IRES-mediated ribosome repositioning directs translation of a +1 overlapping ORF that enhances viral pathogenicity

[†]Craig H Kerr[1,2], [†]Qing S Wang[1], Kyung-Mee Moon[1,2], Kathleen Keatings[3], Douglas W Allan[3], Leonard J Foster[1,2], Eric Jan*[1]

[1]Department of Biochemistry and Molecular Biology, [2]Centre for High-Throughput Biology, [3]Department of Cellular and Physiological Sciences, University of British Columbia, Vancouver BC, V6T 1Z3 Canada

[†]contributed equally

*corresponding author: ej@mail.ubc.ca

keywords: ribosome, frameshift, dicistrovirus, drosophila, virus

ABSTRACT

RNA structures can interact with the ribosome to alter translational reading frame maintenance and promote recoding that result in alternative protein products. Here, we show that the internal ribosome entry site (IRES) from the dicistrovirus Cricket paralysis virus drives translation of the 0-frame viral polyprotein and an overlapping +1 open reading frame, called ORFx, via a novel mechanism whereby a subset of ribosomes recruited to the IRES bypasses downstream to resume translation at the +1-frame 13th non-AUG codon. A mutant of CrPV containing a stop codon in the +1 frame ORFx sequence, yet synonymous in the 0-frame, is attenuated compared to wild-type virus in a *Drosophila* infection model, indicating the importance of +1 ORFx expression in promoting viral pathogenesis. This work demonstrates a novel programmed IRES-mediated recoding strategy to increase viral coding capacity and impact virus infection, highlighting the diversity of RNA-driven translation initiation mechanisms in eukaryotes.

INTRODUCTION

The ribosome mediates translation involving decoding the open reading frame codon by codon through delivery of the correct aminoacyl-tRNAs to the ribosomal A site. This fundamental process occurs with high fidelity for proper gene expression in all species. However, mechanisms exist that can alter the translational reading frame, thus producing alternative protein products from a single RNA (Baranov, Atkins, & Yordanova, 2015). In general, these mechanisms termed recoding, involve a specific RNA structure or element that interacts with the ribosome to cause the translating ribosome to shift reading frame by -1/+1/+2 allowing it to read through stop codons or bypass sequences and restart translation downstream (Dinman, 2012; Firth & Brierley, 2012; Jan, Mohr, & Walsh, 2016). Study of these mechanisms has been enlightening; revealing key ribosome:RNA interactions that alter fundamental processes in the mechanics of ribosome decoding and reading frame maintenance. Importantly, recoding mechanisms are now appreciated as important regulatory processes that can impact the fate of protein expression in cells and virus infections (Au & Jan, 2014; Baranov et al., 2015). Unlike these recoding mechanisms that involve an actively translating ribosome, the intergenic internal ribosome entry site (IRES) within a subset of dicistroviruses has the unusual property to directly recruit the ribosome and initiate translation from overlapping 0 and +1-frame codons to produce two distinct proteins (Ren et al., 2012). Here, we report a novel recoding mechanism and translational initiation pathway whereby a related dicistrovirus IRES directs the ribosome to initiate translation downstream.

Most eukaryotic mRNAs utilize a cap-dependent scanning mechanism involving >12 translation initiation factors to recruit the ribosome and initiate translation from an AUG start codon (Jackson, Hellen, & Pestova, 2010). Alternatively, an IRES, in general, is a structured RNA element that facilitates 5' end-independent translation using subset of translation initiation factors thus providing an advantage during viral infection or under times of stress when cap-dependent translation is compromised (Au & Jan, 2014; Firth & Brierley, 2012; Jan et al., 2016).

Of the different classes of viral IRESs based on factor requirement and mechanism, the intergenic IRES of the *dicistroviridae* family stands out as the most streamlined using a unique mechanism where it directly binds 40S and 80S ribosomes without the need for canonical initiation factors or initiator Met-tRNA$_i$ and initiates translation from a non-AUG codon (Jan, Kinzy, & Sarnow, 2003; Jan & Sarnow, 2002; Pestova & Hellen, 2003; Petrov, Grosely, Chen, O'Leary, & Puglisi, 2016; Wilson, Pestova, Hellen, & Sarnow, 2000). The dicistrovirus IRES is composed of three pseudoknots (PKI, II, and III) that separate into distinct domains; PKII and PKIII fold independently to create the ribosome-binding domain while PKI mediates positioning of the ribosome and establishes the translational reading frame (Jan & Sarnow, 2002; Nishiyama et al., 2003; Wilson et al., 2000). Structural studies have indicated that PKII and PKIII form a compact core structure, and the PKI region adopts a conformation that mimics an anti-codon:codon interaction that binds that conserved core of the ribosome in the A site (Costantino, Pfingsten, Rambo, & Kieft, 2008; Fernandez, Bai, Murshudov, Scheres, & Ramakrishnan, 2014; Pfingsten, Costantino, & Kieft, 2006; Schuler et al., 2006). From here, in an elongation factor 2 dependent manner, the IGR IRES pseudo-

translocates to the ribosomal P site, followed by aminoacyl-tRNA delivery to the A site and a second round of eEF2-dependent pseudo-translocation of the IGR IRES to the E site of the ribosome (Au et al., 2015; Fernandez et al., 2014; Muhs et al., 2015). Altogether, the IGR IRES acts as a complete RNA machine that supersedes initiation factors and commandeers the ribosome, a strategy that is essential for viral protein synthesis in dicistrovirus-infected cells (Kerr, Ma, Jang, Thompson, & Jan, 2016).

In general, the dicistrovirus IGR IRESs are conserved at the structural, but not sequence level and are classified into two sub-groups (termed Type I and II) based on the presence of distinct structural elements; the main distinction comes from a larger L1.1 loop and an additional stem-loop (SLIII) in Type II IRESs (Jan, 2006; Nakashima & Uchiumi, 2009). SLIII allows for the PKI domain of Type II IRESs to mimic the global shape of a tRNA in addition to assisting in reading frame selection and the larger L1.1 region functions to mediate 60S recruitment (Au et al., 2015; Jang, Lo, & Jan, 2009; Pfingsten et al., 2006). The domains of Type I and II IGR IRESs function similarly to directly recruit 80S ribosomes and initiate translation (Cevallos & Sarnow, 2005; Hatakeyama, Shibuya, Nishiyama, & Nakashima, 2004; Hertz & Thompson, 2011; Jang & Jan, 2010). Recent high resolution cryo-EM structures of the IGR IRES bound to the 80S ribosome have demonstrated that the IRES initially binds in the A site (Abeyrathne, Koh, Grant, Grigorieff, & Korostelev, 2016; Fernandez et al., 2014; Muhs et al., 2015). Movement of the IRES involves an eEF2-dependent pseudo-translocation event where the ribosome rotates up to 10° allowing PKI to move into the P site in an inchworm-like manner (Abeyrathne et al., 2016). This allows for the non-AUG initiation codon of the IRES to be presented in the A site for the incoming amino-acyl tRNA. The first pseudo-

translocation event and delivery of the first amino-acyl tRNA are the rate-limiting steps of initiation on the IGR IRES (Zhang, Ng, Chen, & Cooperman, 2016).

Biochemical, phylogenetic, and bioinformatics analyses have demonstrated that a subset of Type II IGR IRESs are able to direct synthesis of a hidden +1 open reading frame (ORF), termed ORFx, within ORF2 of the viral genome (Firth, Wang, Jan, & Atkins, 2009; Ren et al., 2012). The functional role of ORFx during viral infection remains elusive. Extensive mutagenesis of the PKI domain of the *Israeli acute paralysis virus* (IAPV) IGR IRES has revealed that 0 and +1 frame translation can be uncoupled, suggesting that the IGR IRES may adopt specific conformations that govern the translational reading frame (Au et al., 2015; Ren, Au, Wang, Lee, & Jan, 2014). Generally, Type I and II IGR IRESs are thought to operate similarly in mechanism. Specific domains between the two types are functionally interchangeable (Jang & Jan, 2010). In the present study, we investigate the capacity of other IGR IRESs from Dicistroviruses to facilitate +1-frame translation. We show that the IGR IRES from Cricket paralysis virus (CrPV) can synthesize a 43 amino acid long ORFx protein using an unexpected mechanism that involves IRES-mediated ribosome bypassing. Furthermore, we provide insight into the role of ORFx during CrPV infection and show that mutants deficient in ORFx have impaired virulence in adult flies, thus uncovering a novel viral recoding strategy that is essential for viral pathogenesis.

RESULTS

IGR IRES-dependent +1-frame translation is not conserved throughout *Dicistroviridae*

We previously showed that a subset of dicistrovirus IGR IRESs can direct translation in the 0- and +1-frames and that a base pair adjacent to the PKI domain is important for initiation in the +1-frame (Figure 1) (Ren et al., 2014; Ren et al., 2012). The IGR IRESs are classified into two types: Type I, and II, with the main difference being an extra SLIII within the PKI domain of Type II IRES. Since the honeybee and fire ant viruses harbour Type II IGR IRESs that can support +1 frame ORFx translation, we investigated whether Type I IRESs also had the capacity for +1-frame translation. We first surveyed the other dicistrovirus IGR IRESs and their downstream sequences for a potential base pair adjacent to the PKI domain and an overlapping +1 open reading frame both *in silico* (Figure 1A, B). The IAPV and ABPV IGR IRESs contain an adjacent U-G base pair whereas the SINV-1 and KBV IRESs have a C-G base pair (Figure 1A). All but one of the IGR IRESs contains a potential base pair adjacent to the PKI domain, with the majority of them possessing a potential U-G base pair (Figure 1A). As shown previously, the honeybee and fire ant dicistrovirus ORFx proteins are approximately 94-125 amino acids in length (Firth et al., 2009). The predicted +1 ORFx lengths of other dicistroviruses range from 1-53 amino acids in length (Figure 1A). Apart from the honeybee and fire ant dicistrovirus ORFx, the longest putative ORFx sequences are found within the genomes of the CrPV and DCV at 53 amino acids in length.

To determine whether the other IGR IRESs can direct +1-frame translation, we cloned each IGR IRES within the intergenic region of a dual luciferase bicistronic

construct (Figure 1B). The Renilla luciferase (36 kDa; RLuc) monitors scanning-dependent translation whereas firefly luciferase (FLuc) is translated by the IRES. To measure reading frame selection, the firefly luciferase gene is fused in-frame either in the 0 or +1-frame. Translation was measured by incubating the bicistronic constructs in Sf-21 extracts in the presence of [^{35}S]-methionine/cysteine and the resulting proteins were detected by SDS-PAGE and quantified via phosophoimager analysis. Additionally, we generated bicistronic reporter constructs that contain a T2A 'stop-go' peptide, which allows sensitive and accurate measurement of luciferase activity (Wang et al 2013 Methods) (Figure 1- figure supplement 1). Since we previously showed robust IAPV IRES 0 and +1-frame translation (Ren et al., 2012), we chose it as a benchmark to compare the +1-frame activity of the other IGR IRESs. In general, all IGR IRESs can direct 0-frame translation to varying extents in this *in vitro* system. Normalizing to the IAPV 0-frame translational activity, the CrPV IGR IRES showed the highest 0-frame translational activity (~170%) whereas the MCDV IGR IRES had the lowest activity (~4%; Figure 1B). By contrast, only a few IGR IRESs can support +1-frame translation. The Type II IGR IRESs from Taura Syndrome Virus (TSV) and Mud Crab Dicistrovirus (MCDV) cannot support +1 frame translation indicating that only a subset of Type II IGR IRESs can direct 0 and +1 frame translation. Interestingly, besides the IAPV IGR IRES, the BQCV and CrPV IRESs mediated +1-frame translation above background levels, 80% and 5% of 0-frame translation, respectively. In summary, only a subset of IGR IRESs can facilitate both 0 and +1-frame translation.

CrPV +1-frame translation is IGR IRES-dependent and initiates downstream

To explore +1 frame translation mechanisms further, we focused on the CrPV IRES (Figure 1C). We first determined whether the structural integrity of the CrPV PKI domain is important for +1-frame translation. For these assays, we used a bicistronic reporter construct that contains the CrPV IRES where the firefly luciferase gene was subcloned into the +1-frame downstream of the IRES with CrPV nucleotides 6217-6387 (Accession KP974707.1), which includes the predicted +1-frame CrPV ORFx (65 kDa; ORFx-FLuc). A shortened IRES-mediated 0-frame protein (~11 kDa; sORF2) is also translated, thus scanning-dependent and CrPV IRES-mediated 0/+1-frame translation can be simultaneously monitored from this dual luciferase reporter construct (Figure 2A). Synthesis of all three proteins is detected by incubating the bicistronic construct in a Sf21 translation extract in the presence of [^{35}S]-methionine/cysteine (Figure 2B, lane 1) (Ren et al., 2012). As shown in Figure 1A, CrPV IRES +1-frame ORFx translation is approximately 5% of 0-frame translation. Mutating CC_{6214-5} to GG, which disrupts PKI base pairing and abolishes CrPV IRES activity, resulted in negligible or diminished 0- and +1 frame translation whereas a compensatory mutation that restores PKI base pairing rescued translation (Jan & Sarnow, 2002) (Figures 2B, lanes 2 and 3, Figure 1-figure supplement 1), indicating that the integrity of the IRES and the PKI domain is required for CrPV IRES +1-frame translation.

The adjacent U-G base pairing of the IAPV IRES is important for +1-frame translation (Ren et al., 2012). CrPV also has the capacity to form a wobble base pair adjacent to the PKI domain through nucleotides U_{6186} and G_{6217}, potentially directing translation from the first +1-frame CUA leucine codon (Figure 1A). To determine if this base pair is necessary to drive CrPV IRES +1-frame translation, we mutated U_{6186} and

G_{6217} to other bases. Mutating U_{6186} to C or G led to an approximate 18%-23% reduction in +1 frame activity and mutating G_{6217} to C, A, or U resulted in roughly a 39%, 24%, and 36% reduction in +1 frame activity, respectively. Although each mutation reduced +1 frame translational activity to some degree, none of the mutations abolished it (Figures 2B, lanes 4-9). These results suggest that unlike with IAPV, base pairing between nucleotides 6186 and 6217 is not absolutely required for CrPV IRES +1-frame translation.

To determine the potential initiation site of CrPV ORFx, we systematically replaced codons downstream of the IRES with a stop codon and monitored 0 and +1-frame translation *in vitro* using the bicistronic reporter construct (Figure 2B, C). Overall, stop codons placed in the +1-frame did not significantly affect 0-frame translation, indicating that IRES activity was not compromised (Figure 2B, C). Replacing individual codons between the 1st and the 12th +1-frame codon with a stop codon inhibited to varying extents (between 36%-71% reduction compared to wild-type) but did not completely abolish +1 frame translation (Figure 2B, lanes 11-16; 2C). Conversely, +1-frame translation was completely inhibited when the 13th +1-frame codon and codons thereafter was replaced with a stop codon (Figure 2B, lanes 15-16; 2C). Replacing both the 3rd and 5th +1-frame codons with stop codons reduced +1-frame translation by 33% but did not eliminate it, ruling out ribosome read-through. To address the possibility that IRES translation initiates in the 0-frame and then a fraction of translating ribosomes shift into the +1-frame, we inserted a stop codon in the 0-frame downstream of the IRES. As expected, a stop codon in the 0-frame 1st to 4th codons downstream of the IRES abolished 0-frame translation (Figure 2B, lanes 18-21). However, the 0-frame stop

codon insertions reduced by 52-77% but did not eliminate +1 frame translation indicating that a subset of ribosomes do not shift from the 0 to +1 reading frame. Further supporting this, introducing stop codons in both the 0 and +1 frames did not abolish +1 frame translation, though inhibited +1 frame translation by ~80% (Fig. 2C, 0 S4/+1 S5). We noted that the adjacent 14th codon is an AUG methionine. Mutating the AUG to a CGG (M14R) decreased by 63% but did not abolish +1-frame translation (Fig. 2C). In summary, the mutational analysis indicated that CrPV ORFx translation requires an intact IRES and the +1-frame AAA lysine 13th codon. Mutating the 13th AAA codon to UUU slightly reduced but did not abolish +1-frame translation (Fig. 2D), suggesting that there may be flexibility in the codon identity for CrPV ORFx +1-frame translation.

