HHS Public Access

Author manuscript

*Nature*. Author manuscript; available in PMC 2015 September 05.

Published in final edited form as:

*Nature*. 2015 March 5; 519(7541): 110–113. doi:10.1038/nature14219.

# Initiation of Translation in Bacteria by a Structured Eukaryotic IRES RNA

**Timothy M. Colussi**[1,2,4,#], **David A. Costantino**[1,2,#], **Jianyu Zhu**[3,5], **John Paul Donohue**[3], **Andrei A. Korostelev**[3,6], **Zane A. Jaafar**[1], **Terra-Dawn M. Plank**[1,7], **Harry F. Noller**[3], and **Jeffrey S. Kieft**[1,2,*]

[1]Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO, USA, 80045, USA

[2]Howard Hughes Medical Institute, University of Colorado Denver School of Medicine, Aurora, CO, USA, 80045, USA

[3]Center for Molecular Biology of RNA and Department of Molecular, Cell and Developmental Biology, Sinsheimer Labs, University of California at Santa Cruz, Santa Cruz, CA 95064, USA

## Abstract

The central dogma of gene expression (DNA→RNA→protein) is universal, but in different domains of life there are fundamental mechanistic differences within this pathway. For example, the canonical molecular signals used to initiate protein synthesis in bacteria and eukaryotes are mutually exclusive[1,2]. However, the core structures and conformational dynamics of ribosomes that are responsible for the steps of translation following initiation are ancient and conserved across the domains of life[3,4]. We asked whether an undiscovered RNA-based signal might be able to use these conserved features, bypassing mechanisms specific to each domain of life, and initiate protein synthesis in both bacteria and eukaryotes. Although structured internal ribosome entry site (IRES) RNAs can manipulate ribosomes to initiate translation in eukaryotic cells, an analogous RNA structure-based mechanism has not been observed in bacteria. Here, we report our discovery that a eukaryotic viral IRES can initiate translation in live bacteria. We solved the crystal structure



*To whom correspondence should be addressed: Jeffrey S. Kieft, HHMI/Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Mail Stop 8101, Aurora, CO 80045, Telephone: 303-724-3257, Fax: 303-724-3215, Jeffrey.Kieft@ucdenver.edu.

[4]Current address: Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA

[5]Current address: Cocrystal Discovery, Inc., Mountain View, CA 94043, USA

[6]Current address: RNA Therapeutics Institute, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA, 01655, USA

[7]Current address: Department of Reproductive Medicine, University of California at San Diego, La Jolla CA 92093, USA

[#]Authors contributed equally to this work.

**Author contributions**

T.M.C. and J.S.K. designed the experiments and the constructs tested. T.M.C. and D.A.C. conducted the bacterial functional assays. Clones were generated by T.M.C., T-D.M.P. and Z.A.J. J.S.K. performed the ribosome association assays. Ribosomes were purified, crystals grown, and the structure solved by J.P.D., J.Z., and A.A.K. under the supervision of H.F.N. J.S.K. provided overall supervision and guidance, and together with T.M.C. and D.A.C. wrote the manuscript with input from all authors.

Atomic coordinates and structure factor amplitudes have been deposited at the Protein Data Bank under accession number 4XEJ.

**Online Content.** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Supplementary Information** is available in the online version of the paper.

The authors declare no competing financial interests.

of this IRES bound to a bacterial ribosome to 3.8 Å resolution, revealing that despite differences between bacterial and eukaryotic ribosomes this IRES binds directly to both and occupies the space normally used by tRNAs. Initiation in both bacteria and eukaryotes depends on the structure of the IRES RNA but in bacteria this RNA uses a different mechanism that includes a form of ribosome repositioning after initial recruitment. This IRES RNA bridges billions of years of evolutionary divergence as an example of an RNA structure-based translation initiation signal capable of operating in two domains of life.

