initiation at the two initiation sites (see lanes 10 and 11).

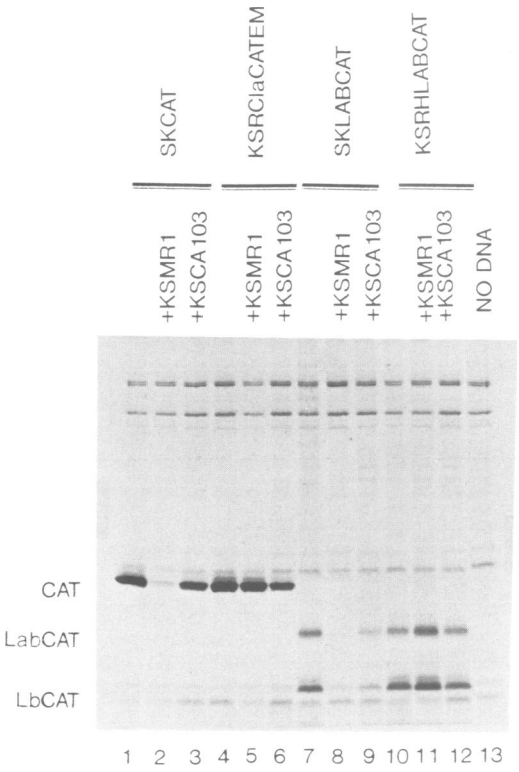
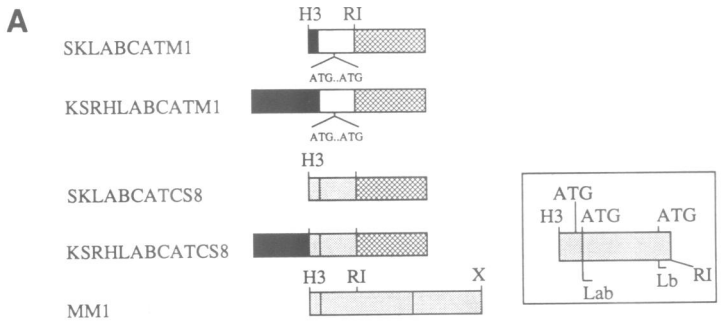


Fig. 2. Effect of the IRES and inhibition of cap-dependent translation on the expression of CAT related proteins. Plasmids were transfected into vTF7-3 infected BSC40 cells alone or with KSMR1 or KSCA103 as indicated. CAT related proteins from the [^{35}S]methionine labelled cells were immunoprecipitated and analysed by SDS−PAGE and autoradiography. The mobilities of the full length CAT protein and the two fusion gene products LabCAT and LbCAT are indicated.

In order to examine further the properties of the site selection system, additional plasmids were constructed and are shown in Figure 3. The homologous *Hin*dIII−*Eco*RI sequence from type C FMDV cDNA (strain CS8) was introduced into the same L−CAT fusion system (these



B wt O1K ATGAATACAACTGACTGTTTTATCGCTTTGGTACAG...ATG

M1 ATGAATACAACTGACTGTTTTATGGCTATGGTACAG...ATG

Fig. 3. Construction of variant forms of the L−CAT fusion genes. Panel **A** indicates the structure of L−CAT fusion genes containing type C FMDV (strain CS8) initiation sites (see inset) or mutagenized O1K sequences. Other details are as in Figure 1, the CS8 sequences are indicated by a speckled fill. Panel **B** shows the sequences from the wild-type and mutagenized construct (M1). The two new initiation codons are underlined.