CrPV +1-frame translation occurs independently of 40S scanning

The current data led us to generate two hypotheses: i) a subset of 40S subunits recruited to the CrPV IRES scan downstream to start translation at the 13th codon (scanning hypothesis) or ii) 40S or 80S ribosomes recruited to the IGR IRES bypass the spacer region to the 13th codon (bypass hypothesis). To address the scanning hypothesis, we took a three-pronged approach. First, we utilized mutants of the CrPV IRES in the L1.1 loop that is known to be deficient in recruitment of the 60S subunit (Jang et al., 2009; Pfingsten et al., 2006). If scanning is occurring, 40S subunits recruited to the L1.1 IRES may scan downstream to the downstream +1-frame initiation codon. Reporter constructs harbouring mutations in the L1.1 loop were deficient in 0 and +1-frame translation (Figure 3A), suggesting that 60S recruitment by the IGR IRES is necessary for translation in the +1-frame.

Secondly, we utilized the translational inhibitor edeine to assess if scanning was occurring (Figure 3B). Edeine prevents the 40S ribosomal subunit from recognizing an AUG start codon (Dinos et al., 2004; Kozak & Shatkin, 1978). Both 0- and +1-frame translation were resistant to edeine relative to that of scanning-dependent translation (Figure 3B). This result further supports that the +1-frame 14th AUG codon is not utilized for +1-frame translation.

Finally, we introduced a stable hairpin (HP1 ΔG = -17.0 or HP2 ΔG = -37.0 kcal/mole) into the spacer region between the CrPV IRES and the downstream 13th codon to prevent 40S scanning (Figure 3C). Introducing HP1 or HP2 into the spacer region did not affect 0-frame translation showing that these mutations did not affect IRES activity (Figure 3C). In contrast, +1-frame translation was decreased by ~50%. However, in these constructs, the hairpins are likely unwound due to 80S ribosomes translating in the 0-frame, thus confounding the interpretation of the +1-frame translation result (Figure 3C). To uncouple these effects, we mutated the 0-frame 3rd codon to a stop codon upstream of the hairpin (0S3), thus blocking translating ribosomes from reaching the hairpin and maintaining hairpin integrity. Surprisingly, placing a stop codon upstream of the hairpin abolished +1-frame translation (Figure 3C), suggesting that 80S ribosomes may scan downstream to the 13th codon. Given the reduction in +1-frame translation observed when stop codons are placed adjacent to PKI (Figure 2A), it is possible that these hairpin insertions, in addition to the nucleotide identity, are disrupting a functional RNA element within the spacer region that mediates translation at the downstream +1 frame codon.

+2 frame translation mediated by the CrPV IRES

Our data thus far suggests that scanning of 40S ribosomes to the 13th position is likely not occurring. Therefore, we next asked whether ribosomes bound to the IGR IRES are repositioned to the downstream codon akin to translational bypass (Huang et al., 1988; Samatova, Konevega, Wills, Atkins, & Rodnina, 2014). We hypothesized that if the ribosome is indeed bypassing to a specific 'landing site' then it should be independent of frame, whereas if the ribosome merely begins translation in the 0-frame before slipping into the +1-frame, then we expect to see no ORFx translation. To address this, we inserted a series of nucleotides into our bicistronic construct that shifts only the ORFx-Fluc into the +2 frame (Figure 4). Specifically, we inserted either a single nucleotide or up to 7 proceeding the 13th AAA codon. As a control, we inserted either 6 or 9 nucleotides in the same position that does not introduce an additional frameshift. Insertion of a U creates a stop codon in the +1-frame while an inserted C does not. To our surprise, we observed ORFx expression with all insertions (Figure 4). These results suggest that repositioning of the ribosome from the IRES to a downstream 13th +1 or +2 frame codon is reading frame-independent.

The integrity of the VLR and pseudo-translocation of the IGR IRES through the ribosome is required for ORFx expression

How do 80S ribosomes recruited to the IRES mediate bypass at the downstream 13th +1-frame codon? Since the IGR IRES, which occupies the ribosomal A site upon ribosome binding, undergoes pseudotranslocation to the ribosomal P site to vacate the A site for delivery of the first aminoacyl-tRNA (Fernandez et al., 2014), we reasoned that the ribosome bound to the CrPV IGR IRES must have an empty ribosomal P or A

site in order to reposition downstream and accommodate the +1 frame start codon. To address this, we introduced mutations in the VLR, which has been shown to disrupt the IGR IRES-mediated pseudotranslocation event (Figure 1C) (Ruehle et al., 2015). Specifically, shortening the length of the VLR by 2 or 3 nucleotides (Δ2 and Δ3, respectively) inhibits the first pseudo-translocation event from the A site to the P site whereas altering the identity of nucleotides A_{6204}, and $AA_{6208\text{-}9}$ to guanosines (G-rich) inhibited IRES translocation from the P site to the E site (Ruehle et al., 2015). Interestingly, all three VLR mutants decreased +1-frame activity (Figure 5); both the G-rich and Δ3 mutants demonstrated little to no activity, while the Δ2 mutant still exhibited roughly 50% activity to that of WT. This result is consistent with previous data that the Δ2 mutant IRESs are still able to accommodate a fraction of aminoacyl-tRNA in the A site (~25%), allowing translation to occur (Ruehle et al., 2015). Altogether, these results indicate that the pseudotranslocation event of the IGR IRES through the ribosome contributes to +1-frame translation downstream.

Given that pseudotranslocation is required for +1 frame translation, we propose two possible models: After ribosomes that are recruited to the CrPV IRES-mediated pseudotranslation, i) a fraction of ribosomes start translation downstream at the 13th +1 frame codon or ii) a fraction of ribosomes initiates translation in the 0- or +1-frame (Ala or Leu, respectively) of CrPV ORF2 before ribosome bypass and resuming translation downstream at the 13th +1 frame codon. To distinguish these possibilities, we attempted to identify the N-terminal amino acids of ORFx. To this end, we fused a 6X His-tag on the C-terminus of ORFx-FLuc in the bicistronic construct lacking the T2A sequence (Figure 1 – figure supplement 1). Using this construct, we used a large scale

(100X) *in vitro* Sf-21 translation reaction in an effort to isolate the ORFx-FLuc-6XHis to appreciable levels (Figure 5 – figure supplement 1A). In a parallel reaction, we incubated radioactive [^{35}S] Met/Cys to reactions in order to follow purification of ORFx. A clear band corresponding to radioactive ORFx-Fluc-6XHis was observed via autoradiography and after purification with Ni^{2+}-NTA beads (Figure 5 – figure supplement 1A, 1B). The purified ORFx-Fluc-6XHis was then subjected to Edman degradation analysis (Figure 5 – figure supplement 1C, 1D). Subsequently, a likelihood score was calculated for each residue within a cycle to determine the appropriate amino acid assignment (see Materials and Methods). From this, we reproducibly observed Lys and Met residues in the 3rd and 4th positions, respectively (Figure 5 – figure supplement 1E). Unfortunately, the amino acids in positions 1 and 2 could not be unambiguously assigned by Edman degradation due to the low yield of +1 frame ORFx translation. Indeed, subjecting low picomole amounts of protein to N-terminal sequencing can result in convoluted traces in early cycles often due to contaminants from sample loading or the sample itself (Speicher, Gorman, & Speicher, 2009). Despite this, the amino acids with the best likelihood score for position 1 are Ala and Thr (note that Ser is a common contaminant and Trp is near background levels) while position 2 is observed as a Leu or Ile residue. Taken together with our biochemical analysis, this data suggests that a fraction of ribosomes recruited to the CrPV IRES undergoes pseudotranslocation in either the 0 or +1 frame and then bypasses downstream to resume translation at the 13th Lys codon.

The spacer region downstream of the IGR IRES is necessary for +1 frame

translation

We next investigated whether the spacer region located between PKI of the IGR IRES and the downstream AAA codon contributes to IRES-mediated ribosome bypass. Our results showed that inserting stop codons in the 'spacer region' (Figure 2A) between the IRES PKI domain and the +1-frame 13th codon of ORFx inhibited but did not completely abolish +1-frame translation, suggesting that an element within this spacer region may promote +1-frame translation. We first addressed whether the CrPV spacer region is sufficient to direct +1-frame translation by generating a chimeric construct whereby the PSIV IRES is fused with the CrPV spacer region (Figure 6). The PSIV IRES directs strong 0-frame but no or relatively weak +1-frame translation (Figure 1). The PSIV-CrPV spacer chimeric reporter resulted in 0-frame translation (Figure 6A), indicating that the spacer region does not affect the activity of the PSIV IRES. In contrast to the PSIV IRES construct, the chimeric PSIV-CrPV reporter resulted in +1-frame translation, implying that the CrPV spacer region is sufficient to drive +1-frame translation in the presence of a functional IGR IRES (Figure 6A).

To examine more closely the molecular determinants of the CrPV spacer region that direct +1-frame translation, we deleted or inserted a randomized sequence within the spacer region to test whether the length was important (Figure 6 and Figure 6 - figure supplement 1A). Insertions within or deletion of the entire spacer region resulted in loss of +1-frame but not 0-frame translation, suggesting that sequence context is important for +1-frame translation (Figure 6A - figure supplement 1B). Additionally, we examined the possibility of the spacer region to form intramolecular interactions. *In silico* analysis using RNA folding algorithms did not reveal any obvious stable secondary

structures in the spacer region (data not shown), however we noticed that nucleotides $AUUU_{6222-25}$ and $AAAU_{6240-43}$ have the potential to interact. Mutating $UUU_{6223-25}$ to AAA (A/U) or $AAAA_{6240-43}$ to UUUU (U/A) drastically reduced +1-frame translation (Figure 6 - figure supplement 1C). However, compensatory mutations did not restore +1-frame translation. To delineate a functional element within the spacer region, we systematically tested deletions from the 3' end of the spacer region (Figure 6B). Interestingly, deleting 3 to 27 nucleotides did not affect +1 frame translation. However, deleting 30 nucleotides and leaving seven nucleotides adjacent to the IRES abolished +1 translation (Figure 6B). Interestingly, Δ3, Δ6 and Δ27 mutants appear to have much higher Fluc/Rluc activity when compared to WT, an observation that requires further examination. In summary, these results suggest that the specific sequences within the spacer region, in particular the sequences immediately downstream of the PKI domain, are important for mediating IRES-dependent +1 frame translation.

CrPV +1-frame ORFx is expressed yet not required for infection in *Drosophila* S2 cells

Our *in silico* and biochemical data indicate that ribosomes recruited to the CrPV IRES can bypass downstream to resume +1-frame translation from the 13th codon. We termed the predicted +1 frame open frame CrPV ORFx, which is predicted to be 43 amino acids in length (Figure 7). To determine whether CrPV ORFx is synthesized during infection, we used a proteomics approach to identify ORFx peptides in CrPV-infected S2 cells. Specifically, *Drosophila* S2 cells infected with CrPV (MOI 10) were harvested at 6 hours post infection and lysed. Proteins were subsequently digested with

trypsin and peptides were analyzed by LC-MS/MS. We identified two peptides that correspond to CrPV ORFx both of which were located downstream of the +1-frame 13th codon (Figure 7). Importantly, these peptides were not identified in mock-infected S2 cells. Thus, ORFx is expressed in CrPV-infected S2 cells.

We next sought to determine if CrPV ORFx expression influences the outcome of viral infection. Using a recently developed a CrPV infectious clone, termed CrPV-2, we introduced mutations that would abolish ORFx expression (Kerr et al., 2015). To this end, we created two separate mutant clones: the +1-frame 12th codon (UUG_{6251-3}) was altered to an amber stop codon (UAG; CrPV-S12) and the +1-frame 19th codon (UUA_{6272-4}) changed to an ochre stop codon (UAA; CrPV-S19) (Figure 8A). Both mutations are synonymous in the 0 frame. Based on our translation data (Figure 2), the +1-frame S19 would inhibit ORFx expression whereas the +1-frame S12 would inhibit but not abolish ORFx expression.

We first examined whether ORFx influences viral protein synthesis *in vitro*. Incubation of *in vitro* transcribed CrPV-2 RNA in Sf21 translation extracts led to synthesis and processing of all viral proteins as reported previously (Figure 8B) (Kerr et al., 2015). Introduction of stop codons in the 0 frame of ORF1 or ORF2 inhibited synthesis of either viral non-structural or structural proteins, respectively (Figure 8B) (Kerr et al., 2015). Both CrPV-S12 and CrPV-S19 RNAs resulted in viral protein synthesis that was indistinguishable compared to wild-type CrPV -2 *in vitro* (Figure 8B). Note that CrPV ORFx is predicted to be ~4.9 kDa and is undetectable in this analysis. These results demonstrate that CrPV ORFx is not necessary for viral protein synthesis *in vitro*.

We next assessed the viability of the CrPV-S12 and -S19 viruses in cell culture. To determine if ORFx plays a role in the first replication cycle of CrPV, we transfected S2 cells with *in vitro* transcribed RNA from wild-type or mutant S12 or S19 CrPV-2 to generate virus. After 48 hours, virus was harvested and used to infect naïve S2 cells at an MOI of 10 to follow the first round of infection. We monitored viral protein synthesis via metabolic [^{35}S]-Met/Cys labeling as well as viral protein and RNA accumulation by Western and Northern blotting, respectively. Infection with wild-type CrPV-2, CrPV-S12 or CrPV-S19 all resulted in accumulation of viral proteins and RNA and shutdown of host translation in a similar manner (Figure 8C). Furthermore, there was no visible difference in viral RNA and protein levels between wild-type and mutant virus infections. These results indicated that ORFx does not play a role in the first round of CrPV replication in S2 cells.

To determine if ORFx influences subsequent rounds of infection, we infected S2 cells at an MOI of 1 and monitored viral titres (Figure 8D). Neither infection at an MOI of 1 produced significantly different titres between wild type CrPV-2 and either stop codon mutant at any time point, apart from both CrPV-S12 and CrPV-S19 resulting in higher titres than CrPV-2 after 24 hours post infection at MOI 1. A similar result was observed in CrPV-infected S2 cells at an MOI 0.1 (data not shown). Taken together, ORFx has no observable effect on the life cycle of CrPV in S2 cells.