**Keywords**

Internal ribosome entry site (IRES); X-ray crystallography; 70S ribosome; RNA structure; bacterial initiation

Bacteria cannot recognize the “cap” on the 5’ end of eukaryotic mRNAs and eukaryotic ribosomes cannot use the Shine-Dalgarno sequence (SDS)[1] (Extended Data Fig. 1a). Although non-canonical mechanisms exist[2,3], there is no known translation initiation signal that can operate in multiple domains of life at any location in an mRNA. Despite this divergence there is strong conservation in the functional core of the ribosome, where mRNA and tRNAs interact and move[4]. In fact, the tRNAs used in elongation from bacteria and eukaryotes are interchangeable[5]. Therefore, we asked whether a structured RNA embedded in an mRNA sequence could interact with conserved ribosome features in the decoding groove and initiate translation in both bacteria and eukaryotes. Candidates for such RNAs are the intergenic region internal ribosome entry sites (IGR IRESs) from *Dicistroviridae* viruses. In eukaryotes, these IRESs act independently of a 5’ cap[6], adopt a functionally essential compact fold that docks within the ribosome[7-9] without initiation factors or a start codon[10-16], and partially mimic tRNA (Extended Data Fig. 1b&c)[12,17-19]. It is proposed that they drive translation initiation by co-opting the ribosome’s conserved elongation cycle[17,19-22], and they operate in diverse eukaryotic systems[6,23].

We generated an inducible expression vector encoding a single mRNA containing two independent luciferase (LUC) reporters (Extended Data Fig. 1d)[24], and verified that it allowed simultaneous measurement of initial rates of production of each protein (Extended Data Fig. 2&3). We used this construct to test if an IGR IRES RNA can drive translation in live bacteria. The *Renilla* luciferase (RLUC) was placed to initiate translation from a SDS (and “enhancer” sequence), and the Firefly luciferase (FLUC) was placed after a Wild-type (WT) *Plautia stali* intestine virus (PSIV) IGR IRES. There was some production of both LUCs prior to induction (due to expected “leaky expression”, Extended Data Fig. 4), but induction resulted in marked increase in both reporters; the production of FLUC is consistent with translation beginning at the IRES (Fig. 1c; Extended Data Fig. 2). Removing the RLUC-driving SDS (Upstream SDS_K/O; all mutants shown in Extended Data Fig. 5) diminished production of RLUC, but FLUC production increased >10-fold (Fig. 1b; all raw LUC data in Extended Data Table 1a), attributable to decreased competition for ribosomes and with ribosomes initiating independently at the IRES. Replacing the IGR IRES with the IRES from classical swine fever virus (CSFV) resulted in negligible FLUC production (Extended Data Fig. 2), demonstrating specificity for the IGR IRES.

A source of initiation from the IGR IRES could be a “cryptic” SDS in the purine-rich sequence between the IRES and the FLUC start codon (Extended Data Fig. 6). FLUC production from this SDS-like sequence alone was at ~30% of the WT IRES, not enough to account for all FLUC produced from the IRES. Mutating this SDS-like sequence in the context of the full IRES decreased FLUC production, but translation was still higher than from an SDS or the SDS-like sequence. Thus, the structured IRES can drive FLUC production without the SDS-like sequence, but both likely contribute to function when present.

To determine the structural basis for IGR IRES activity in bacteria, we solved the crystal structure of the full-length IRES RNA•70S ribosome complex to 3.8 Å resolution. In eukaryotes, IGR IRES domain 1+2 contacts both subunits, while domain 3 mimics an mRNA/tRNA interaction on the small subunit (Extended Data Fig. 1b)[7,8,10,11,19,25]. We observed electron density for domain 3 in the P site as in the crystal structure of isolated domain 3 bound to 70S ribosomes[19] (Fig. 2a; Extended Data Fig. 7); this may represent an initiation-state or translocated IRES. Domain 1+2 density was weak but its location could be modeled using the crystal structure of unbound PSIV IGR IRES domain 1+2[26] (Fig. 2a). Domain 1+2’s location in the 70S ribosome differs from IGR IRES•80S ribosome complexes with domain 3 in the A site[22,27]. In 80S ribosomes, domain 1+2 interacts with eukaryotic-specific ribosomal protein eS25 and the L1 stalk[10,11,28,29], which is structurally distinct from that in bacterial ribosomes[30]. In the full-length IRES•70S structure, the L1 stalk is displaced ~15Å compared to the structure containing domain 3 only. (Fig. 2b). The absence of eS25 and differences in the L1 stalk may be responsible for the partial disorder and location of the IRES. Nonetheless, the structure clearly illustrates that the compactly folded IRES can bind in the tRNA binding sites of bacterial ribosomes.