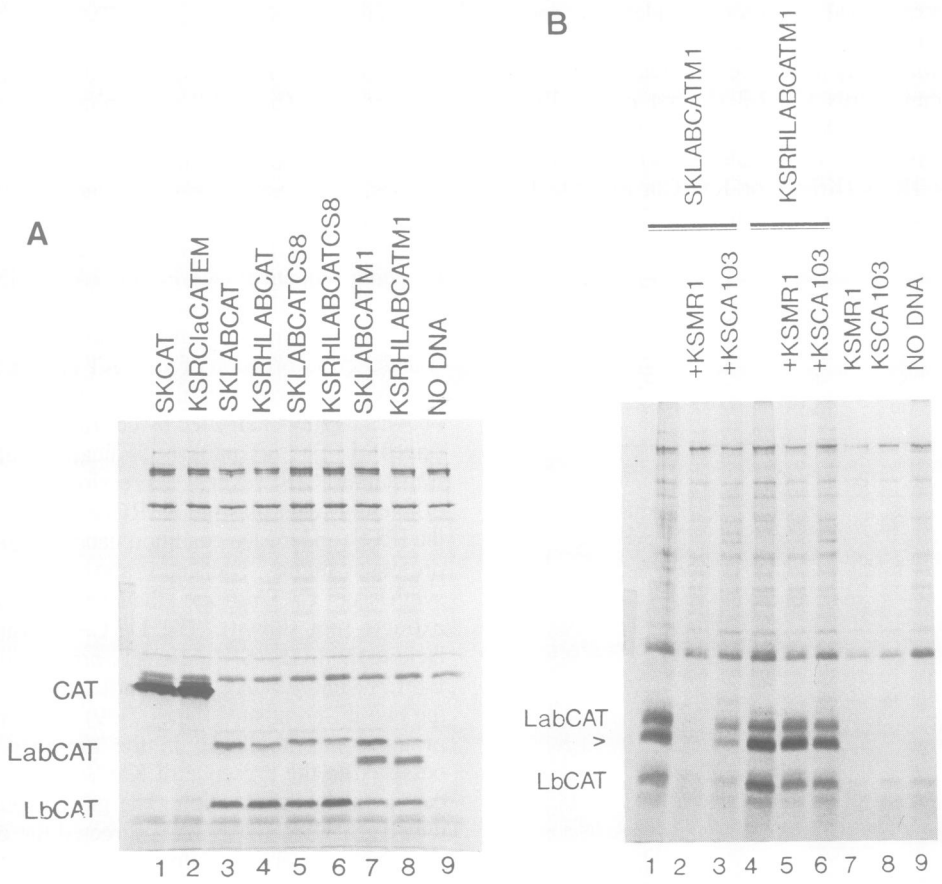


Fig. 4. Effect of additional ATG codons on the usage of the two authentic initiation codons. Transient expression of the CAT related proteins was assessed by immunoprecipitation and analysis by SDS−PAGE and autoradiography. Panel **A**. The plasmids indicated were assayed alone as described in Figure 2. Panel **B**. Indicated plasmids were assayed alone, with KSMR1 or with KSCA103 as in Figure 2. The position of the novel L−CAT fusion products from the M1 mutants is arrowed.

plasmids are suffixed by CS8). This sequence encodes an addtional AUG codon eight nucleotides upstream of the first initiation site (see Figure 5). This AUG is in the same position as the AUG 10 in EMCV which is not used in the presence of the IRES but is recognized in the absence of this element (Kaminski *et al.*, 1990). Thus it was thought that in the absence of the IRES (plasmid SKLABCATCS8), the initiation at the authentic start sites may be attenuated. In fact the pattern of initiation from this construct appeared very similar to that for the O1K version (Figure 4A, compare lanes 3 and 5). The LabCAT product containing the amino-terminus of the CS8 Lab protein runs slightly more slowly than the analogous product containing the O1K Lab amino-terminus (Figure 4A, compare lanes 3, 4 and 5, 6) presumably as a result of the sequence differences between these two proteins. The ratio of initiation at the two sites in the CS8 sequence is slightly biased by the presence of the IRES (in plasmid KSRHLABCATCS8) towards the production of the LbCAT product (Figure 4A, lanes 5 and 6) as observed for the O1K construct. It appears that the upstream mini-cistron is ignored when the RNA is translated under both IRES-dependent or cap-dependent conditions. This probably reflects the poor sequence context of this upstream AUG and the fact that the ORF is very short i.e. it terminates after three codons. As expected the synthesis of the L−CAT proteins from SKLABCATCS8 was abolished in the presence of KSMR1 but maintained from KSRHLABCATCS8 under these conditions (data not shown).