ORFx contributes to CrPV pathogenesis in adult flies and associates with membranes

We next addressed whether ORFx contributes to CrPV pathogenesis in adult fruit

flies. To test this, we injected adult flies intrathoracically with PBS, CrPV-2, CrPV-S12, or CrPV-S19 and monitored fly mortality daily. By day 5, flies injected with CrPV-2 or CrPV-S12 exhibited 50% mortality and were 100% dead by day 11 and 12, respectively (Figure 9A). By contrast, flies injected with CrPV-S19 did not reach 50% mortality until day 9 and reached 100% mortality at day 14 (Figure 9A). These results demonstrate that ORFx contributes to CrPV pathogenesis in adult flies.

To determine if the effect seen on CrPV pathogenesis is a result of defects in viral replication, we measured viral titres and assessed viral protein levels in adult flies. At 5 days post infection, flies were homogenized and lysates were analyzed via Western blot for viral proteins and viral titres were determined. Interestingly, viral titres showed no significant differences between wild type CrPV-2, CrPV-S12, and CrPV-S19 (Figure 9B). As well, protein levels in infected flies were unaffected by the S19 mutation compared to wild-type or the S12 mutant (Figure 9C). Finally, to ensure that the S12 and S19 mutations in the mutant CrPV clones are stable, we performed RT-PCR analysis and sequenced this region. Our results showed that the S12 and S19 mutations are stable during virus propagation in S2 cells (data not shown). In summary, our results indicate that the defect in viral pathogenesis observed in the S19 mutant virus is not due to a defect in viral replication.

From our data, CrPV ORFx likely expresses a 43 amino acid protein whereby ribosomes bypass to resume translation from the 13th +1-frame AAA codon. Comparing the sequence of ORFx from CrPV and other species shows no appreciable homology to other proteins. Nevertheless, *in silico* topology predictions suggest that ORFx has the propensity to form an alpha helical transmembrane segment at its C-terminus (amino

acids 22-39; Figure 10 – figure supplement 1). To examine ORFx function, we generated Drosophila expression constructs containing either N- or C-terminal HA-tagged ORFx . We first assessed whether expression of HA-ORFx affected cell viability by a trypan blue exclusion assay. Transfection for 10 or 24 hours with the ORFx-HA and HA-ORFx construct did not result in a dramatic decrease in cell viability (Figure 10 – figure supplement 1). However, by 48 hours, transfection of ORFx-HA and HA-ORFx led to a statistically significant but relatively minor reduction (13% and 15% decrease, respectively) in cell viability, suggesting that ORFx expression has a slight toxic effect in S2 cells. Immunoblotting for HA showed that HA-ORFx is expressed in S2 cells (Figure 10 - figure supplement 2). To determine the subcellular localization in S2 cells, we transfected the HA-tagged ORFx constructs and monitored ORFx localization (pORFx-HA) by HA-antibody immunofluorescence staining in comparison with cytoplasmic, ER, Golgi, and nuclear marker protein antibodies. We also mutated two pairs of amino acids, LV and LI, to KK and KR respectively, which would disrupt the transmembrane domain (pORFx-TMmut-HA). Co-staining showed that the wild-type HA-tagged ORFx overlaps mainly with ER protein marker Calnexin, and partially overlaps with Golgi- and nuclear-associated proteins, Golgin84 and nuclear Lamin, respectively (Figure 10, pORFx-HA). The wild-type HA-ORFx displayed little to no overlap with α-Tubulin staining. By contrast, the mutant transmembrane HA-tagged ORFx staining showed no overlap with Calnexin, Golgin84, nor Lamin (Figure 10, pORFx-TmMut-HA). Collectively, these results suggest that ORFx associates with membranous organelles and potentially the ER specifically (Figure 10). To examine this further using another approach, we transfected S2 cells with the HA-tagged ORFx expression plasmid and used differential

centrifugation on cells to separate subcellular components. As controls, we monitored protein markers that should be enriched in different subcellular fractions. As expected, tubulin is found in the cytoplasmic fraction and the ER-associated KDEL protein and cytochrome C are enriched in nuclei and ER fractions (Figure 10- figure supplement 2). Both HA-tagged ORFx are detected within membranous fractions and not within the cytoplasmic fraction (Figure 10- figure supplement 2), further supporting that ORFx resides within the membrane of cells that may be important for its function.

DISCUSSION

Recoding mechanisms have illuminated diverse RNA structural elements that interact with the ribosome to affect reading frame maintenance (Dinman, 2012). In this study, we have demonstrated a novel translation recoding mechanism by which an IRES can promote ribosome repositioning at a downstream codon. In contrast to the IAPV IGR IRES, which directs ribosome reading frame selection (Ren et al., 2012), the CrPV IGR IRES can facilitate the expression of a downstream +1 overlapping frame, which we termed ORFx. We provide extensive analysis using mutagenesis that translation of CrPV ORFx occurs 37 nucleotides downstream at the 13th AAA (Lys) codon. Moreover, we show that ORFx is expressed in CrPV-infected cells by mass spectrometry and that ORFx is required for promoting CrPV pathogenesis in a *Drosophila* injection model. Our data suggest a model whereby after 40S and 60S ribosomal subunits binding to the CrPV IGR IRES, the majority of the ribosomes translate in the 0-frame by delivery of the incoming Ala-tRNA^{Ala} to the 0-frame GCU codon whereas a fraction of ribosomes undergoes pseudotranslocation in either the 0 or +1 frames and then bypasses or "slides" 37 nucleotides downstream to resume translation at the +1-frame AAA (Lys) codon (Figure 11). From our results, the following rules appear to apply to +1-frame translation in CrPV: i) the PKI must be intact, ii) both 40S and 60S ribosomal subunits are required to bind to the IRES, iii) pseudotranslocation of the IRES through the ribosome is necessary, iv) the spacer region between PKI and the 13th AAA codon is essential, and v) the nucleotide identity of the spacer region is crucial for efficient +1 frame translation.

A unique feature of this repositioning mechanism is that an intact IRES is essential for +1-frame translation (Figure 2). The current model is that the PKI domain

of the IRES occupies the ribosomal A site upon ribosome binding to the IRES, followed by a pseudotranslocation event in order to vacate the A site to allow delivery of the first aminoacyl-tRNA in the 0-frame (Fernandez et al., 2014). Given this model, it is difficult to envision how IRES can direct the ribosome to bypass 37 nucleotides to translate at a +1-frame AAA Lys codon. A potential clue comes from our mutagenesis analysis that suggests that translocation of the CrPV IRES through the ribosome is a prerequisite for downstream +1-frame translation (Figure 5). This is intuitive as the PKI domain must vacate the A site in order to allow translation in both the 0 and +1 frames. Building on this, N-terminal sequencing of the *in vitro* translated CrPV ORFx clearly showed that the 3rd and 4th N-terminal amino acids are Lys and Met, respectively, which are likely encoded by the 13th and 14th +1 frame codons from the CrPV PKI domain (Figure 5 – Figure supplement 1). Although the first two N-terminal residues of ORFx were not confidently assigned, it is clear that two translocation cycles occur prior to translation of the 13th and 14th codons. Thus, the ribosome bound to the IRES must undergo two pseudotranslocation events before ribosome bypass and 'landing' at the 13th +1 frame codon to resume translation (Figure 11). Stop codon insertions upstream of +1 frame 13th codon do not support +1 frame translation and insertion of a stop codon in the 12th codon of the CrPV infectious clone does not affect viral pathogenesis (Figures 2 and 9), thus ribosomes bound to the IRES must initiate translation then bypass to resume translation at the 13th codon. Given this, how can +1 frame translation still occur in the mutational analysis with 0- or +1-frame stop codons insertions directly downstream of the IRES (Figure 2)? A recent *in vitro* study using single molecule fluorescence spectroscopy by the Puglisi group demonstrated that the CrPV IGR IRES can facilitate

+1-frame translation approximately 5% of the time compared to 0-frame translation whereby reading frame selection is dictated by the kinetics of tRNA binding in the first 0- or +1-frame codon (Petrov et al., 2016). Thus, in the context of the stop codons in the 0 or +1 frames, a fraction of ribosomes recruited to the IRES can still initiate translation in either frame and then undergoes ribosome bypass downstream to the 13th +1 frame AAA codon (Figure 11). Determination of the aminoacyl-tRNA(s) that are delivered during the initial CrPV IGR IRES pseudotranslocation event should provide insights whether the ribosomal A site is vacant prior to ribosome sliding downstream to the 13th +1 frame codon.

Another key feature is that the downstream spacer region is necessary for efficient CrPV IRES-mediated +1-frame synthesis (Figures 3C, 6, and Figure 6 - figure supplement 1). Interestingly, the majority of mutations within the spacer region causes a reduction ranging from 10-80% in the amount of CrPV IRES-mediated +1-frame synthesis (Figure 2), suggesting that nucleotide or codon identity is crucial for triggering the bypass event. There are no obvious RNA secondary structures within the spacer region, however, we cannot rule out any long distance RNA:RNA interactions that may contribute to +1-frame translation. Furthermore, sequences immediately downstream of the IRES PKI domain are important for +1 frame translation (Figure 3B). Investigation into how the spacer sequence influences ribosomes bound to the IRES to start translation in the 0 frame or at the downstream +1 frame initiating codon should shed light onto this ayptical ribosome bypass mechanism.

Our results indicate that the 13th +1 frame AAA Lys codon is not a strict requirement as appreciable CrPV IRES +1-frame translation still occurs when the AAA

is altered to other codons (Figures 2C and 6B). Although there is an AUG codon at the 14th codon, altering this codon did not eliminate +1-frame translation. In addition, +1-frame translation is edeine-insensitive and requires an intact IRES PKI domain and 60S joining, thus ruling out a 40S scanning mechanism per se (Figures 2 and 3). Lastly, another interesting feature is that once ribosomes 'leave' the IRES, translation of the downstream codon is frame-independent, similar to that observed in the ribosome bypass event in *gene 60* of T4 bacteriophage (Huang et al., 1988; Samatova et al., 2014).

How does the ribosome reposition to the downstream +1-frame initiation codon after IRES binding? It is possible that the repositioning of the ribosomes occurs via mechanism similar to that observed with prokaryotic 70S 'scanning' (Yamamoto et al., 2016) or 70S 'sliding' that occurs in coupled translational reinitiation (Adhin & van Duin, 1990; Nomura, Gourse, & Baughman, 1984) (Figure 11). Indeed, a related dicistrovirus PSIV IRES can direct translation using a scanning-like mechanism in prokaryotes, which may suggest a similar property observed with our studies on the CrPV IRES (Colussi et al., 2015). Moreover, it is has been reported and proposed that energy-independent scanning or diffusion of the ribosome or ribosomal subunits (i.e. phaseless wandering) can occur to locate an AUG codon (Abaeva, Pestova, & Hellen, 2016; Adhin & van Duin, 1990; Sarabhai & Brenner, 1967; Skabkin, Skabkina, Hellen, & Pestova, 2013)(Figure 11). Nevertheless, our study shows that CrPV IRES +1-frame translation requires 80S ribosome binding to the IRES (Figure 3A) and is edeine-insensitive (Figure 3B), thus we favour a model that 80S ribosomes reposition to the 13th +1-frame codon. This warrants comparison is the translational bypassing observed in *gene 60* of T4

bacteriophage (Wills et al., 2008). In *gene 60*, translating ribosomes stall in a non-canonical rotated state at a 'take-off' Gly codon with a peptidyl-tRNA^{Gly}, which dissociates from the anticodon, and 'lands' at a matching Gly codon 50 nucleotides downstream and allowing translation to resume. This bypass mechanism is dependent on a post-translocation step requiring a downstream 5' ‘take-off’ stem-loop structure and a nascent translated peptide (Chen et al., 2015; Samatova et al., 2014). Although CrPV does not have an obvious RNA structure within the spacer region, it is possible that the highly complex structure of the IGR IRES itself may contribute to downstream +1 frame translation (Figure 11). Moreover, it is known that the ribosome bound to the IGR IRES is in a rotated state and that the first pseudotranslocation step is rate limiting, which may contribute to CrPV IRES reading frame selection (Fernandez et al., 2014; Zhang et al., 2016). Finally, it is possible is that CrPV IRES bypass could be occurring through an RNA looping event (Paek et al., 2015); the downstream RNA is brought into close proximity with the 80S ribosome allowing it to transition to 13th AAA codon potentially by an unknown protein factor or a long-range RNA:RNA interaction. Taken together, it is likely that it is a combination of tRNA kinetics and conformational changes of the IRES and the downstream spacer region that lead to bypassing, of which the contributions of each element require further investigation.

The biological relevance of CrPV +1-frame translation was initially evidenced by the detection of ORFx peptides in CrPV-infected S2 cells (Figure 7). To our surprise, disruption of ORFx synthesis by stop codon insertion in the +1-frame did not perturb viral infection in tissue culture cells but showed retarded mortality in adult flies even though viral load remained similar between wild type and mutant viruses (Figures 8 and

9). Naturally, CrPV infection is thought to cause death through paralysis, subsequently leading to dehydration or starvation of the host (Chtarbanova et al., 2014; Reinganum, 1975). CrPV can infect several tissues in the fly including the trachea, midgut, and central nervous system although the latter has not been demonstrated directly (Bonning & Miller, 2010; Chtarbanova et al., 2014; Paek et al., 2015). How ORFx contributes to CrPV pathogenesis is an outstanding question. Our results indicate that ORFx is membrane associated (Figure 10) and does not contribute directly to viral replication but rather to the pathogenesis of CrPV infection in fruit flies (Figure 9). Furthermore, expression of ORFx is slightly cytotoxic in Drosophila cells, a property that may also contribute to pathogenesis of CrPV infection. Future studies into the localization of ORFx, potential interacting partners, and tissue tropism of wild-type versus mutant virus infection in the fly should provide insights into the role of ORFx.

Viruses continue to surprise us with their ability to manipulate the ribosome in remarkable ways. Here, in addition to the previous findings on the honey bee dicistrovirus IRES, we have revealed another recoding mechanism utilizing an IRES, thus highlighting the strong selection to increase the coding capacity of the dicistroviridae genome. Furthermore, an IRES that can direct ribosome repositioning to facilitate the translation of a hidden +1 overlapping ORFx adds to the growing list of diverse pathways of ribosome translational recoding. The CrPV IGR IRES may serve as a model for other IRESs such as those from Hepatitis C virus and Theiler's murine encephalomyelitis virus that can direct translation of alternative reading frames (Vassilaki & Mavromara, 2009; Yamasaki, Weihl, & Roos, 1999). It will be of considerable interest to investigate whether other IRESs direct reading frame selection

by a similar CrPV IGR IRES-mediated ribosome repositioning mechanism. Ribosome repositioning or bypass is not specific to bacteria or mitochondria (Herr, Gesteland, & Atkins, 2000; Lang et al., 2014) but may be a more general phenomenon in eukaryotes that initially thought.

