Identification of the minimal sequences from the *Rhopalosiphum padi* virus 5’ untranslated region required for internal initiation of protein synthesis in mammalian, plant and insect translation systems

Elisabetta Groppelli¹, Graham J Belsham²³ and Lisa O Roberts¹*  

¹ School of Biomedical and Molecular Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK  
² BBSRC Institute for Animal Health, Ash Road, Pirbright, Surrey GU24 0NF, UK.  
³ Present address: National Veterinary Institute, Technical University of Denmark, Lindholm, 4771 Kalvehave, Denmark  

* Author for correspondence. Tel: 44 1483 686499; Fax: 44 1483 300374; e-mail: l.roberts@surrey.ac.uk  

Running title: Deletion analysis of the RhPV 5’ UTR  

Word count (summary): **141**  
Word count (body text, inc. figure legends, not inc. cited references): **2584**  
No. of figures: **3**
SUMMARY

*Rhopalosiphum padi* virus (RhPV) is a member of the *Dicistroviridae*. The genomes of these viruses contain two open reading frames, each preceded by distinct internal ribosome entry site (IRES) elements. The RhPV 5’ IRES is functional in mammalian, insect and plant translation systems and can form 48S initiation complexes *in vitro* with just the mammalian initiation factors eIF2, eIF3 and eIF1. Large regions of the 5’ UTR can be deleted without affecting initiation complex formation. We have now defined the minimal sequences required for directing internal initiation in mammalian (RRL), plant (WGE) and insect (Sf21 cells) translation systems. A fragment (nt 426-579) from the 3’ portion of the 5’UTR can direct translation in each of these translation systems. In addition, a distinct region (nt 300-429) is also active. Thus, unstructured regions within the 5’UTR seem to be critical for IRES function.

Initiation of translation on most cellular mRNAs occurs through a cap-dependent mechanism which involves the binding of the cap-binding complex (eukaryotic initiation factor complex 4F) to the cap structure (m\(^7\)GpppN) at the 5’ end of the mRNA, followed by recruitment of the 40S ribosomal subunit (reviewed in Hershey and Merrick, 2000). However, a number of viral and cellular mRNAs use a cap-independent mechanism termed internal initiation (reviewed in Belsham and Jackson, 2000; Carter et al., 2000; Hellen and Sarnow, 2001). In this case, initiation of protein synthesis is directed by an internal ribosome entry site (IRES) element within the 5’ untranslated region (UTR) of the mRNA. IRES elements have been
identified in viral and cellular mRNAs, however, there is much variation in their structures and requirements for initiation factors.

Some of the best-characterised IRES elements are located in the 5’ UTRs of picornavirus genomes. The picornavirus IRES elements have generally been grouped into two major classes according to their secondary structure and activity in vitro. The cardio-and aphthovirus IRES elements form one group, which function very efficiently in the rabbit reticulocyte lysate (RRL) system. In contrast, the entero- and rhinovirus IRES elements make up the second group which functions poorly in RRL. However, the activity of these IRES elements in RRL is greatly enhanced if it is supplemented with HeLa cell extracts (Brown & Ehrenfeld, 1979; Dorner et al., 1984). A third type of picornavirus IRES element has recently been described in the porcine teschovirus-1, porcine enterovirus-8 and simian virus 2 genomes; these IRES elements show significant similarity to the IRES from hepatitis C virus (a member of the Flaviviridae) (Pisarev et al., 2004; Chard et al., 2006a; 2006b). It is now clear that picornavirus IRES elements require cellular (trans-acting) factors for their function, in addition to the canonical translation initiation factors, and that these factors may play an important role in IRES (and virus) tropism. Examples of these proteins are the La autoantigen, polypyrimidine tract-binding protein (PTB), poly(C)-binding protein and unr (upstream of N-Ras) (reviewed in Belsham & Jackson, 2000).