Since it is probable that, in the presence of the IRES, ribosomes bind very close to the first FMDV initiation codon, as in EMCV, two different mechanisms can be envisaged whereby ribosomes reach the second initiation site. One possibility is that the ribosomes could also bind immediately upstream of the second initiation site i.e. there are two distinct entry sites for ribosomes. The second mechanism would be for the ribosomes to scan past the first initiation site, which has a rather poor context, and then to initiate translation at the next AUG, i.e. 'leaky' scanning. In order to distinguish between these possibilities, two new in-frame ATG codons were introduced into the coding sequence cDNA between the two start sites of the O1K sequence (Figure 3). These codons were introduced by site-directed mutagenesis and required just two point mutations. The first new codon is followed by a G residue but the −3 position is sub-optimal while the second new codon has a purine at both −3 and +4 positions and hence has a good match to the consensus sequence. The mutated sequences were reconstructed back into the CAT fusion vectors (plasmids suffixed M1) and assayed as above. In the absence (SKLABCATM1) or presence of the IRES (KSRHLABCATM1), the introduction of the two new start sites led to the production of two new L−CAT products which migrate on SDS−PAGE between the LabCAT and

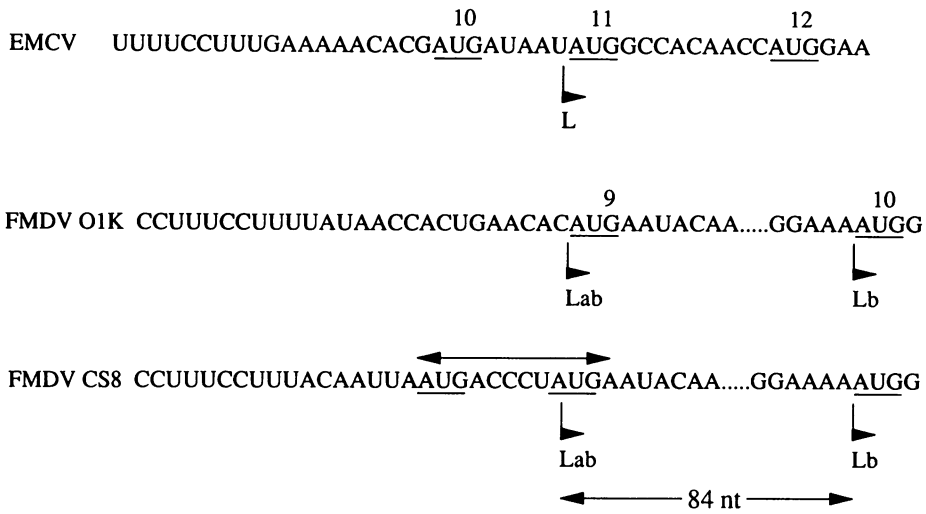


Fig. 5. Sequence comparison of EMCV and FMDV (strains O1K and CS8) around the initiation codons. The start sites for the L protein of EMCV and the two forms of the L protein (Lab and Lb) are indicated. The upstream minicistron in the CS8 sequence is shown by the double-headed arrow. The full sequence of the FMDV CS8 cDNA will be published separately (C.Escarmis, J.Diez, M.Medina and E.Domingo, unpublished results).

LbCAT proteins (Figure 4A, lanes 7 and 8) (on the original autoradiograph, the two new species are resolved but may merge in the printing process: they differ by only two amino acids). Less LbCAT product is produced from these constructs (Figure 4A, compare lanes 7, 8 with 3, 4). However, the bias towards the LbCAT product induced by the presence of the IRES is maintained (Figure 4A, lanes 7 and 8). Thus it appears that the internal entry of ribosomes to initiate protein synthesis at the first site is followed by ribosome scanning prior to initiation occurring at the second site. The IRES-dependent initiation at all four sites from plasmid KSRHLABCATM1 was unaffected by the co-expression of the FMDV L protein (Figure 4B, compare lanes 4 and 5) but was abolished under these conditions in the absence of the IRES (Figure 4B, compare lanes 1 and 2).