The IGR IRES’ compact structure is essential for function in eukaryotes[25,26], and the IRES•70S structure suggested this is true in bacteria. To test this, we disrupted two pseudoknots essential for the IRES’ compact structure, both individually (PK1_K/O, PK2_K/O) and together (PK1+PK2_K/O) and measured activity (Fig. 3a&b; Extended Data Fig. 8a)[10]. FLUC production decreased in all three, with the double mutant at a level where activity could be accounted for by the cryptic SDS-like sequence. Indeed, disruption of both pseudoknots and the SDS-like sequence (Downstream SDS-like_K/O+PK1+PK2_K/O) abrogated IRES activity (Extended Data Fig. 6). Isolated IRES domain 3 operated similarly to the domain 1+2-disrupting mutant (PK2_K/O). Thus, IGR IRES translation in bacteria depends on a compact RNA structure and although domain 1+2 is poorly ordered in the crystal, it may be required to form transient interactions with the ribosome.

We explored the putatively transient IGR IRES•70S interactions using translationally competent cell-free extracts. In rabbit reticulocyte lysate (RRL; positive control) the IRES forms 80S ribosomes both in the presence and absence of a non-hydrolyzable GTP analog (GMPPNP) (Fig. 3c). In contrast, 70S formation on the IRES in *E. coli* lysate was virtually undetectable (Fig. 3d). We repeated the experiment with an IRES RNA containing the FLUC AUG and several codons downstream of the IRES to allow initiation to occur and stabilize the resultant complexes. Both IRES•70S complexes and IRES•30S complexes formed in the presence of elongation inhibitor hygromycin B (Fig. 3e). In the *E. coli* lysate,

the amount of IRES-ribosome complex is low compared to that observed for the RRL, consistent with formation of an unstable or transient complex.

In eukaryotes, IGR IRES-driven translation begins directly on the IRES and is proposed to co-opt the ribosome's elongation cycle[17,19,21,22]; we asked if this is true in bacteria where the IRES-ribosome interactions appear different and transient. Removal of the FLUC start codon located 15 nucleotides downstream of the IRES structure (ΔAUG) resulted in a complete loss of FLUC production, while a stop codon placed upstream of the FLUC start codon (uSTOP) had little effect (Fig. 4a&b; Extended Data Fig. 8b). Removal of 1 or 2 nucleotides just upstream of the FLUC AUG [F-SHIFT(−1) and F-SHIFT(−2)] had little effect. These results are consistent with translation in bacteria beginning on the FLUC AUG, not directly at the IRES on a non-AUG codon. This implies a repositioning of the ribosome from the IRES to the FLUC start codon. To explore this, we created a construct with an out-of-frame start codon between the IRES and the start codon (uAUG); this mutation decreased activity but not to the degree expected if this codon were being used efficiently. The source of this discrimination is not clear, but we posit that selection of the FLUC AUG is assisted by the nearly ideally positioned cryptic SDS-like sequence upstream. Constructs with alterations between the IRES and FLUC start codon all had decreased activity in the context of the PK1+PK2_K/O mutation (Extended Data Fig. 9), indicating that IRES structural integrity remains necessary for their function.

The mechanism of this IRES in bacteria is more primitive than in eukaryotes. We propose that the structured IRES RNA forms interactions with bacterial ribosomes that are transient and weaker than the highly-tuned interactions that occur in eukaryotes, but allow internal entry of the ribosome to the message. Recruited subunits or ribosomes are repositioned to a downstream start codon where protein synthesis starts. That a compact IRES RNA can use this primitive mechanism suggests that RNA structure-driven or -assisted initiation may be used in potentially in all domains of life, driven by diverse RNAs perhaps possessing tRNA-like character or decoding groove binding capability.

## Methods

### Plasmid construction

DNA containing the *Plautia stali* intestine virus (PSIV) IGR IRES (nts 6000-6195) between the RLUC and FLUC genes was ligated into the KpnI and SacI sites of a pET30a vector (Novagen) using T4 DNA Ligase (New England Biolabs). The resultant construct contained 15 nucleotides of sequence between the 3' end of the IRES (designated as the 3' end of pseudoknot 1, nt 6195) and the AUG start codon of the FLUC ORF.

Mutants were generated using several methods:

1. PCR with appropriate forward and reverse primers (IDT) was used to generate two halves of the desired sequence. The halves were annealed and amplified by PCR using the T7 and T7 terminator sequencing primers. The resultant DNA was then ligated into the same pET30a vector using the above restriction sites.