RhPV belongs to the Dicistroviridae family that also includes Cricket paralysis virus (CrPV), Plautia stali intestinal virus (PSIV) and Drosophila C virus (DCV). RhPV has a single-stranded, positive sense RNA genome which contains two separate open reading frames (ORFs; Moon et al., 1998). ORF1 encodes the non-structural proteins and ORF2 encodes the structural proteins. The 5’ UTRs and the intergenic regions (IGRs) of the dicistrovirus genomes each contain an IRES element
(Sasaki & Nakashima, 1999; Domier et al., 2000; Wilson et al., 2000a; Woolaway et al., 2001). The IGR IRES of CrPV does not require any of the initiation factors for the assembly of an initiation complex on the mRNA, which occurs at a non-AUG codon (Wilson et al., 2000b). The properties of the 5’ IRES elements are very different and they direct translation initiation from AUG codons (Sasaki and Nakashima, 1999; Domier et al., 2000; Wilson et al., 2000a; Woolaway et al., 2001). We have previously shown that the RhPV 5’ IRES functions in mammalian (RRL), insect (Drosophila lysate, Sf21 lysate and Sf21 cells) and plant (wheat germ extract) systems (Woolaway et al., 2001; Kubick et al., 2003; Royall et al., 2004). The ability of the IRES to function in all three systems suggested a mechanism of internal initiation with a less specific requirement for translation initiation factors. Indeed, we have recently shown that the RhPV 5’ IRES requires only the mammalian initiation factors eIF2, eIF3 and eIF1 to form 48S complexes in vitro (Terenin et al., 2005). The factors eIF1A and eIF4F stimulate the assembly of 48S complexes on the RhPV RNA but are not essential. We also found that it was possible to delete large regions of the 5’ UTR without affecting initiation complex formation. However, deletion of an unstructured region in the 380 nt proximal to the initiation codon significantly reduced 48S complex formation. As this previous work was carried out using only mammalian factors, we set out to define the minimal RhPV sequences required for directing internal initiation in translation systems from mammals (RRL), plants (WGE) and insects (Sf21 cells). Here, we report that a small fragment from the 3’ end of the RhPV 5’ UTR, corresponding to nt 425 to 579, still retains significant IRES activity in all three translation systems whereas nt 475 to 579 lacked IRES activity. Intriguingly, a fragment corresponding to nts 300 to 429 is also capable of directing internal initiation in all three systems. The ability of different regions or fragments of
the RhPV 5’ UTR to display IRES activity is distinct from viral IRES elements previously described and is reminiscent of some cellular IRES elements.

In order to define the minimal sequence that can direct internal initiation of protein synthesis in different translation systems, we created truncated versions of the cDNA corresponding to the RhPV 5’ UTR and inserted the fragments between two reporter sequences. Fragments lacking sequences from either the 5’ or 3’ end were obtained by PCR using specific primers (supplementary information). All primers contained BamHI restriction enzyme sites to facilitate cloning into the pGEM-CAT/LUC vector between the two open reading frames, as previously described (van der Velden et al., 1994). The plasmid pGEM-CAT/RhPV∆1/LUC, which contains the entire 5’ UTR (nt 1-579; Woolaway et al., 2001), was used as template for each PCR. Deletions of 50, 100, 150, 200, 250 and 300 nt from the 5’ end were created (Figure 1). Similarly, deletions of 50, 100, 150 and 200 nt from the 3’ end of the RhPV IRES were also generated. Further deletions at the 3’ end were not made since we have shown previously that a 200 nt deletion from the 3’ end significantly reduced IRES activity (Woolaway et al., 2001). All the resulting constructs were verified by sequencing.

To examine the activity of the RhPV sequences in RRL, the plasmids were used to programme in vitro coupled transcription/translation (TNT) reactions (Promega). Reactions contained [35S]-methionine and the products were analysed using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography. LUC expression was also measured by luciferase assay (Promega) according to the manufacturer’s instructions. Plasmid pGEM-CAT/LUC which lacks any IRES sequence, and pGEM-CAT/∆1/LUC containing the wt RhPV IRES, were used as negative and positive controls, respectively. All the constructs expressed
similar levels of the upstream cistron, CAT as expected (Figure 2A). All the constructs containing regions of the RhPV 5’ UTR were also able to direct expression of LUC, although with different efficiencies. Deletion of up to 300 nt from the 5’ end had little effect on IRES activity, with between 74% and 100% of wt activity being observed.

As removal of 300 nt from the 5’ end of the RhPV UTR had little effect on IRES activity, further deletions were made of 350, 400, 425, 450, 475 and 500 nt (primer sequences are given in supplementary data). The fragments lacking up to 450 nt from the 5’ end still retained significant IRES activity in RRL (Figure 2A, right panel) but further deletion to nt 475 reduced activity to near background level.

We then explored the 3’ boundary of the IRES. Deletion of 50 nt from the 3’ end of the RhPV 5’ UTR (construct 3’Δ50) did not have any major affect on IRES activity but a further deletion of 50 nt (construct 3’ Δ100) reduced activity to about 35 % of wt. Deletion of 150 nt (construct 3’ Δ150) further reduced activity and depletion of the 3’ 200 nt resulted in a complete loss of activity (c.f. no IRES control). Most of these results are consistent with previous experiments (Terenin et al., 2005). However, in the previous study we observed that deletions from the 3’ end of the IRES only marginally affected IRES activity in RRL. There is a key difference between the two studies. The earlier experiments were performed using monocistronic transcripts with a hairpin located at the 5’ end rather than using dicistronic constructs (as used in the current study), the latter appear more stringent (see also Robertson et al, 1999).