Discussion

The assay system developed in this study allowed examination of the utilization of the two initiation sites in FMDV under a variety of conditions. A key question addressed was whether initiation of protein synthesis at the second initiation site followed direct ribosome entry adjacent to this site, analagous to the situation for the first initiation site, or followed ribosome scanning from a single ribosome entry site. These studies showed that introduction of two new in-frame initiation codons between the two start sites lead to the production of two new L−CAT proteins and diminished production of the LbCAT protein. The most likely explanation for this result is that ribosomes scan through the region between the initiation sites following internal entry just upstream of the first initiation codon at which only a fraction of the ribosomes initiate synthesis. Subsequently some of the ribosomes initiate translation at the two new start sites and hence less are available for initiation at the Lb start site. This is consistent with the fact that the sequence context for the first initiation codon is rather poor in comparison with the Kozak consensus sequence, whereas at the second site the match is rather better (see Figure 5). In the absence of the IRES the mRNA behaves like a normal capped cellular mRNA and initiation occurs according to the normal rules. Hence the first AUG is only partially recognized and a significant fraction of the ribosomes scan through this site and reach the second site. In the presence of the IRES the ribosomes enter internally, presumably just to the 5′ side of the first initiation site (within eight bases) as in EMCV, and then proceed to scan and initiate with essentially the same characterstics as for the mRNA lacking the IRES. A slight bias towards initation at the second start site was consistently observed in the presence of the IRES compared with the constructs lacking this element. It may be that the IRES makes ribosome scanning more 'leaky' around the first AUG codon since the ribosomes will have only just commenced the process whereas in the constructs lacking the IRES they will have scanned through ~60 bases at this point. When the FMDV L protein is expressed, the initiation events are inhibited in the absence of the IRES but continue in the presence of the IRES which confirms that the ribosome entry (but not the scanning process) occurs by separate mechanisms in the two different situations. Since it appears that ribosomes do not bind at a second site it is of interest that a second pyrimidine rich tract is present upstream of the second initiation codon (see Sangar *et al.*, 1987), the role of this sequence is unknown but it is found close to the ribosome entry site in poliovirus, EMCV and FMDV. It can be argued that the experiments presented do not preclude the possibilty that some internal entry of ribosomes occurs at a second site (adjacent to the second start site) but the data suggest that this has to be a minor component since only a modest change in utilization of the two sites occurs in the presence or absence of the IRES.

It was thought that the presence of the upstream AUG in the CS8 sequence may have inhibited the translation from the authentic initiation sites in the absence of the IRES; however, it was found that the pattern and efficiency of site utilization in the presence and absence of the IRES closely followed that observed with the O1K constructs which lack this extra AUG. Thus it appears that the upstream AUG is ignored; this probably reflects the poor context of this upstream AUG and also perhaps the fact that the ORF from this site is extremely short (only two amino acids would be joined) (see Figure 5). It may be that secondary structure

within the RNA assists selection of the first authentic initiation by slowing the passage of ribosomes at this point since the context of this site is also poor. These features meant that these experiments did not allow the verification in FMDV of the ribosome entry site predicted from the work with EMCV by Kaminski *et al.* (1990). However, other studies using plasmids provided from R.Jackson's laboratory confirm that the selection of AUG codons identified in EMCV using *in vitro* translation assays also applies in the intracellular transient expression system employed here (G.J.Belsham, A.Kaminski and R.Jackson, unpublished observations).

It is not clear why FMDV requires the synthesis of two different forms of the L protein, but this feature is maintained in all seven serotypes of the virus. It has been shown (Vakharia *et al.*, 1987; Devaney *et al.*, 1988; M.Medina and G.J.Belsham. unpublished observations) that the Lb protein has protease activity to cleave the L/P1−2A junction (in *cis* and *trans*) and also will initiate the cleavage of p220 and (hence) the inhibition of cap-dependent translation. These are all of the activities so far attributed to the L protein, thus no distinct activity of Lab is known.

Materials and methods

Plasmid DNA constructions were performed using standard methods (Sambrook *et al.*, 1989). Plasmids SKCAT, SKRClaCAT, KSMRHMR1.Cla, KSMR1 and KSCA103 have been published previously (Belsham and Brangwyn, 1990; Belsham *et al.*, 1990). Plasmid pMM1 (containing the *Hin*dIII−*Xba*I fragment of FMDV CS8 cDNA including the two initiation sites was kindly given by Miguel Medina and Esteban Domingo (Madrid). Plasmid KSRClaCATEM was constructed as follows: pKSRCla (pKS+ containing the *Eco*RI−*Cla*I IRES-containing fragment from KSMRHMR1.Cla) was linearized with *Eco*RI, blunt ended and religated to produce KSRClaEM; this was then cut with *Hin*dIII and *Hin*cII and ligated to the *Hin*dIII−*Bam*HI (blunt ended) fragment from SKRClaCAT containing the CAT gene and the 3′ end of the IRES; plasmids with the correct structure (containing the complete IRES and the CAT gene) were isolated. The plasmids SKCAT and KSRClaCATEM contain unique *Eco*RI and *Hin*dIII sites and the variants of the FMDV initiation site cDNA were introduced into these vectors using these sites (see Figures 1 and 3).