2. Site-directed mutagenesis using the QuikChange (Agilent) mutagenesis strategy using appropriate primer pairs.

3. Insertion of PCR amplified DNA or synthesized gBlock gene fragments (IDT) into the dual-LUC- containing pET30a vector between the SpeI and NcoI sites (between the *Renilla* and Firefly genes) using a ligation-independent cloning (LIC) method, In-Fusion HD Cloning Plus (Clontech Laboratories). PCR products or gBlocks contained sequence overlapping 12 base pairs on the 5' of the SpeI site and 12 base pairs 3' of the NcoI site of the vector. Assembled constructs maintained both restriction sites.

4. For the T7 knockout construct, a pET30a vector containing a mutated T7 promoter (TAAATGGTGTCTGAATTC) was synthesized (DNA 2.0) and DNA coding for the WT PSIV flanked by the two LUC genes was amplified by PCR. The PCR product was inserted between the KpnI and SacI sites in the mutated T7 vector by LIC.

5. The mutant in which the PSIV IGR IRES was replaced by a SDS (without enhancer sequence) was generated by ligating the DNA fragment into the pET30a/dual-LUC vector using the SpeI and NcoI sites.

### Bacterial cell culture

Rosetta DE3 cells (Novagen) were transformed with the plasmids described above and grown overnight in 5 mL Luria Broth (LB) with kanamycin (Fisher) at 37°C with constant agitation to generate a starter culture. To start the experiment, 50 mL of LB containing kanamycin was inoculated with 1 mL of the overnight starter culture. The 50 mL cultures were grown with agitation at 37°C to an absorbance at 600 nm of 0.6 (measured on a Thermo Scientific NanoDrop 2000c spectrophotometer). The cultures were induced with 1 mM IPTG (Gold Bio) and allowed to grow for 4 hours. Samples (50 μL) were taken at 10-30 minute intervals.

### Measurement of LUC activity

At each time point, 50 μL of cell culture was removed, the cells were pelleted by centrifugation, and the supernatant was removed. Cells were resuspended in 300 μL 1X Passive Lysis Buffer (PLB, Promega). 20 μL of the resultant cell lysate was added to a 96-well microplate (Greiner Bio-One). The dual-LUC assay was performed by first adding 100 μL LAR II (Promega) to measure FLUC activity, then 100 μL of Stop & Glo reagent (Promega) was added to measure RLUC activity. The assay was performed and measurements were taken using a Promega Glomax Multi+ detection system.

### Determination of IRES activity

FLUC and RLUC activity (expressed as Relative Light Units; RLU's) were graphed as a function of time for each culture using the program KaleidaGraph. The initial rate of FLUC and RLUC production was determined using the data from the first 30-40 minutes post-induction. LUC production was generally linear over this time scale after a 5-10 minute lag. IRES activity was then calculated as the ratio of the initial rate of FLUC to RLUC for each

culture. Ratios from individual independent cultures were averaged. Bar graphs represent averages from at least three independent cultures; error bars depict one standard error from the mean. This method corrects for variation in growth, induction, and potential protein stability differences between cultures.

### RNA transcription and purification for ribosome assembly assay

DNA templates for *in vitro* transcription were generated by PCR using a plasmid containing the WT PSIV IRES as the template and primers designed to amplify just the DNA of interest under control of a T7 RNA polymerase promoter. The resultant PCR-generated DNA template was used in *in vitro* transcription reactions. RNA was purified from raw transcription reactions by HPLC. The first RNA used in assembly assays contained nt 6000-6195 of the PSIV IGR IRES, and the second contained this same sequence, plus the sequence GAAAAAGAATTTACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCGCTGGAGAGC downstream of the IRES.