In order to define the minimal 5’ UTR sequence required for IRES activity in plant systems, the plasmids were also assayed in the WGE TNT system (Promega). Reactions were analysed by SDS-PAGE/autoradiography and LUC assay as described above (Figure 2B). All the constructs expressed similar levels of CAT as expected.
As in the RRL system, the 5’ end deletions were tolerated much more than deletions from the 3’ end. However, in the WGE, five of the six 5’ end mutants directed LUC expression more efficiently than the full length 5’ UTR in the ∆1 construct. The 5’∆200 mutant displayed IRES activity of 150% compared to the full-length 5’ UTR. This suggests that the 5’ sequences may actually be inhibitory in WGE and their removal creates a more efficient IRES element. As seen in RRL, the 3’ deletions had a major effect on activity and removal of the 3’ 200 nt resulted in very little internal initiation.

The same reporter plasmids were also assayed in an insect cell system using a transient expression system in Sf21 cells as described in Royall et al., (2004). Cells (60 mm dishes) were infected with the recombinant baculovirus AcT7N that expresses T7 RNA polymerase (a gift from Monique van Oers, University of Wageningen; van Poelwijk et al., 1995) and subsequently transfected with plasmid DNA using lipofectin. Lysates (prepared 48 h later) were analysed for CAT and LUC expression using a CAT ELISA (Roche) and LUC expression as described above. LUC activities were normalised against CAT expression to take account of any differences in transfection efficiency between the plasmids (Figure 2C). Consistent with the observations made in RRL and WGE, the 5’ end deletion mutants displayed higher IRES activity than the 3’ end mutants. As observed in the WGE system, four of the six 5’ end mutants directed higher LUC expression than the full length 5’ UTR, the 5’∆200 construct exhibited about twice this activity. Interestingly, removal of the 3’ 50 nt resulted in a dramatic decrease in LUC expression (to about 10 % of the ∆1 value), which contrasts with the activity of this sequence in the RRL and WGE systems. However, a further deletion of 50 nt (construct 3’∆100) restored the activity. This suggests that the 3’ proximal structures may play a more critical role in
translation initiation in the insect cell system. It should be noted that in all 3 systems
the presence of the 5'UTR sequences in the antisense orientation resulted in no
significant IRES activity (Figure 2).

From these results it was apparent that deletion of up to 300 nt from the 5’ end
and 150 nt from the 3’ end of the RhPV 5’ UTR could be tolerated without
completely losing IRES activity in any of these three assay systems. We therefore
created three further mutants, with deletions at both ends, to define the minimal
sequence that can display IRES activity. Removal of the 5’ 150 and 3’ 150 nt resulted
in construct M150/429 and was obtained by PCR using the primers 5’∆150 F and
3’∆150 R. Deletion of the 5’ 250 and 3’ 150 nt created construct M250/429 and was
generated with primers 5’∆250 F and 3’∆150 R. Loss of the 5’ 300 and 3’ 150 nt
created construct M300/429 and this was generated using primers 5’∆300F and
3’∆150 R. All the plasmids were assayed in RRL, WGE and Sf21 cells. Remarkably,
the combined deletion mutants still retained at least 30-40% of the full length ∆1
sequence in all three systems tested, and up to about 80% of the wt IRES activity in
WGE (Figure 3). The smallest combined deletion mutant (M300/429) retained about
half the activity of ∆1, indicating the region from nt 300 to 429 can function as an
IRES. The same results were obtained in RRL programmed with in vitro-derived
RNA transcripts (data not shown). Since, as shown above, nt 426-579 in 5’∆425 also
functions, this data suggests that more than one region of the RhPV 5’ UTR can
function to direct internal initiation.