Mutagenesis of the FMDV O1K cDNA was performed on the *Hin*dIII−*Eco*RI fragment from KSCA103 cloned into similarly digested M13tg130 (Amersham) using the synthetic oligonucleotide GTTTTATGGCTATGGTACAGGC (bases changed from the wild-type sequence are underlined) with a T7-GEN *in vitro* mutagenesis kit (United States Biochemical). The oligonucleotide was synthesized on an Applied Biosystems 381A machine and used without further purification. M13 phage DNAs were screened by hybridization with the mutagenic oligonucleotide and positive clones were sequenced to confirm the presence of the correct mutations. The *Hin*dIII−*Eco*RI fragment from the mutant (M1) Rf DNA was constructed into both SKCAT and KSRClaCATEM as indicated above.

Transient expression studies were performed as described previously (Belsham and Brangwyn, 1990) using lipofectin (BRL) mediated transfection of plasmid DNA (5 μg) into vTF7-3 (Fuerst *et al.*, 1986) infected BSC40 cells (35 mm dish). After 20 h, cells were incubated with [^{35}S]methionine (50 μCi/dish), cell extracts were prepared, CAT related proteins were immunoprecipitated with polyclonal anti-CAT antibodies (5prime-3prime Inc) and proteins were analysed by SDS−PAGE (Laemmli, 1970) using 12.5% acrylamide gels.

Acknowledgements

I thank Julia Brangwyn for excellent technical assistance, Doros Panayi for photography, Martin Ryan for synthesis of the oligonucleotide and Andrew King for encouragement and critical reading of the manuscript. I also thank Miguel Medina and Esteban Domingo for plasmid pMM1 and the sequence of the FMDV strain CS8 cDNA prior to publication.

References

Belsham,G.J. and Brangwyn,J.K. (1990) *J. Virol.*, **64**, 5389−5395.
Belsham,G.J., Brangwyn,J.K., Ryan,M.D., Abrams,C.C. and King,A.M.Q. (1990) *Virology*, **176**, 524−530.
Devaney,M.A., Vakharia,V.N., Lloyd,R.E., Ehrenfeld,E. and Grubman,M.J. (1988) *J. Virol.*, **62**, 4407−4409.
Forss,S., Strebel,K., Beck,E. and Schaller,H. (1984) *Nucleic Acids Res.* **12**, 6587−6601.
Fuerst, T.R., Niles,E.G., Studier,F.W. and Moss,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8122−8126.
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.
Kaminski,A., Howell,M. and Jackson,R. (1990) *EMBO J.*, **9**, 3753−3759.
Kozak,M. (1989) *J. Cell Biol.*, **108**, 229−241.
Kuge,S., Kawamura,N. and Nomoto,A. (1989) *J. Virol.*, **63**, 1069−1075.
Kuhn,R., Luz,N. and Beck,E. (1990) *J. Virol.*, **64**, 4625−4631.
Laemmli,U.K. (1970) *Nature*, **227**, 680−685.
Pelletier,J. and Sonenberg,N. (1988) *Nature*, **334**, 320−325.
Pelletier,J., Flynn,M.E., Kaplan,G., Racaniello,V. and Sonenberg,N. (1988) *J. Virol.*, **62**, 4486−4492.
Pilipenko,E.V., Blinov,V.M., Chernov,B.K., Dimitrieva,T.M. and Agol,V.I. (1989) *Nucleic Acids Res.*, **17**, 5701−5711.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sangar,D.V., Newton,S.E., Rowlands,D.J. and Clarke,B.E. (1987) *Nucleic Acids Res.*, **15**, 3305−3315.
Sonenberg,N. and Pelletier,J. (1989) *BioEssays*, **11**, 128−132.
Trono,D., Andino,R. and Baltimore,D. (1988) *J. Virol.*, **62**, 2291−2299.
Vakharia,V.N., Devaney,M.A., Moore,D.M., Dunn,J.J. and Grubman,M.J. (1987) *J. Virol.*, **61**, 3199−3207.

Received on October 21, 1991; revised on November 19, 1991