### Ribosome assembly assay

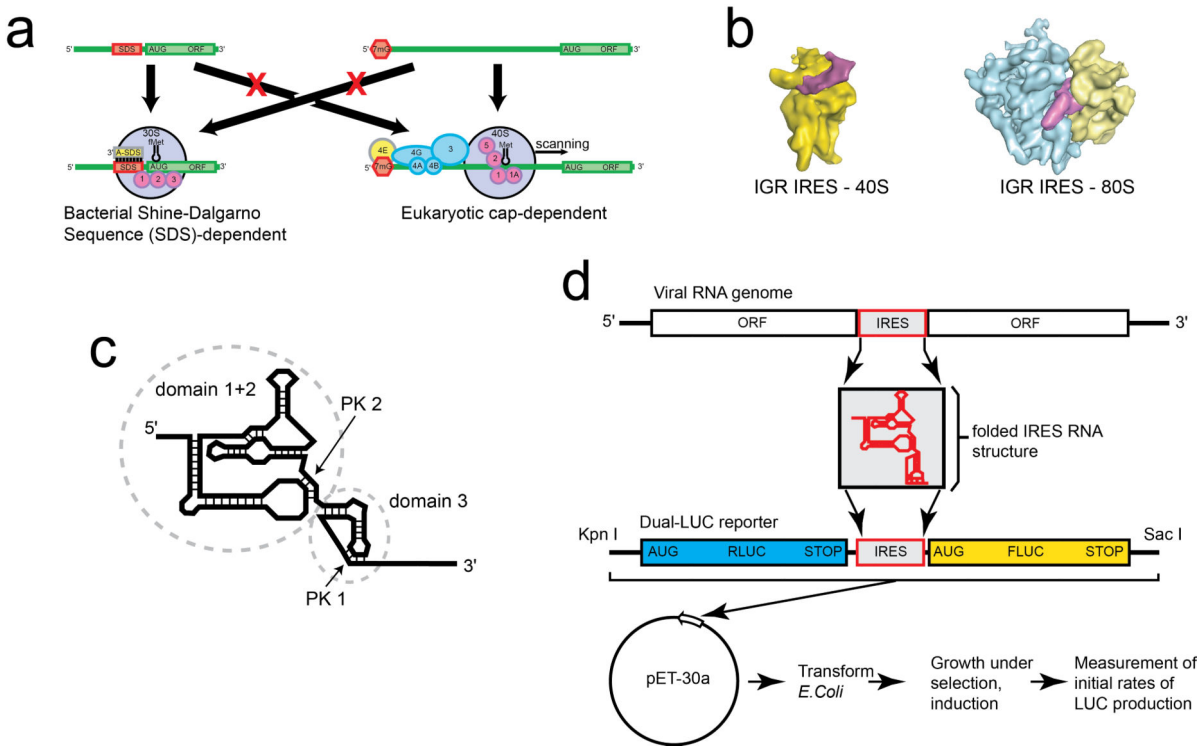
RNA for use in assembly assays was 5’-end radiolabeled with P-32 using T4 polynucleotide kinase (NEB), purified by gel electrophoresis, and diluted to 1000 CPM/μL. For the assays in RRL, 1 μL radiolabeled RNA was combined with 30 μL of lysate supplemented with amino acids. For the reaction with GMPPNP, 5 μL of a 20 mM stock of the analog was added to achieve a final concentration of 2 mM, and an equimolar amount of $MgCl_2$ was added. For the reaction with hygromycin B, 2 μL of a 50 mg/mL stock was added to a final concentration of 2 mg/mL. RNase-free water was added to a total final volume of 50 μL. For the reactions in *E. coli* lysate, Promega product #L1030 was used. 1 μL of labeled RNA was added to 15 μL of lysate and 20 μL of S30 premix supplemented with 5 μL of the amino acid mix and 1 μL of RNasin RNase inhibitor (Promega). For the reactions with GMPPNP or antibiotic, the same amounts were added as for the RRL reactions. Reactions were incubated at 30°C for 5 minutes, then 250 μL of ice-cold dilution buffer (40 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT) was added and the reactions were immediately loaded on 15-30% sucrose gradients in dilution buffer. Gradients were centrifuged at 35,000 RPM for 4 hours in an SW41 rotor, then fractionated using a BioComp system. The amount of radiation in each fraction was measured and used to generate the plots. According to the manufacturer, this lysate contains substantial RNase activity; we attempted to mitigate this effect using RNase inhibitors. However, we were unable to fully eliminate the activity.

### Crystallographic data collection and structure determination

70S ribosomes were purified and the 70S•PSIV IRES complex was prepared and crystallized essentially as previously described[19]. The IRES RNA used contained nucleotides 6000-6195 of the PSIV viral RNA. X-ray diffraction data were collected at beamline 23 ID-B at the Advanced Photon Source at Argonne National Laboratory, using an X-ray wavelength of 1.033 Å and an oscillation angle of 0.2°. For determining the structure of the 70S•PSIV IRES complex, one data set obtained from a single crystal was integrated and scaled using XDS[31]. Two percent of the reflections were marked as test-set ($R_{free}$ set)