We have previously demonstrated that the RhPV 5’ IRES functions efficiently
in mammalian, plant and insect translation systems (Woolaway et al., 2001; Royall et
al., 2004). The unusual characteristics of this IRES element were studied further by
analysing the factors required for initiation complex formation and the secondary
structure of the IRES. We showed that large deletions could be made in the RhPV 5’ UTR sequence without losing IRES function using mammalian initiation factors and ribosomes (Terenin et al., 2005). In this study we now demonstrate that such large deletions (up to 425 nt from the 5’ end) can also be tolerated without significantly affecting activity in WGE and Sf21 cells. In fact, the 3’ proximal sequences functioned at least as well as the complete 5’ UTR in the WGE and Sf21 systems, sometimes more efficiently. Furthermore, we have shown that small internal fragments of the RhPV 5’ UTR (130 nt in the case of the smallest fragment, nt 300-429) are able to direct internal initiation in all three systems to about 40% of the efficiency of the wt ΔI sequence. From previous work (Terenin et al., 2005) we know that these regions are largely unstructured. Other viral IRES elements, for example those from the picornaviruses, cannot tolerate such large deletions. However, certain IRES elements from cellular mRNAs have this characteristic. For example, fragments of the immunoglobulin heavy-chain binding protein (BiP) and c-myc IRES elements retain function, however, they do not work in RRL (Yang & Sarnow, 1997; Stoneley et al., 2000).

We have previously proposed (Terenin et al., 2005) that the RhPV 5’ IRES probably consists of several regions that are able to bind ribosomes and factors, and the absence of structure facilitates this process. The studies presented here support this idea and show that minimal sequences of approximately 130 nt are important for IRES activity in mammalian, plant and insect origin. It is worth noting however, that this “cross-kingdom” activity has not been demonstrated for other 5’ IRES elements from the Dicistroviridae. For example, the CrPV 5’ UTR does not function in WGE (Wilson et al., 2000a). It may be that the ability of the RhPV 5’ UTR to function in
all three systems allows (or has historically allowed) the virus mRNA to be translated within cells of different origin, and has therefore aided survival and competition. Finally, we have previously reported the utility of the RhPV 5’ IRES element in the baculovirus expression system (Pijlman et al., 2006) and \textit{in vitro} protein expression systems (Royall et al., 2004). Use of smaller fragments of the RhPV IRES in such systems, rather than the full-length version, may be beneficial in the construction of such vectors.

\textbf{Acknowledgements}

This work was supported by the BBSRC. EG gratefully acknowledges receipt of a School of Biomedical and Molecular Sciences studentship award. We also thank Marion Chadd for excellent technical assistance and Elizabeth Royall for critical reading of the manuscript.
Figure legends

**Figure 1. Schematic representation of the truncated versions of the RhPV 5’ IRES.** The cDNA fragments were obtained by PCR and inserted in the intercistronic space of the dicistronic vector pGEM-CAT/LUC. The first cistron, CAT, encodes chloramphenicol acetyltransferase and is translated by a cap-dependent mechanism, while the second cistron encodes firefly luciferase (LUC) and is translated by an IRES-dependent mechanism.

**Figure 2. Delimitation of the 5’ and 3’ borders of the RhPV 5’ UTR sequence required for IRES activity in mammalian, plant and insect systems.** Plasmids which express mRNAs of the form CAT/IRES/LUC were analysed in RRL (A), WGE (B) and Sf21 cells (C) as described in the text. For RRL and WGE assays the samples were analysed by SDS-PAGE and autoradiography. LUC assays were also performed and values relative to the wt Δ1 construct are shown. LUC values were normalised against CAT expression. Results are representative of three separate experiments. For the Sf21 cells, CAT and LUC expression was measured as described in the text and LUC values normalised against CAT expression. Results are the means (±/− standard error) from three separate experiments.

**Figure 3. Internal fragments of the RhPV 5’ IRES display IRES activity.** Indicated plasmids of the form CAT/IRES/LUC were analysed in RRL (A), WGE (B) and Sf21 cells (C) as described in the text and legend to Figure 2. The names of the plasmids refer to the nts from the RhPV 5’ UTR contained within them.
References


van der Velden, A., Kaminski, A., Jackson, R. J. and Belsham, G. J. (1995). Defective point mutants of the encephalomyocarditis virus internal ribosome entry site can be complemented in trans. Virology 214, 82-90.


Figure 1

Δ1  1 ———————————— 579
Anti Δ1  579 ———————————— 1
5' Δ50  51 ———————————— 579
5' Δ100  101 ———————————— 579
5' Δ150  151 ———————————— 579
5' Δ200  201 ———————————— 579
5' Δ250  251 ———————————— 579
5' Δ300  301 ———————————— 579
3' Δ50  1 ———————————— 529
3' Δ100  1 ———————————— 479
3' Δ150  1 ———————————— 429
3' Δ200  1 ———————————— 379
M300/429  301 ——— 429
M250/429  251 ——— 429
M150/429  151 ——— 429

T7  CAT  LUC
Figure 2
Figure 3