MATERIALS AND METHODS

un

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
strain, strain background ()	CrPV-2	PMID: 25810541	Accession: KP974706	Infectious clone of CrPV
strain, strain background ()	CrPV-ORF1-STOP	PMID: 25810541		
strain, strain background ()	CrPV-S12	This paper		
strain, strain background ()	CrPV-S19	This paper		
genetic reagent (D. melanogaster)	w[1118]	Bloomington Drosophila Stock Center	FLYB: FBst0003605	Maintained by the D. Allan Lab at UBC
cell line ()	Schneider 2 cells	Invitrogen	Catalog #: R690-07; FLYB:FBtc0000181; RRID:CVCL_Z992	Cell line maintained in E. Jan lab; FlyBase symbol: S2-DRSC.
antibody	CrPV anti-ORF2	PMID: 19889774	NA	Dilution: 1:1000
antibody	Golgin84	Developmental Studies Hybridoma Bank	Golgin84 12-1	Dilution: 1:1000
antibody	Calnexin99a	Developmental Studies Hybridoma Bank	Cnx99A 6-2-1	Dilution: 1:1000
antibody	Tubulin	Santa Cruz	sc-8035	Dilution: 1:1000
antibody	Lamin	Developmental Studies Hybridoma Bank	ADL84.12	Dilution: 1:1000
antibody	Cytochrome C	abcam	ab13575	Dilution: 1:1000
antibody	KDEL	Enzo Life Sciences	ADI-SPA-827	Dilution: 1:1000
other	Hoeschst Stain	Molecular Probes		Dilution: 1:10000
recombinant DNA reagent	pEJ551 (plasmid)	PMID: 23009811		Bicistronic construct lacking T2A sequence. Used in Figure 1 & 2.
recombinant DNA reagent	pEJ588 (plasmid)	this paper		Bicistronic construct containing T2A sequence allowing for luciferase activity monitoring used in Figure 2 and onwards.
recombinant DNA reagent	pORFx-HA	this paper		C-terminally HA-tagged CrPV ORFx sequence expressed from pAc5.1/V5-HisB (Thermo)
recombinant DNA reagent	pHA-ORFx	this paper		N-terminally HA-tagged CrPV ORFx sequence expressed from pAc5.1/V5-HisB (Thermo)
commercial assay or kit	Sf-21 TnT Extract	Promega	Catalog #: L1102	
chemical compound, drug	Edeine	NA	NA	Limited reagent held in the Jan lab

Cell culture and virus. *Drosophila* Schneider line 2 (S2; Invitrogen) cells were maintained and passaged in Shield's and Sang medium (Sigma) supplemented with 10% fetal bovine serum.

Propagation of CrPV in *Drosophila* S2 cells has been previously described (Garrey, Lee, Au, Bushell, & Jan, 2010). CrPV-2 and mutant viruses were generated from *Drosophila* S2 cells using an adapted protocol (Krishna, Marshall, & Schneemann, 2003). Briefly, 5.0 x 10^7 S2 cells were transfected with *in vitro* transcribed RNA derived from pCrPV-3 or mutant plasmids and incubated for 48h. Cells were dislodged into the media, treated with 0.5% Igepal CA-630 (Nonidet P-40) and 0.1% 2-mercaptoethanol, and incubated on ice for 10 min. Cell debris was cleared by centrifugation at 13,800 RCF for 15 min at 4°C. Viral particles were then concentrated by ultracentrifugation at 141,000 RCF for 2.5 h at 4°C. The pellet was resuspended in PBS and sterilized

through a 0.2 µM filter. Viral titres and yield were determined as previously described (Kerr et al., 2016). All viruses were sequence verified via RT-PCR with primers directed against the CrPV IGR IRES.

Plasmids and bicistronic reporter constructs. Each IGR IRES with flanking upstream and downstream sequences was cloned between the EcoRI and the NcoI sites within the intergenic region of pRΔDEF as described (Johannes, Carter, Eisen, Brown, & Sarnow, 1999). The Firefly luciferase (FLuc) open reading frame was fused in frame to either the 0 or +1 frame. The sequences that were cloned are as follows: nucleotides 6372–6908 of Israeli acute paralysis virus (IAPV; NC_009025.1); nucleotides 6741-6969 of Taura syndrome virus (TSV; NC_003005.1); nucleotides 6860-7243 of Mud crab dicistrovirus (MCDV; HM777507.1); nucleotides 5974–6372 of Cricket paralysis virus (CrPV; NC_003924.1); nucleotides 6060-6422 of Drosophila C Virus (DCV; NC_001834); nucleotides 5982-6324 of Plautia stali intestine virus (PSIV; NC_003779.1); nucleotides 151-771 of Big Sloux River virus (BSRV; JF423197.1); nucleotides 5626-5917 of Black queen cell virus (BQCV; NC_003784.1); nucleotides 3453-3712 of Aphid lethal paralysis virus (ALPV) strain brookings non-structural polyprotein and capsid protein precursor (HQ871932.1); nucleotides 6546-7159 of Rhopalosiphum padi virus (RhPV; NC_001874.1). The PSIV_CrPV chimera contains the nucleotides 5981-6192 of PSIV and 6217-6371 of Cricket paralysis virus isolate CrPV-2 (KP974706.1)(Kerr et al., 2015).

For CrPV IRES-containing bicistronic constructs, the AUG start codon of the FLuc gene in the +1 frame was removed by PCR-based site-directed mutagenesis

(Stratagene). For the T2A-containing constructs, the Thosea asigna virus (accession: AF062037) 2A sequence was inserted in frame preceding the FLuc gene into plasmid pEJ566 between EcoRI and NdeI restriction sites as described previously (Wang, Au, & Jan, 2013).

Hairpin structure sequences HP1 and HP2 were inserted between the 6th codon AAC and 7th codon AAG of CrPV (+1) frame. The HP1 sequence is 5'-CAGCGTTATCG GGAACGTCGAAGGGGTACGCCATTGTTCGACCACCGGCTCGAGGGGCCC-3' (ΔG = -17.0 kcal/mole). The HP2 sequence is 5'- ACTGCAGCCACAACCGGCCCCAAGCTT GGGCCGTGGTGGCTGCAG-3' (ΔG = -32.70 kcal/mole).

Other dicistrovirus sequences used for analysis in in Table S1 are: Solenopsis invicta virus-1 (SINV-1) NC_006559; Kashmir bee virus (KBV) NC_004807; Acute bee paralysis virus (ABPV) NC_002548; Homalodisca coagulata virus-1 (HoCV-1) NC_008029; Himetobi P virus (HiPV) NC_003782; Cripavirus NB-1/2011/HUN (Bat guano dicistrovirus) NC_025219.1; Aphid lethal paralysis virus (ALPV) NC_004365; Aphid lethal paralysis virus strain brookings (ALPV-brookings) HQ871932; and Triatoma virus (TrV) NC_003783.

The constructs encoding HA-tagged CrPV ORFx were generated as follows. For the C-terminally tagged ORFx-HA, a 3X HA tag was inserted between XbaI and ApaI sites in pAc5.1/V5-HisB (Thermo). CrPV ORFx (nucleotides 6254-6376) was subsequently cloned in-frame using KpnI and XbaI sites. For the N-terminally tagged HA-ORFx a synthesized fragment of DNA (gBlock; IDT) containing a 3X HA tag followed by the CrPV Orfx sequence was cloned into the pAc5.1 vector with KpnI and ApaI. The transmembrane mutant of ORFx was generated via site-directed

mutagenesis on the C-terminally HA-tagged construct. All constructs were verified via sequencing.

***In vitro* transcription and RNA transfection**. pCrPV-2 (Kerr et al., 2015) and derivative plasmids were linearized with Ecl136II. RNA was transcribed using a T7 RNA polymerase reaction and subsequently purified with a RNeasy kit (Qiagen). The integrity and purity of the RNA were confirmed on a 1.2% denaturing formaldehyde agarose gel.

Transfection of *in vitro* synthesized RNA into S2 cells was performed using Lipofectamine 2000 (Invitrogen) as per the manufacturer's instructions. 3 µg of RNA derived from either pCrPV-2 or its cognate mutants were used for transfection using 2.5 x 10^6 cells.

***In vitro* translation assays**. Plasmids were linearized with BamHI and then purified using QiaQuick Purification columns (Qiagen). The Sf-21 TnT Coupled Transcription/Translation System was utilized for all translation assays (Promega). In this system, linearized plasmid is added to the extracts and RNA is synthesized by T7 polymerase and subsequently translated. Total reaction volume was 10 µL containing: 6.7 µL of Sf-21 cell extract (Promega), 0.3 µL of L-[^{35}S]- methionine/cysteine (PerkinElmer, >1000 Ci/mmol) and 1 µg linearized plasmid DNA. Each reaction mixture was incubated at 30°C for 2 hours and then resolved on a 15% SDS-PAGE. Gels were dried and radioactive bands were monitored via phosphoimager analysis. For quantitation of bands, the number of methionines and cysteines were accounted for and

normalized for each protein. For T2A containing constructs (Wang et al., 2013), luciferase activity was monitored using a Dual-Luciferase reporter assay system (Promega) and an Infinite M1000 PRO microplate reader (Tecan).

Northern blot analysis. Total RNA was isolated from cells using TRIzol reagent. Equal amounts of RNA were loaded on a denaturing agarose gel and subsequently transferred to Zeta-probe blotting membrane (Bio-Rad). DNA probes were radiolabelled with $\alpha[P^{32}]$-ATP (MP Biomedicals) using a DecaLabel DNA labeling kit (Fermentas). Probes were incubated with membrane in hybridizatn buffer (7% SDS, 0.5 M Sodium phosphate, 1 mM EDTA) at 65°C overnight. Membranes were rinsed with H_2O and radioactive bands were detected via phosphoimager analysis (Storm; GE Healthcare). The DNA probe to detect CrPV RNA was generated by PCR amplification using primers 5'-TCCTCAAGCCATGTGTATAGGA-3' and 5'-GTGGCTGAAATACTATCTCTGG-3'.

Western blots. Equal amounts of S2 protein lysates were resolved on a 12% SDS-PAGE gel and then transferred to a polyvinylidene difluoride Immobilon-FL membrane (Millipore). Membranes were blocked for 30 min at room temperature with 5% skim milk in TBST (50 mM Tris, 150 mM NaCl, 1% Tween-20, pH 7.4). Blots were incubated for 1 h at room temperature with the following antibodies: CrPV ORF1 (raised against CrPV RdRp) rabbit polyclonal (1:10,000), CrPV ORF2 (raised against CrPV VP2) rabbit polyclonal (1:10,000) (Garrey et al., 2010), mouse anti-tubulin (1:1000; Santa Cruz), mouse anti-KDEL (1:1000;), or mouse anti-cytochrome C (1:1000; AbCam).

Membranes were washed 3 times with TBST and incubated with either goat anti-rabbit IgG-HRP (1:20,000; GE Healthcare), goat anti-mouse IgG-HRP (1:5000; Santa Cruz), IRDye 800CW goat anti-mouse (1:10,000; Li-Cor Biosciences), or IRDye 800CW goat anti-rabbit (1:20,000; Li-Cor Biosciences) for 1 h at room temperature. An Odyssey imager (Li-Cor Biosciences) or enhanced chemiluminescence (Thermo) were used for detection.

Protein isolation and N-terminal sequencing. An ORFx-6XHis-tag was introduced into the bicistronic construct lacking the T2A sequence. The construct was linearized and protein was synthesized in an *in vitro* translation extract as previously described (see above). Reactions were pooled and diluted with Buffer A (50 mM NH_4HCO_3, 500 mM NaCl, 15 mM Imidazole, pH 8.0) before incubating with Ni-NTA beads (QIAGEN) overnight at 4°C. After incubation, beads were then washed 3 times with Buffer A before eluting with Buffer B (50 mM NH_4HCO_3, 500 mM NaCl, 500 mM Imidazole, pH 8.0). Eluted protein was concentrated over a 30 kDa MWCO filter, separated via Tris-glycine SDS-PAGE, and transferred to a PVDF membrane. Membranes were stained with Coomassie brilliant blue dye and the corresponding band was analyzed by Edman sequencing (SickKids SPARC Biocentre). Likelihood scores were calculated as follows: area under the curve (AUC) for each amino acid in each cycle was determined using ImageJ software. Each AUC was then divided by the AUC from the previous cycle subtracted by one to obtain a difference ratio (i.e. (AUC X_n / AUC X_{n-1}) – 1; where X = a given amino acid and n = cycle number).

LC-MS/MS Analysis. Cell pellets harvested from CrPV-infected cells at 6 hpi were solubilized in 1% sodium deoxy cholate and 50 mM NH_4HCO_3. Protein concentrations were determined via BCA assay (Thermo). Proteins (100 µg) were reduced (2 µg DTT, 37°C, 30 min) and alkylated (5 µg iodoacetamide, RT, 20 min). Samples were digested with trypsin overnight at room temperature. Peptides were acidified with 1% TFA to pH <2.5 and the precipitated deoxycholate was remove via centrifugation. Peptides were desalted and concentrated on C18 STAGE-tips, eluted in 80% acetonitrile, 0.5% acetic acid, and dried in a vacuum concentrator (Eppendorf)(Rappsilber, Ishihama, & Mann, 2003). Samples were resuspended in 20% acetonitrile and 0.1% formic acid before loading on an Agilent 6550 mass spectrometer.

Data were searched using MaxQuant (v1.5.3.30)(Cox & Mann, 2008). Parameters included: carbamidomethylated cysteine (fixed), methionine oxidation (variable), glutamine and asparagine deamidation (variable), and protein N-terminal acetylation (variable); trypsin specific; maximum 2 missed cleavages; 10 ppm precursor mass tolerance; 0.05 Da fragment mass tolerance; 1% FDR; +1 to +7 charge states; and common contaminants were included. Both the *Drosophila* and CrPV protein databases used were the most recent annotations downloaded from UniProt (www.uniprot.org).

Subcellular fractionation by differential centrifugation. 3 X 10^7 S2 cells were transfected with either Drosophila expression vectors (pAc5.1) containing ORFx-HA or HA-ORFx (pORFx-HA or pHA-ORFx) for 48 h at 25°C (XtremeGene; Roche). Cells were then collected by centrifugation at 200 x g for 5 min and washed twice with ice

cold PBS. Cell pellets were then resuspended in SF Buffer (20 mM HEPES [pH 7.4], 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA) and incubated on ice for 10 min. Cells were subsequently disrupted by passaging through a 25 gauge needle 20 times. Whole cells were removed by centrifugation at 200 x g for 5 min and nuclei were then collected by centrifugation at 600 x g for 10 min. The mitochondria and endoplasmic reticulum (ER) were pelleted by centrifugation at 3000 x g for 15 min. Finally, remaining membrane was pelleted at 16,000 x g for 30 min. The remaining supernatant was collected as the cytosolic fraction.

Immunofluorescence. S2 cells (1.5 x 10^6) were transfected for 48 h with 2 µg of pAc5.1/V5-HisB plasmid constructs expressing N- or C-terminally HA-tagged ORFx. Cells were seeded onto coverslips and fixed with 3% paraformaldehyde for 15 minutes. Subsequently, cells were permeabilized with 0.2% Triton X-100 in 1X PBS for 1 h before blocking coverslips with 3% BSA in 1X PBS for 1 h. Coverslips were then incubated with rabbit anti-HA (1:1000; Cell Signalling) and either mouse anti-Lamin (1:1000; DSHB), mouse anti-Golgin84 (1:1000, DSHB), or mouse anti-Calnexin 99A (1:1000, DSHB) in blocking solution overnight at 4°C. Cells were then washed three times with 1X PBS. Finally, slides were analyzed using a Leica SP5 confocal microscope with a 63x oil objective lens and a 2x digital zoom. Z-stacks of 15 slices each were taken of for each condition.

Fly stocks and viral injections. Flies (*Isogenic* w^{1118}; Bloomington Drosophila Stock Center) were maintained on standard cornmeal food at 25°C and 70% humidity with a

12 h light-dark cycle. Freshly eclosed virgin males and females were separated and collected in groups of 10 each. Flies (10 males and 10 females) were injected with 200 nL of PBS, CrPV-2, CrPV-S12, or CrPV-S19 (5000 FFU) using a PV830 PicoPump (World Precision Instruments) and transferred to standard food. Mortality was monitored daily.

ACKNOWLEDGEMENTS

We thank John Atkins for insightful discussions. Kathy Pan and Seonghoon Lee contributed a subset of data in Figure 1B and 2B, respectively. This study was supported by a Canadian Institutes of Health Research (PJT-148761) to E.J. and to D.A. (MOP-130517), an NSERC Discovery Grant to L.F. and an NSERC CGS-D fellowship to C.K.

COMPETING INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Biochemical analysis of potential +1-frame translation by Dicistrovirus IGR IRESs. A) *In silico* analysis for potential +1-frame translation mediated by IGR IRESs in Dicistroviruses. List of type I and type II IGR IRESs based on their secondary structures. Potential +1-frame base pair immediately downstream of their IGR IRES PKI region as in IAPV IGR IRES, together with the predicted length of ORFxs and the first amino acids, are shown. **B) IGR IRES mediated +1-frame translation using an *in vitro* translation assay.** The translation of firefly luciferase (FLuc), which is fused in the 0-frame or +1-frame, is driven by the individual IGR IRES within the bicistronic reporter construct. Linearized reporter constructs are incubated in Sf21 *in vitro* translation insect cell extract at 30 °C for 2 hours in the presence of [^{35}S] methionine/cysteine. Translation of FLuc and RLuc was monitored by phosphorimager analysis after resolving on a 16% SDS PAGE. On the bottom panel, the ratio of FLuc/RLuc are quantified and normalized to the IAPV 0-frame translation. Shown are averages from at least three independent biological experiments (± SD). **C) The secondary structure of CrPV IGR IRES.** The ΔPKI, ΔPKII, ΔPKIII, L1.1A, and L1.1B mutants are shown. Comp denotes compensatory mutations combining ΔPKI CC/GG and ΔPKI GG/CC. DM - double mutation of ΔPKI and ΔPKII. TM - triple mutation of ΔPKI, ΔPKII, and ΔPKIII.

Figure 1 - figure supplement 1. **Construction of a T2A construct +1-frame reporter construct that recover FLuc enzyme activity. A)** The schemes of bicistronic constructs in which firefly luciferase gene (FLuc) is in frame with the CrPV ORFx are

shown. The T2A sequence (dark grey) is inserted between the CrPV ORFx and the FLuc gene. The arrow indicates the 'self-cleavage' or 'stop-go' site. A mutation within the T2A peptide (D to E) inactivates T2A 'self-cleavage' activity. T2A-minus (-T2A) and T2A-containing ((+)T2A) +1-frame bicistronic constructs were incubated in Sf21 extracts for 2 h in the presence of [^{35}S] methionine and cysteine. The left panel is a representative SDS-PAGE monitored by autoradiography, whereas the right panel is the quantitation of relative luciferase activity (RLA) measured by Dual-Luciferase reporter assay (Promega). RLA are normalized to that observed with the +1-frame T2A-containing reporter construct. Shown are averages from at least three independent biological experiments (± SD). **B)** The quantitation of luciferase activities of IGR IRES mutants compared to WT T2A containing +1-frame constructs are shown. Displayed are averages from at least three independent biological experiments (± SD). **C)** The comparison between +1-frame and 0-frame T2A containing constructs. The ratio of +1/0-frame translation *in vitro* is 7.1 ± 0.5 %, consistent with that monitored by autoradiography using T2A-minus constructs (Figure 1B). Note that the FLuc activity levels from both the 0-frame ΔPKI mutant and the +1-frame ΔPKI mutant indicates to some extend IGR IRES independent translation. Shown are averages from at least three independent biological experiments (± SD).

Figure 2. Translation of +1 frame CrPV ORFx is IGR IRES-dependent and initiates downstream of the PKI region. A) Schematic of mutations introduced downstream of the IGR IRES. The PKI region and downstream 'spacer' sequence of CrPV IGR IRES are shown. A series of mutations and stop codons were introduced either on the

0-frame or +1-frame. **B) Analysis of the IGR IRES translation *in vitro*.** Linearized reporter constructs are incubated in Sf21 translation extract at 30°C for 2 hours in the present of [^{35}S]-methionine/cysteine. Translation of FLuc and RLuc was monitored by autoradiography after resolving on a 16% SDS PAGE. A representative gel from at least three independent biological experiments is shown. * = p-value < 0.05. **C) Quantitation of IRES-mediated translation.** Translation of FLuc and RLuc monitoring FLuc/RLuc enzymatic activity were quantified and normalized to wild-type (WT) IAPV IRES. RLuc monitors scanning-dependent translation acting as an internal control. RLA- relative light units. Shown are averages from at least three independent biological experiments (± SD). * = p-value < 0.05.

Figure 3. Translation of CrPV ORFx is dependent on 80S binding and scanning-independent. A) Effect of L1.1 mutation on +1 frame translation. Comparison of luciferase activities between WT IGR IRES and L1.1 mutants (L1.1A and L1.1B) that do not bind the 60S subunit. The value of RLuc and FLuc are normalized to WT. **B) +1-frame translation is relatively insensitive to edeine.** CrPV IRES-mediated 0- and +1-frame translation is monitored in the absence or presence of 2 μM edeine. The ΔPKI mutants in both 0- and +1-frame are used as controls for IGR IRES dependent translation. The luciferase activities of RLuc and FLuc on the left panel are normalized to the WT of the 0-frame, whereas the luciferase activities of RLuc and FLuc on the right panel are normalized to the WT of the +1-frame. **C) Effect of RNA hairpin structures on +1 frame translation.** Translation of 0- and +1-frame translation of bicistronic reporter constructs that contain RNA hairpins structures (HP1 ΔG = -17.00 kcal/mole

and HP2 ΔG = -32.70 kcal/mole) between the nucleotides 6135 and 6136. To block the ribosome scanning the downstream sequence driven by the 0-frame initiation, we introduced a stop codon at the 3rd codon on the 0-frame on both HP1 and HP2 constructs (0S3). Shown are averages from at least three independent biological experiments (± SD).

Figure 4. CrPV IRES-mediated +2 frame translation. Mutations and insertions that shift the ORFx into the +2 frame are shown in the schematic above. Translation assays were monitored and normalized to that of the wild-type CrPV IRES. Shown are averages from at least three independent biological experiments (± SD).

Figure 5. Pseudotranslocation is necessary for +1-frame translation. Constructs containing mutations in the variable loop region (VLR) of the IGR IRES known to disrupt pseudotranslocation were monitored for translation activity. Luciferase activities were compared between WT IGR IRES and VLR mutants after *in vitro* translation assays. Shown are averages from at least three independent biological experiments (± SD).

Figure 5 – figure supplement 1. N-terminal sequencing of ORFx-Fluc-6XHis. (A) Schematic of the predicted ORFx sequence. 0- and +1-frame protein sequence of ORF2 (green) and ORFx (red) are shown. **(B) *In* vitro translation of ORFx-FLuc-6XHis.** Constructs were linearized with BamHI, purified, and incubated with Sf-21 extract including [S^{35}]-Met/Cys. Proteins were separated via SDS-PAGE and visualized by autoradiography. **(C) Pulldown of radiolabeled ORFx-FLuc-6XHis.** Sf-21

translation extracts were incubated with Ni-NTA beads (see Materials and Methods). Gels were exposed using autoradiography. Shown is an representative gel from two independent experiments. **(D) Large scale isolation of ORFx-Fluc-6XHis.** Proteins from elution fractions in (C) were pooled and concentrated before transferring to a PVDF membrane. Membranes were stained with Coomassie before band excision and N-terminal sequencing. The band corresponding to ORFx-FLuc-6XHis is marked with an asterisk. Shown is a representative pulldown result from two independent experiments. **(E) Elution profiles of N-terminal sequencing (Edman degradation) of ORFx-Fluc-6XHis.** The N-terminal elution profiles of amino acid residues (1-4) from Edman degradation sequencing. Amino acids are shown as individual letter codes. dpu = diphenolurea; dptu = diphenylthiourea. **(F) Calculated likelihood scores for amino acids at each position of ORFx-FLuc-6XHis.** Scores were calculated as in Materials and Methods. Amino acid assignments with the highest confidence are encompassed by a blue box. Arrows denote which amino acids likely reside in the first position.

Figure 6. The downstream sequence of CrPV IGR IRES promotes +1-frame translation. A) Chimeric IRES displays +1-frame translation activity. The PSIV IGR IRES, which shows no +1-frame translation activity (Figure 1 and Figure 3A), is combined with the downstream sequence of CrPV IGR IRES PKI region. The luciferase activities are normalized to the wild-type CrPV constructs in both +1- and 0-frame. The PSIV-CrPV chimera in the 0-frame is used as control to demonstrate that PSIV IGR IRES functions under this context. **B) Truncation of the spacer region impairs +1-frame translation from the CrPV IGR IRES.** Schematic displaying truncations within

the spacer region between the PKI domain and the +1 frame start site is shown above. *In vitro* translation assay of the CrPV IRES bicistronic construct with the respective mutations were monitored by luciferase activities. Both RLuc and FLuc activities are normalized to that of WT constructs. Results are shown as a normalized ratio of FLuc/RLuc activity (left) and as separated RLuc and Fluc activities (right) Shown are averages from at least three independent biological experiments (± SD). RLA = relative luciferase activity.

Figure 6 - figure supplement 1. Effects of deletion and insertion mutations in the CrPV spacer region on +1-frame translation. Schematic of the spacer region between the PKI domain and the +1 frame start site is shown above. *In vitro* translation assay of the CrPV IRES bicistronic construct with the indicated mutations or an insertion mutation (Ins) of 30 random nucleotides between G_{6217} and C_{6218} or between C_{6235} and A_{6236} are monitored by luciferase activities. Both 0- and +1-frame activities are normalized to that of WT constructs. Shown are averages from at least three independent biological experiments (± SD). RLA = relative luciferase activity.

Figure 7. ORFx is expressed in CrPV infected S2 cells. The predicted +1 frame amino acid sequence is shown above. Residues that are italicized represent the amino acid sequence if initiation occurred adjacent to the IGR IRES, whereas the bolded K lysine residue denotes the position of the 13th amino acid. *Drosophila* S2 cells were infected with CrPV (MOI 10). Proteins were extracted, digested with trypsin and subjected to LC-MS/MS analysis. Peptide fragment spectra were searched against a

Drosophila uniprot database plus CrPV proteins. Two peptides (highlighted) were detected from the trypsin digestion of S2 cell lysate at 6 hours post infection. Individual fragment ions are annotated in the spectra and in the sequence representation.

Figure 8. Role of CrPV ORFx in CrPV-infected S2 cells. A) Schematic of mutations in the CrPV infectious clone, CrPV-2. The downstream nucleotide sequence of CrPV-2 IGR IRES and its potential amino acid sequence of +1-frame ORFx are shown. The ORF2-Stop mutant depletes the synthesis of 0-frame viral structural proteins. Mutants S12 and S19 place a stop codon on the +1-frame with a synonymous mutation in the 0-frame. Note that CrPV-2 (Accession: KP974706) sequence has a $C_{6279}A$ mutation (denoted with a black arrowhead) compared to the original CrPV (Accession: AF218039) sequence. Residues that are italicized, underlined and in bold are the peptides identified by mass spectrometry in Figure 7. **B) CrPV ORFx has no effect on viral protein synthesis *in vitro*.** The *in vitro* transcribed RNA from CrPV-2 or the cognate mutants was incubated in Sf21 cell extract at 30°C for 2 hours in the present of [^{35}S]-methionine/cysteine. Translation of viral proteins were monitored by autoradiography using 12% SDS PAGE. A representative gel of at least three independent biological experiments is shown. **C) CrPV ORFx has no effect on viral protein synthesis and RNA replication in S2 cells.** The *in vitro* transcribed RNA of CrPV-2, S12, or S19 mutants was transfected into S2 cells. Virus was harvested and used to infect naïve S2 cells at MOI 10. Cells were metabolically labeled with [^{35}S]-methionine/cysteine for 30 min at 2, 4, 6 and 8 h.p.i. Lysates were collected and resolved on a 12% SDS PAGE. Western blot analysis against the CrPV RdRp and VP2

and Northern blot analysis of viral RNA are shown. Shown is a representative gel from three independent biological experiments. **D) CrPV ORFx has no effect on viral replication in S2 cells.** S2 cells were infected with CrPV-2, CrPV-S12, or CrPV-S19 viruses at MOI 1.0. Viral titres were measured at 6, 9, 12, 24, and 30 h.p.i for MOI 1. Shown are average from at least three independent biological experiments (±SD).

Figure 9. Pathogenesis of a CrPV mutant lacking ORFx is attenuated in adult *Drosophila melanogaster* flies. A) Adult flies (Iso w^{1118}; 10 male, 10 female) were injected intrathoracically with 5000 FFU of CrPV-2, CrPV-S12, CrPV-S19, or PBS. Subsequently, flies were flipped onto standard media and survival was monitored daily. Shown is a graph representing the average from three separate biological experiments (±SEM). **B)** Flies were harvested at 5 days post injection (dpi) and homogenized in PBS. Viral titres were determined as described in materials and methods. Error bars represent the standard error of the mean. **C)** Western blot against CrPV VP2 on lysates from flies 5 dpi. Shown is a representative gel from at least three independent biological experiments.

Figure 10. Subcellular localization of CrPV ORFx. S2 cells were transfected with a construct expressing C-terminally HA-tagged ORFx or a mutant version of ORFx in the transmembrane region and incubated for 48 h. Following incubation, cells were fixed, permeabilized and co-stained with HA antibody and antibodies against nuclear lamina (Lamin), the Golgi apparatus (Golgin84), the ER (Calnexin 99A), and cytoplasm (alpha-

Tubulin). Shown are representative Z-stack micrographs from three independent experiments. Scale bars represent 15 µm.

Figure 10 – figure supplement 1. Cytotoxicity of CrPV ORFx expression in Drosophila S2 cells. A) The transmembrane helix property of CrPV ORFx. The predicted CrPV ORFx sequence stemming from the 13th codon was analyzed for transmembrane helices using ProtScale and HMMTOP. The mutated form of ORFx is shown on the right. Shown above are hydropathy plot outputs from ProtScale and below are HMMTOP predictions. The red amino acid region indicates the putative transmembrane helix predicted by HMMTOP. **B) Ectopic expression of CrPV ORFx causes slight cytotoxicity in S2 cells.** S2 cells (1.5×10^6) were transfected with expression constructs containing N- or C-terminally HA-tagged ORFx or empty vector and monitored over the course of 48 hours for cell death via trypan blue staining. * = p-value < 0.05.

Figure 10 – figure supplement 2. CrPV ORFx associates with membrane fractions. S2 cells were transfected with expression constructs encoding HA-tagged ORFx at the N- or C-terminus for 48 hours. Cells were then lysed and fractionated by differential centrifugation (see material and methods). Fractions were then analyzed by Western blot to detect the presence of CrPV ORFx or cellular markers. Shown is a representative gel from two independent biological experiments.

Figure 11. Model of CrPV IGR IRES mediated +1-frame ORFx translation. The majority of ribosomes bounds to the IGR IRES initiate translation in the 0-frame from the GCU start codon. A subset of ribosomes bound to the CrPV IGR IRES (~5%) initiates translation in the 0 frame and undergoes two pseudotranslocation events before repositioning downstream at the +1 frame 13th codon (AAA) to translate ORFx. The model proposes the AAA codon is vacant to allow delivery of the Lys-tRNA.

REFERENCES

Abaeva, I. S., Pestova, T. V., & Hellen, C. U. (2016). Attachment of ribosomal complexes and retrograde scanning during initiation on the Halastavi arva virus IRES. *Nucleic Acids Res, 44*(5), 2362-2377. doi: 10.1093/nar/gkw016

Abeyrathne, P. D., Koh, C. S., Grant, T., Grigorieff, N., & Korostelev, A. A. (2016). Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome. *Elife, 5*. doi: 10.7554/eLife.14874

Adhin, M. R., & van Duin, J. (1990). Scanning model for translational reinitiation in eubacteria. *J Mol Biol, 213*(4), 811-818. doi: 10.1016/S0022-2836(05)80265-7

Au, H. H., Cornilescu, G., Mouzakis, K. D., Ren, Q., Burke, J. E., Lee, S., . . . Jan, E. (2015). Global shape mimicry of tRNA within a viral internal ribosome entry site mediates translational reading frame selection. *Proc Natl Acad Sci U S A, 112*(47), E6446-6455. doi: 10.1073/pnas.1512088112

Au, H. H., & Jan, E. (2014). Novel viral translation strategies. *Wiley Interdiscip Rev RNA, 5*(6), 779-801. doi: 10.1002/wrna.1246

Baranov, P. V., Atkins, J. F., & Yordanova, M. M. (2015). Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. *Nat Rev Genet, 16*(9), 517-529. doi: 10.1038/nrg3963

Bonning, B. C., & Miller, W. A. (2010). Dicistroviruses. *Annu Rev Entomol, 55*, 129-150. doi: 10.1146/annurev-ento-112408-085457

Cevallos, R. C., & Sarnow, P. (2005). Factor-independent assembly of elongation-competent ribosomes by an internal ribosome entry site located in an RNA virus that infects penaeid shrimp. *J Virol, 79*(2), 677-683.

Chen, J., Coakley, A., O'Connor, M., Petrov, A., O'Leary, S. E., Atkins, J. F., & Puglisi, J. D. (2015). Coupling of mRNA Structure Rearrangement to Ribosome Movement during Bypassing of Non-coding Regions. *Cell, 163*(5), 1267-1280. doi: 10.1016/j.cell.2015.10.064

Chtarbanova, S., Lamiable, O., Lee, K. Z., Galiana, D., Troxler, L., Meignin, C., . . . Imler, J. L. (2014). Drosophila C virus systemic infection leads to intestinal obstruction. *J Virol, 88*(24), 14057-14069. doi: 10.1128/JVI.02320-14

Colussi, T. M., Costantino, D. A., Zhu, J., Donohue, J. P., Korostelev, A. A., Jaafar, Z. A., . . . Kieft, J. S. (2015). Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature, 519*(7541), 110-113. doi: 10.1038/nature14219

Costantino, D. A., Pfingsten, J. S., Rambo, R. P., & Kieft, J. S. (2008). tRNA-mRNA mimicry drives translation initiation from a viral IRES. *Nat Struct Mol Biol, 15*(1), 57-64. doi: nsmb1351 [pii]10.1038/nsmb1351

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol, 26*(12), 1367-1372. doi: 10.1038/nbt.1511

Dinman, J. D. (2012). Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip Rev RNA, 3*(5), 661-673. doi: 10.1002/wrna.1126

Dinos, G., Wilson, D. N., Teraoka, Y., Szaflarski, W., Fucini, P., Kalpaxis, D., & Nierhaus, K. H. (2004). Dissecting the ribosomal inhibition mechanisms of edeine and pactamycin: the universally conserved residues G693 and C795 regulate P-site RNA binding. *Mol Cell, 13*(1), 113-124.

Fernandez, I. S., Bai, X. C., Murshudov, G., Scheres, S. H., & Ramakrishnan, V. (2014). Initiation of Translation by Cricket Paralysis Virus IRES Requires Its Translocation in the Ribosome. *Cell*. doi: 10.1016/j.cell.2014.04.015

Firth, A. E., & Brierley, I. (2012). Non-canonical translation in RNA viruses. *J Gen Virol, 93*(Pt 7), 1385-1409. doi: 10.1099/vir.0.042499-0

1073 Firth, A. E., Wang, Q. S., Jan, E., & Atkins, J. F. (2009). Bioinformatic evidence for a stem-loop structure
1074 5'-adjacent to the IGR-IRES and for an overlapping gene in the bee paralysis dicistroviruses. *Virol*
1075 *J, 6*, 193. doi: 1743-422X-6-193 [pii]10.1186/1743-422X-6-193
1076 Garrey, J. L., Lee, Y. Y., Au, H. H., Bushell, M., & Jan, E. (2010). Host and viral translational mechanisms
1077 during cricket paralysis virus infection. *J Virol, 84*(2), 1124-1138. doi: JVI.02006-09
1078 [pii]10.1128/JVI.02006-09
1079 Hatakeyama, Y., Shibuya, N., Nishiyama, T., & Nakashima, N. (2004). Structural variant of the intergenic
1080 internal ribosome entry site elements in dicistroviruses and computational search for their
1081 counterparts. *RNA, 10*(5), 779-786.
1082 Herr, A. J., Gesteland, R. F., & Atkins, J. F. (2000). One protein from two open reading frames:
1083 mechanism of a 50 nt translational bypass. *EMBO J, 19*(11), 2671-2680. doi:
1084 10.1093/emboj/19.11.2671
1085 Hertz, M. I., & Thompson, S. R. (2011). In vivo functional analysis of the Dicistroviridae intergenic region
1086 internal ribosome entry sites. *Nucleic Acids Res, 39*(16), 7276-7288. doi: 10.1093/nar/gkr427
1087 Huang, W. M., Ao, S. Z., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R., . . . Fang, M. (1988). A persistent
1088 untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science,*
1089 *239*(4843), 1005-1012.
1090 Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation
1091 and principles of its regulation. *Nat Rev Mol Cell Biol, 11*(2), 113-127. doi: 10.1038/nrm2838
1092 Jan, E. (2006). Divergent IRES elements in invertebrates. *Virus Res, 119*(1), 16-28.
1093 Jan, E., Kinzy, T. G., & Sarnow, P. (2003). Divergent tRNA-like element supports initiation, elongation,
1094 and termination of protein biosynthesis. *Proc Natl Acad Sci U S A, 100*(26), 15410-15415.
1095 Jan, E., Mohr, I., & Walsh, D. (2016). A Cap-to-Tail Guide to mRNA Translation Strategies in Virus-
1096 Infected Cells. *Annu Rev Virol, 3*(1), 283-307. doi: 10.1146/annurev-virology-100114-055014
1097 Jan, E., & Sarnow, P. (2002). Factorless ribosome assembly on the internal ribosome entry site of cricket
1098 paralysis virus. *J Mol Biol, 324*(5), 889-902.
1099 Jang, C. J., & Jan, E. (2010). Modular domains of the Dicistroviridae intergenic internal ribosome entry
1100 site. *RNA, 16*(6), 1182-1195. doi: rna.2044610 [pii]10.1261/rna.2044610
1101 Jang, C. J., Lo, M. C., & Jan, E. (2009). Conserved element of the dicistrovirus IGR IRES that mimics an E-
1102 site tRNA/ribosome interaction mediates multiple functions. *J Mol Biol, 387*(1), 42-58. doi:
1103 S0022-2836(09)00093-X [pii]10.1016/j.jmb.2009.01.042
1104 Johannes, G., Carter, M. S., Eisen, M. B., Brown, P. O., & Sarnow, P. (1999). Identification of eukaryotic
1105 mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA
1106 microarray. *Proc Natl Acad Sci U S A, 96*(23), 13118-13123.
1107 Kerr, C. H., Ma, Z. W., Jang, C. J., Thompson, S. R., & Jan, E. (2016). Molecular analysis of the factorless
1108 internal ribosome entry site in Cricket Paralysis virus infection. *Sci Rep, 6*, 37319. doi:
1109 10.1038/srep37319
1110 Kerr, C. H., Wang, Q. S., Keatings, K., Khong, A., Allan, D., Yip, C. K., . . . Jan, E. (2015). The 5' untranslated
1111 region of a novel infectious molecular clone of the dicistrovirus cricket paralysis virus modulates
1112 infection. *J Virol, 89*(11), 5919-5934. doi: 10.1128/JVI.00463-15
1113 Kozak, M., & Shatkin, A. J. (1978). Migration of 40 S ribosomal subunits on messenger RNA in the
1114 presence of edeine. *J Biol Chem, 253*(18), 6568-6577.
1115 Krishna, N. K., Marshall, D., & Schneemann, A. (2003). Analysis of RNA packaging in wild-type and mosaic
1116 protein capsids of flock house virus using recombinant baculovirus vectors. *Virology, 305*(1), 10-
1117 24.
1118 Lang, B. F., Jakubkova, M., Hegedusova, E., Daoud, R., Forget, L., Brejova, B., . . . Nosek, J. (2014).
1119 Massive programmed translational jumping in mitochondria. *Proc Natl Acad Sci U S A, 111*(16),
1120 5926-5931. doi: 10.1073/pnas.1322190111

1121 Muhs, M., Hilal, T., Mielke, T., Skabkin, M. A., Sanbonmatsu, K. Y., Pestova, T. V., & Spahn, C. M. (2015).
1122 Cryo-EM of ribosomal 80S complexes with termination factors reveals the translocated cricket
1123 paralysis virus IRES. *Mol Cell, 57*(3), 422-432. doi: 10.1016/j.molcel.2014.12.016
1124 Nakashima, N., & Uchiumi, T. (2009). Functional analysis of structural motifs in dicistroviruses. *Virus Res,*
1125 *139*(2), 137-147. doi: S0168-1702(08)00230-X [pii]10.1016/j.virusres.2008.06.006
1126 Nishiyama, T., Yamamoto, H., Shibuya, N., Hatakeyama, Y., Hachimori, A., Uchiumi, T., & Nakashima, N.
1127 (2003). Structural elements in the internal ribosome entry site of Plautia stali intestine virus
1128 responsible for binding with ribosomes. *Nucleic Acids Res, 31*(9), 2434-2442.
1129 Nomura, M., Gourse, R., & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal
1130 components. *Annu Rev Biochem, 53*, 75-117. doi: 10.1146/annurev.bi.53.070184.000451
1131 Paek, K. Y., Hong, K. Y., Ryu, I., Park, S. M., Keum, S. J., Kwon, O. S., & Jang, S. K. (2015). Translation
1132 initiation mediated by RNA looping. *Proc Natl Acad Sci U S A, 112*(4), 1041-1046. doi:
1133 10.1073/pnas.1416883112
1134 Pestova, T. V., & Hellen, C. U. (2003). Translation elongation after assembly of ribosomes on the Cricket
1135 paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes*
1136 *Dev, 17*(2), 181-186.
1137 Petrov, A., Grosely, R., Chen, J., O'Leary, S. E., & Puglisi, J. D. (2016). Multiple Parallel Pathways of
1138 Translation Initiation on the CrPV IRES. *Mol Cell, 62*(1), 92-103. doi:
1139 10.1016/j.molcel.2016.03.020
1140 Pfingsten, J. S., Costantino, D. A., & Kieft, J. S. (2006). Structural basis for ribosome recruitment and
1141 manipulation by a viral IRES RNA. *Science, 314*(5804), 1450-1454. doi: 1133281
1142 [pii]10.1126/science.1133281
1143 Reinganum, C. (1975). The isolation of cricket paralysis virus from the emperor gum moth, Antheraea
1144 eucalypti Scott, and its infectivity towards a range of insect species. *Intervirology, 5*(1-2), 97-102.
1145 Ren, Q., Au, H. H., Wang, Q. S., Lee, S., & Jan, E. (2014). Structural determinants of an internal ribosome
1146 entry site that direct translational reading frame selection. *Nucleic Acids Res, 42*(14), 9366-9382.
1147 doi: 10.1093/nar/gku622
1148 Ren, Q., Wang, Q. S., Firth, A. E., Chan, M. M., Gouw, J. W., Guarna, M. M., . . . Jan, E. (2012). Alternative
1149 reading frame selection mediated by a tRNA-like domain of an internal ribosome entry site. *Proc*
1150 *Natl Acad Sci U S A, 109*(11), E630-639. doi: 10.1073/pnas.1111303109
1151 Ruehle, M. D., Zhang, H., Sheridan, R. M., Mitra, S., Chen, Y., Gonzalez, R. L., . . . Kieft, J. S. (2015). A
1152 dynamic RNA loop in an IRES affects multiple steps of elongation factor-mediated translation
1153 initiation. *Elife, 4*. doi: 10.7554/eLife.08146
1154 Samatova, E., Konevega, A. L., Wills, N. M., Atkins, J. F., & Rodnina, M. V. (2014). High-efficiency
1155 translational bypassing of non-coding nucleotides specified by mRNA structure and nascent
1156 peptide. *Nat Commun, 5*, 4459. doi: 10.1038/ncomms5459
1157 Sarabhai, A., & Brenner, S. (1967). A mutant which reinitiates the polypeptide chain after chain
1158 termination. *J Mol Biol, 27*(1), 145-162.
1159 Schuler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., . . . Spahn, C. M.
1160 (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat Struct Mol Biol,*
1161 *13*(12), 1092-1096. doi: nsmb1177 [pii]10.1038/nsmb1177
1162 Skabkin, M. A., Skabkina, O. V., Hellen, C. U., & Pestova, T. V. (2013). Reinitiation and other
1163 unconventional posttermination events during eukaryotic translation. *Mol Cell, 51*(2), 249-264.
1164 doi: 10.1016/j.molcel.2013.05.026
1165 Speicher, K. D., Gorman, N., & Speicher, D. W. (2009). N-terminal sequence analysis of proteins and
1166 peptides. *Curr Protoc Protein Sci, Chapter 11*, Unit11 10. doi: 10.1002/0471140864.ps1110s57
1167 Vassilaki, N., & Mavromara, P. (2009). The HCV ARFP/F/core+1 protein: production and functional
1168 analysis of an unconventional viral product. *IUBMB Life, 61*(7), 739-752. doi: 10.1002/iub.201

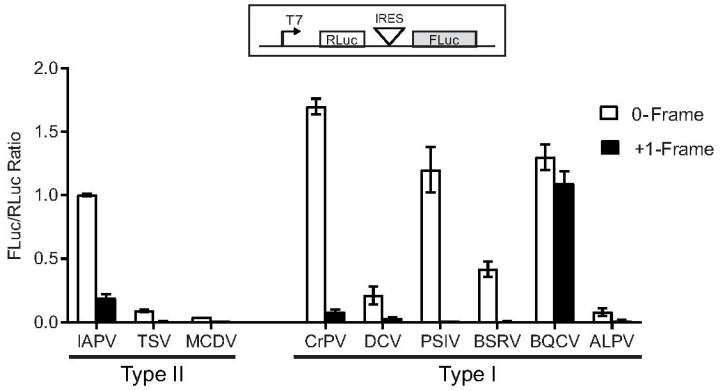
1169 Wang, Q. S., Au, H. H., & Jan, E. (2013). Methods for studying IRES-mediated translation of positive-
1170 strand RNA viruses. *Methods, 59*(2), 167-179. doi: 10.1016/j.ymeth.2012.09.004
1171 Wills, N. M., O'Connor, M., Nelson, C. C., Rettberg, C. C., Huang, W. M., Gesteland, R. F., & Atkins, J. F.
1172 (2008). Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA.
1173 *EMBO J, 27*(19), 2533-2544. doi: 10.1038/emboj.2008.170
1174 Wilson, J. E., Pestova, T. V., Hellen, C. U., & Sarnow, P. (2000). Initiation of protein synthesis from the A
1175 site of the ribosome. *Cell, 102*(4), 511-520.
1176 Yamamoto, H., Wittek, D., Gupta, R., Qin, B., Ueda, T., Krause, R., . . . Nierhaus, K. H. (2016). 70S-
1177 scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria.
1178 *Proc Natl Acad Sci U S A, 113*(9), E1180-1189. doi: 10.1073/pnas.1524554113
1179 Yamasaki, K., Weihl, C. C., & Roos, R. P. (1999). Alternative translation initiation of Theiler's murine
1180 encephalomyelitis virus. *J Virol, 73*(10), 8519-8526.
1181 Zhang, H., Ng, M. Y., Chen, Y., & Cooperman, B. S. (2016). Kinetics of initiating polypeptide elongation in
1182 an IRES-dependent system. *Elife, 5*. doi: 10.7554/eLife.13429
1183
1184

Kerr_Figure 1

A

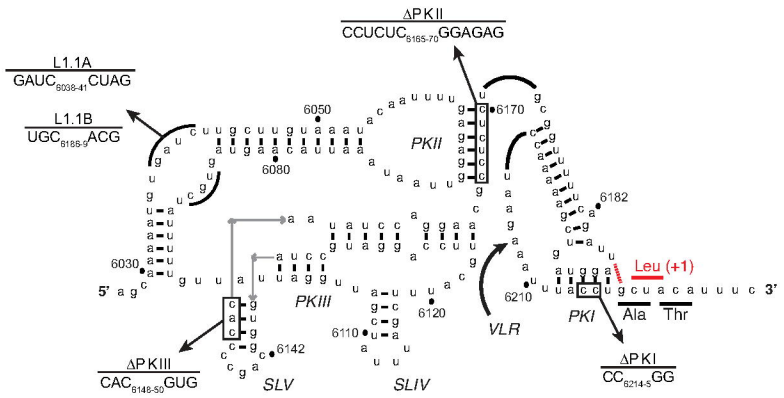
Dicistrovirus	IGR IRES Type	Potential +1 basepair	Predicted +1 ORFx Length (aa)	First +1 frame amino acid
Solenopsis invicta virus-1 (SINV-1)	II	C-G	125	L
Israeli acute paralysis virus (IAPV)	II	U-G	94	A
Kashmir bee virus (KBV)	II	C-G	93	L
Acute bee paralysis virus (ABPV)	II	U-G	92	P
Cricket paralysis virus (CrPV)	I	U-G	53	L
Drosophila C virus (DCV)	I	U-G	51	L
Plautia stali intestine virus (PSIV)	I	U C (?)	43	K
Big Sioux River virus (BSRV)	I	U-G	40	L
Black queen cell virus (BQCV)	I	U-G	26	L
Homalodisca coagulata virus- 1	I	U-G	17	Q
Taura Syndrome virus (TSV)	II	U-G	16	L
Himetobi P virus (HiPV)	I	U-G	14	Q
Bat guano dicistrovirus	I	U-G	14	Q
Mud crab dicistrovirus (MCDV)	II	U-G	9	L
Aphid lethal paralysis virus (ALPV)	I	U-G	7	L
ALPV like Brookings virus	I	U-G	7	Q
Rhopalosiphum padi virus (RhPV)	I	U-G	2	Q
Triatoma virus (TrV)	II	C-G	1	L

B

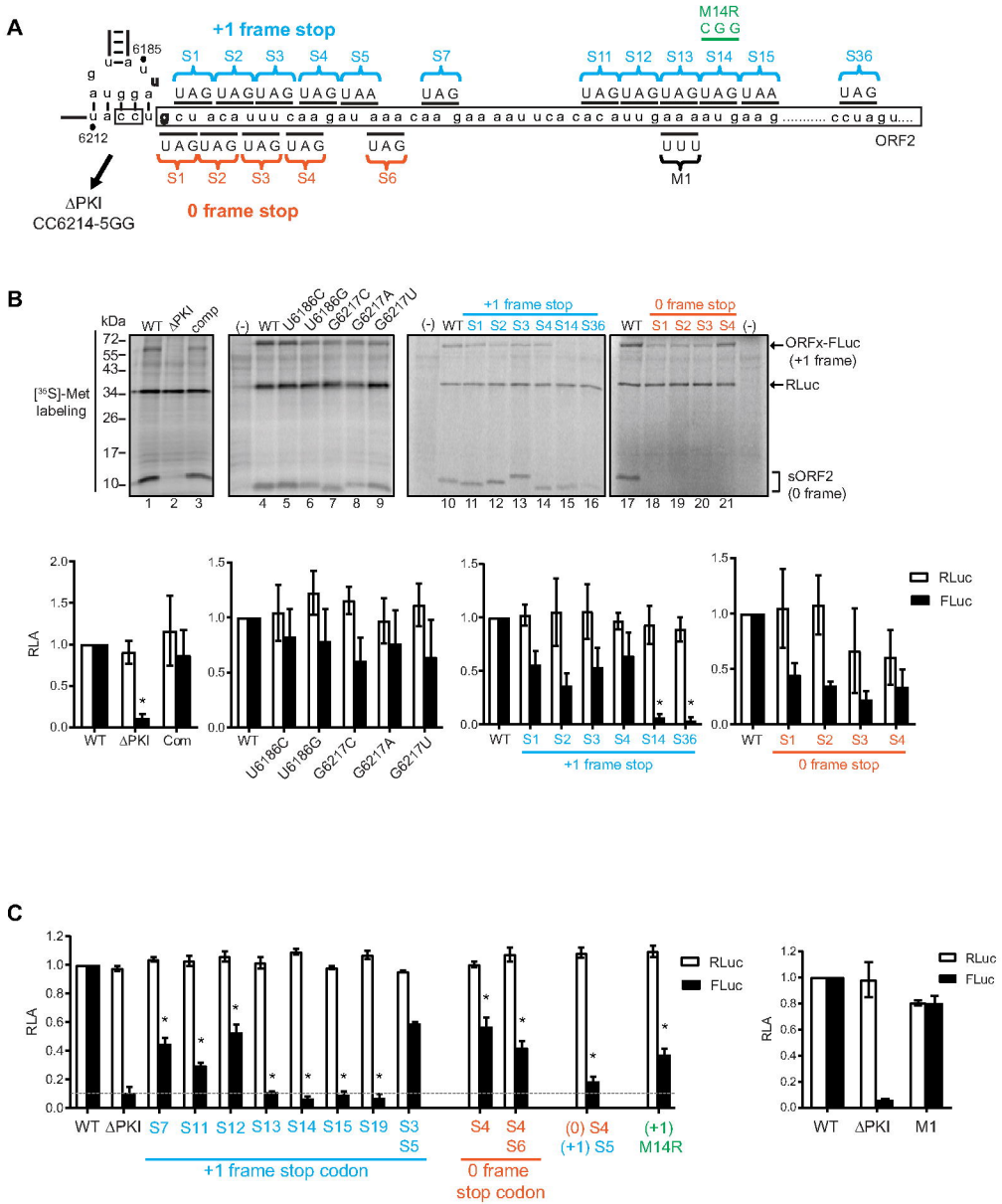


C

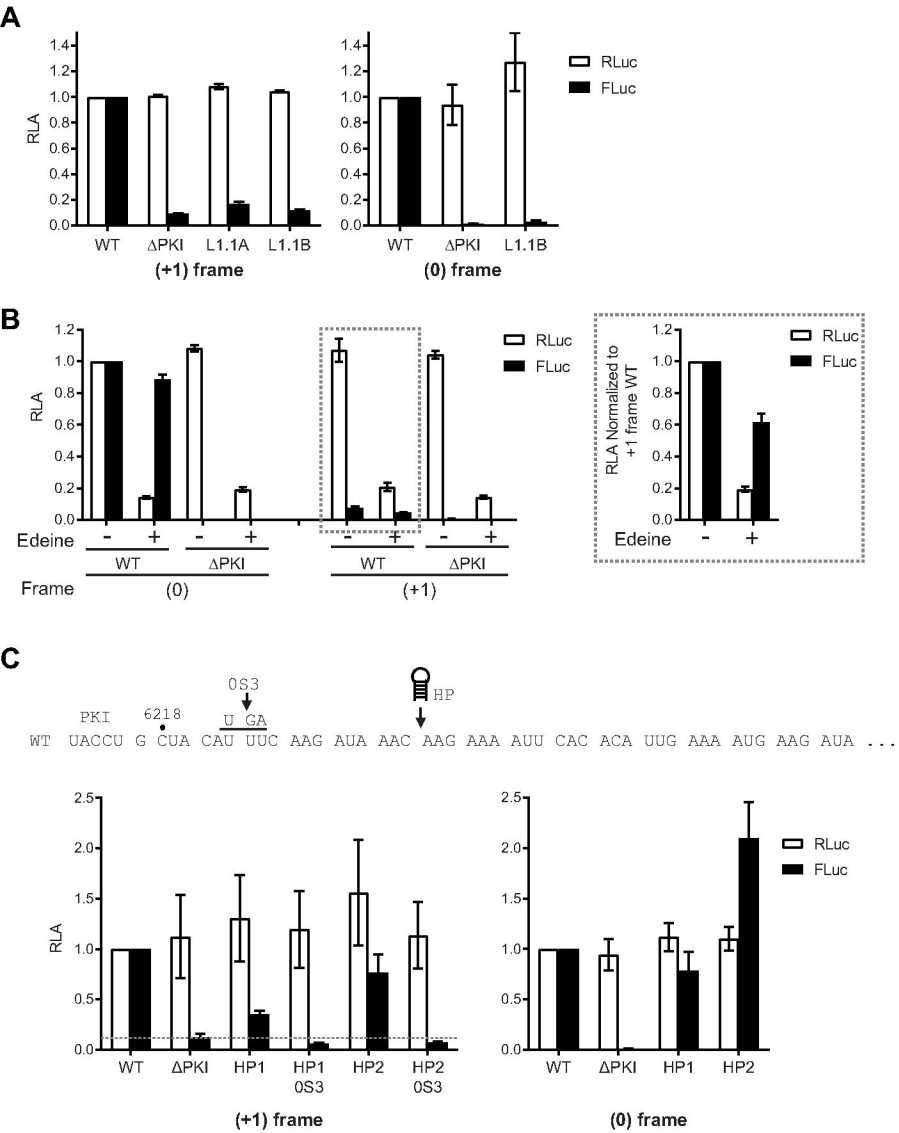
Cricket paralysis virus IGR IRES



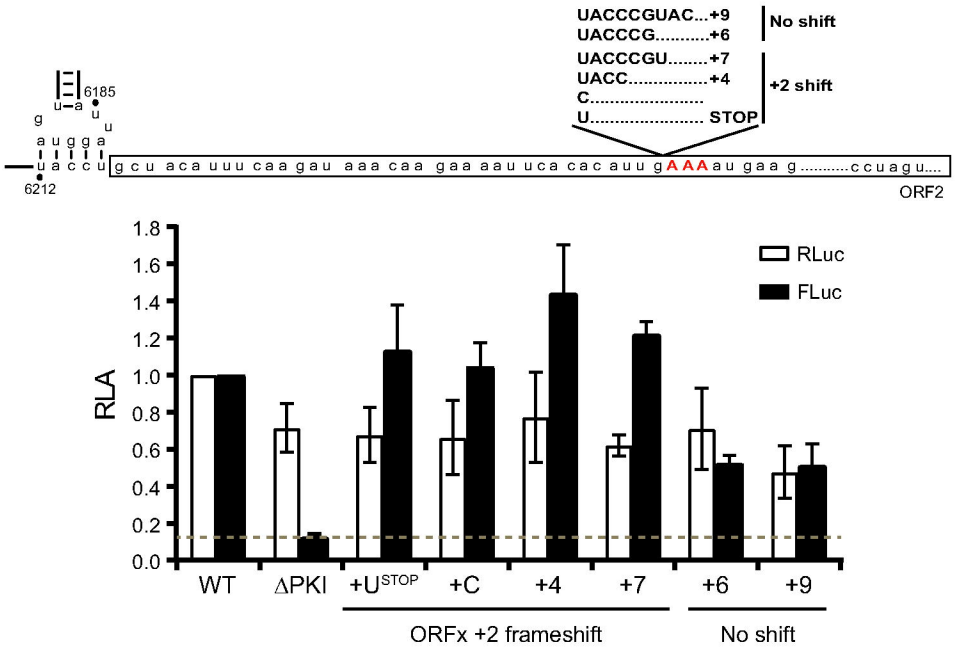
Kerr_Figure 2



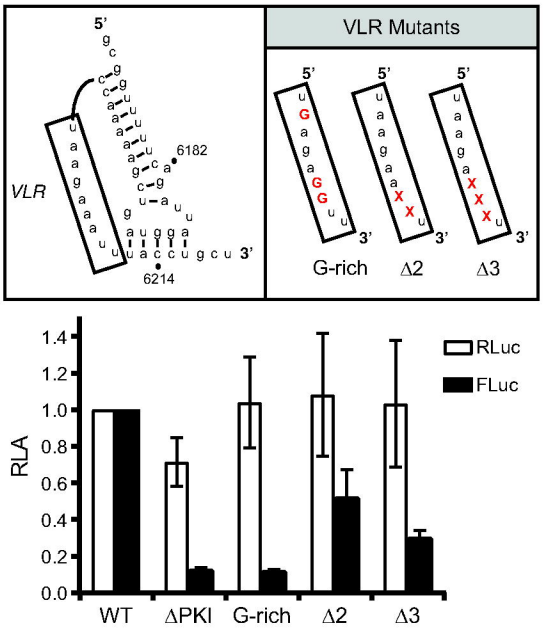
Kerr_Figure 3



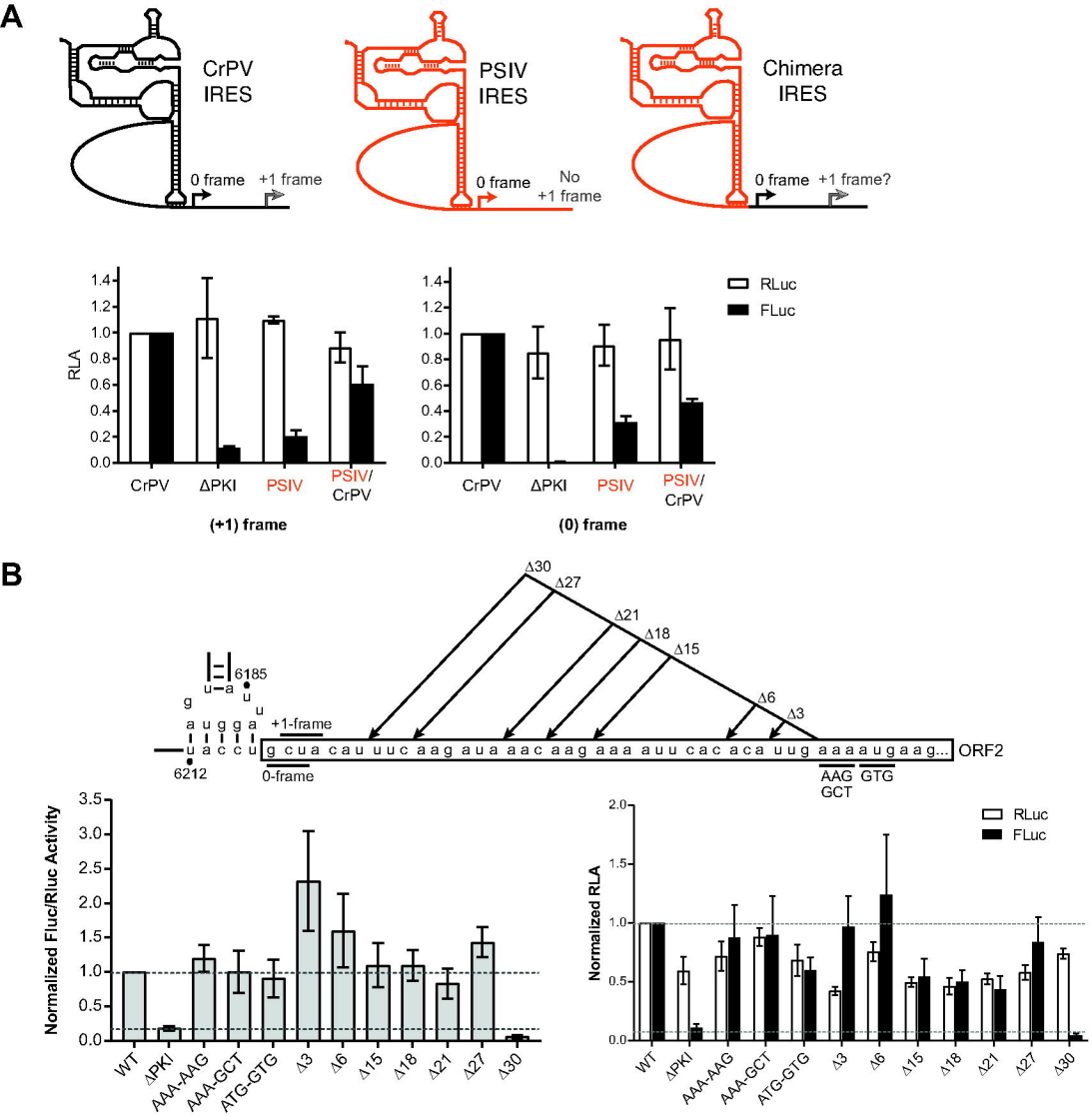
Kerr_Figure 4



Kerr_Figure 5



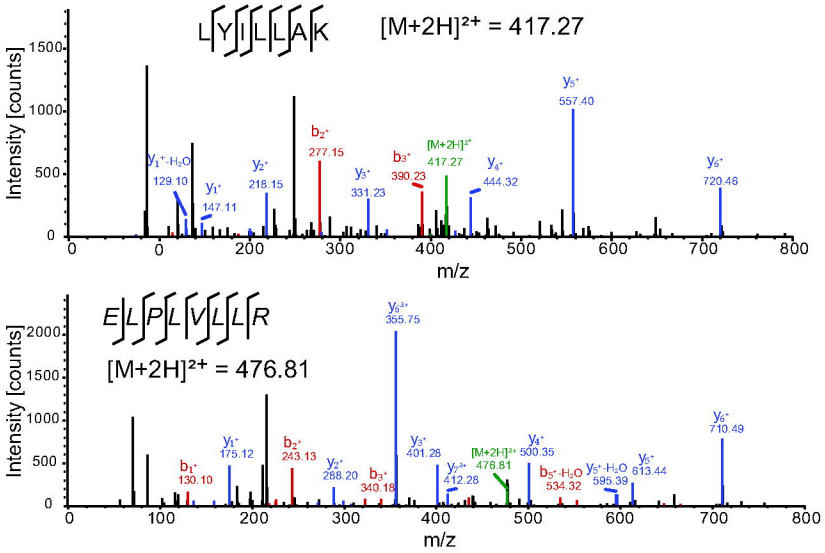
Kerr_Figure 6



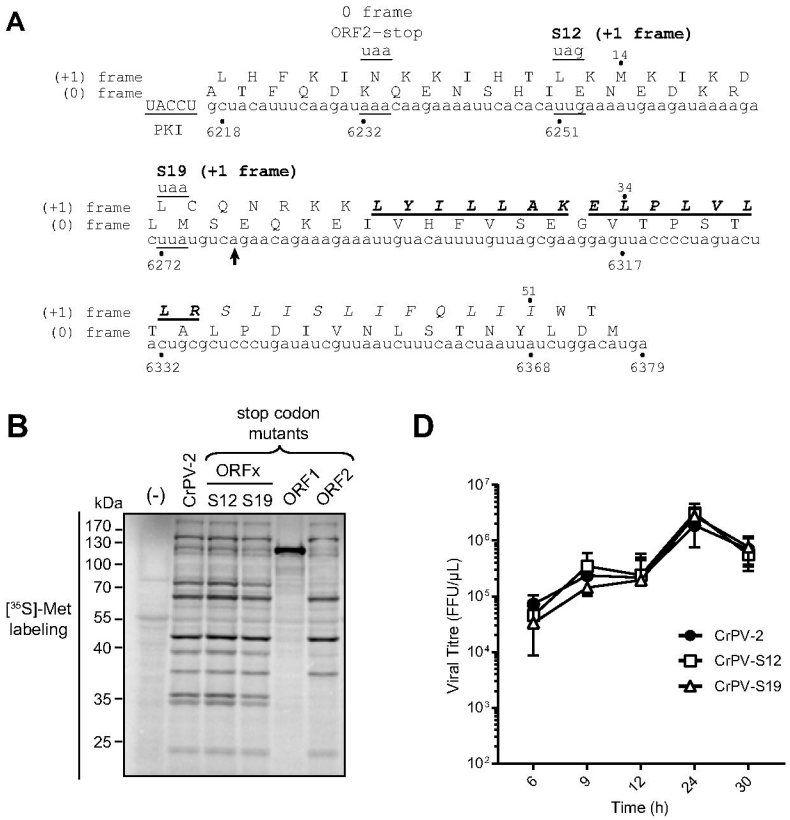
Kerr_Figure 7

Peptide #1 Peptide #2

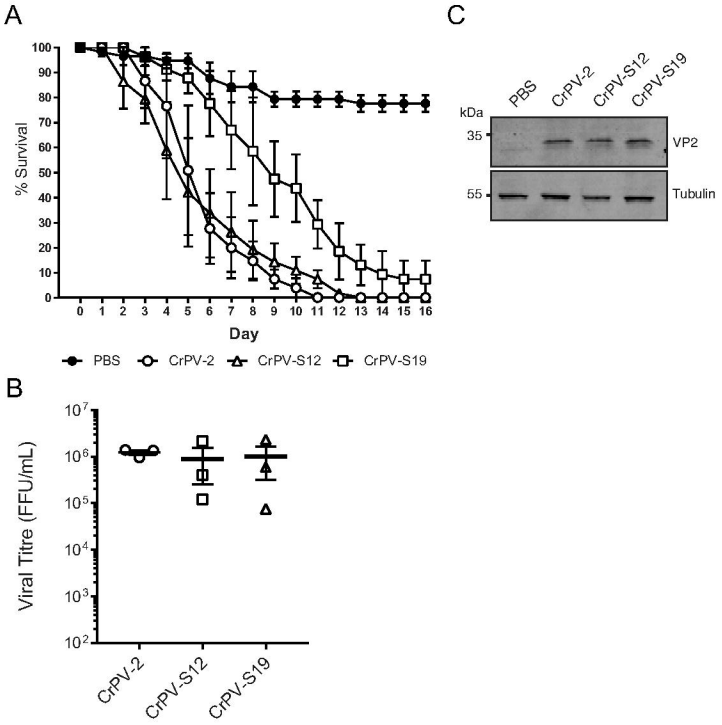
Predicted Sequence: N - *AL* **K**MKIKDLCQNRKK **LYILLAK ELPLVLLR** SLISLIFQLIIW - **C**



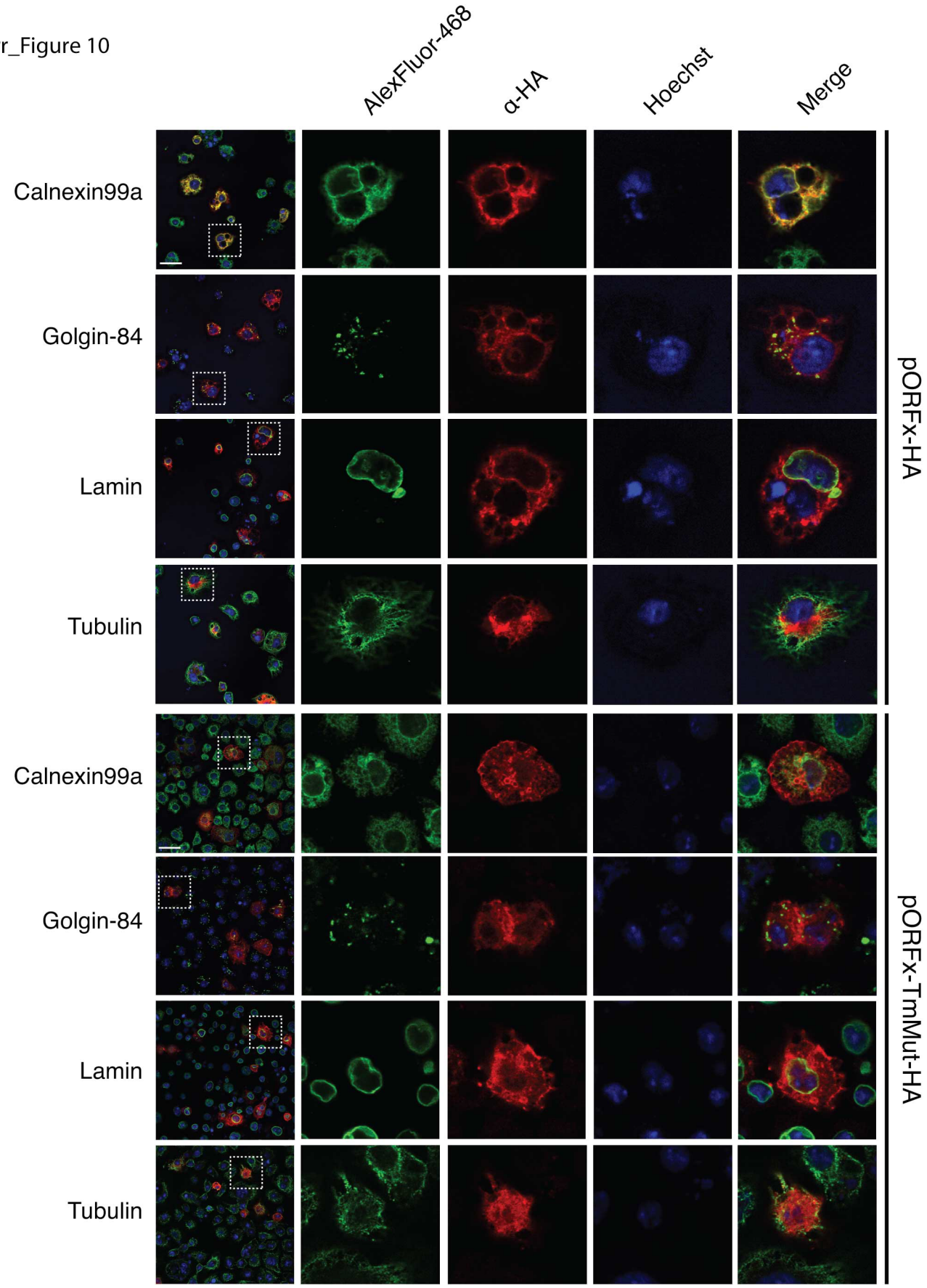
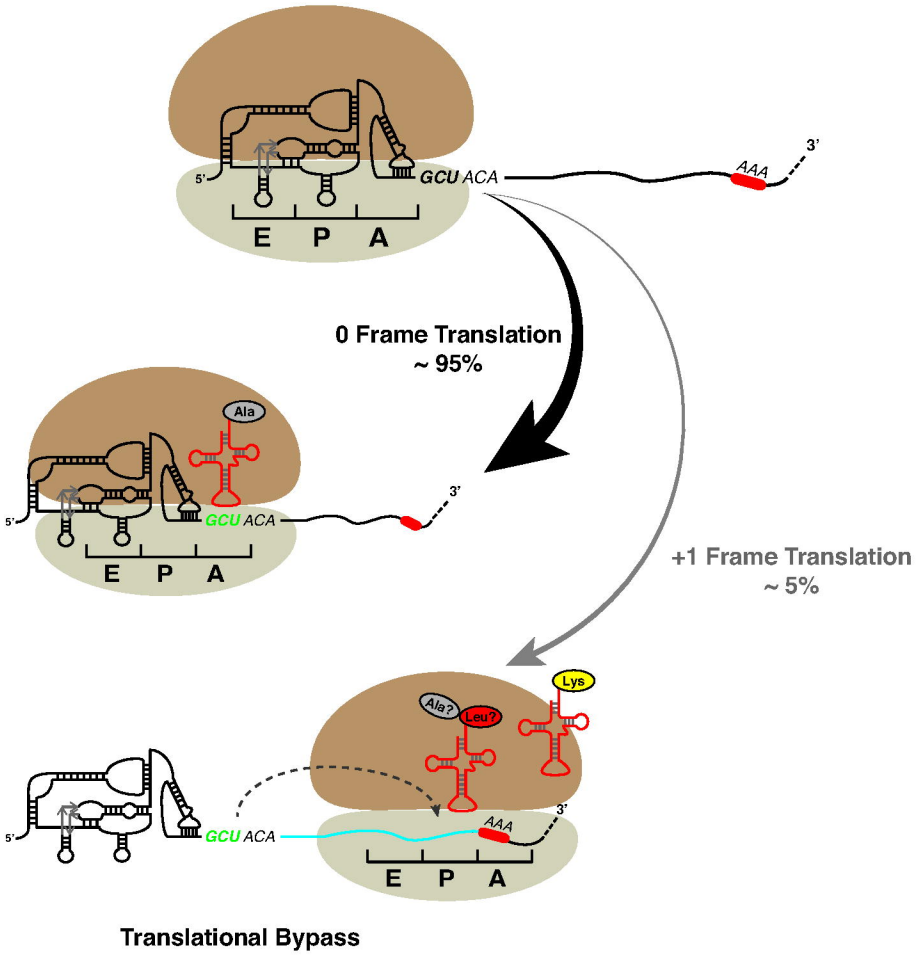
Kerr_Figure 8



Kerr_Figure 9



Kerr_Figure 10



Kerr_Figure 11