reflections and used for cross-validation throughout refinement. The previously determined X-ray structure of the 70S ribosome bound with domain 3 of the PSIV IRES, obtained from the same crystal form, was used as a molecular replacement model[19]. Domain 3 of the IRES and L1 stalk were removed from this starting model. Initial $F_{obs}$-$F_{calc}$ difference maps were calculated after rigid-body and simulated-annealing refinement performed using two-fold non-crystallographic symmetry (NCS) restraints for the ribosome as previously described[27]. The difference maps revealed the positions of the L1 stalk and domain 3 of the PSIV IRES, allowing us to position the models for these parts of the structure. The density corresponding to domain 1+2 of the IRES revealed the approximate positioning for this domain but was not sufficient to allow unambiguous building of the structural model. NCS-restrained structure refinement was carried out using PHENIX[32], as described[27]. Coot[33] was used for structure visualization and calculation of NCS-averaged maps. Figures were rendered using PyMOL[34]. Information on data collection and refinement statistics is summarized in Extended Data Table 1b.

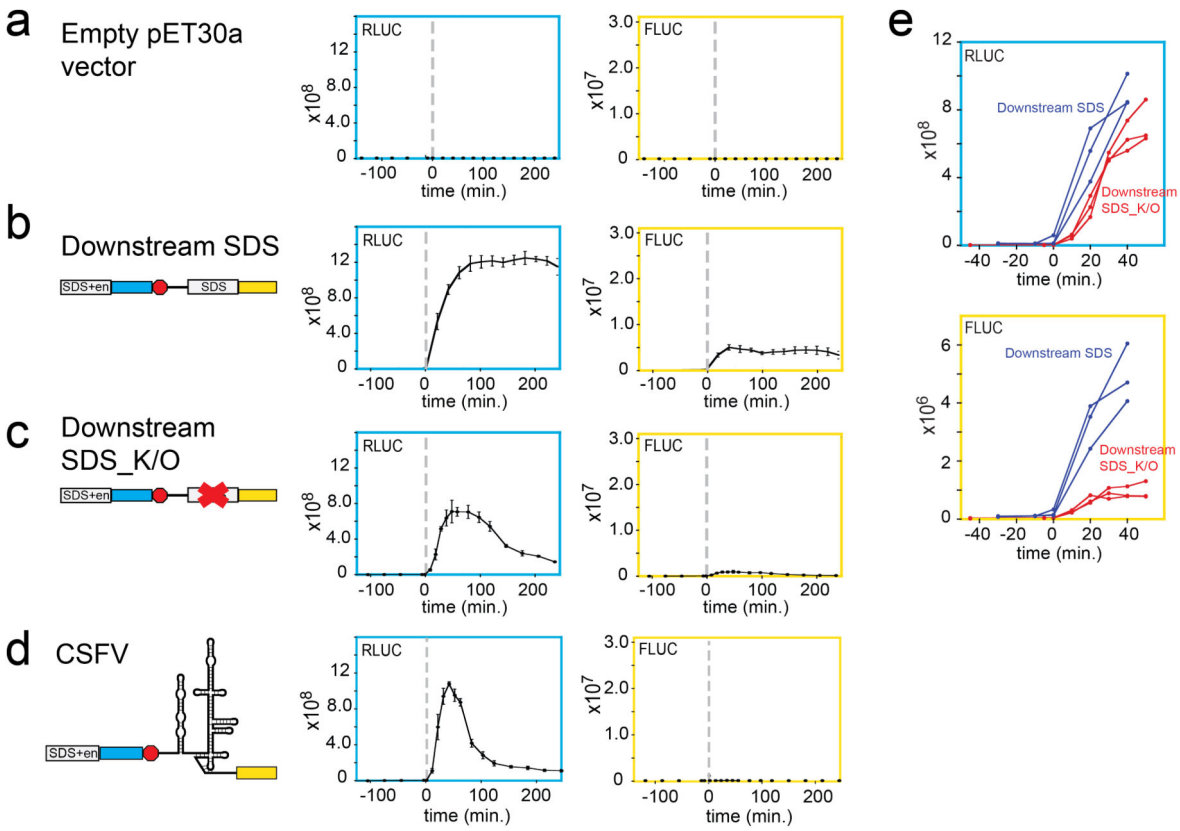
## Extended Data




**Extended Data Figure 1. Canonical translation initiation signals, characteristics of IGR IRESs, and experimental design**

**a**, Bacterial mRNAs (left) use a Shine-Dalgarno Sequence (SDS, red) upstream of the AUG start codon and open reading frame (green) to recruit the 30S subunit (grey). The interaction is through the anti-Shine-Dalgarno Sequence (A-SDS, yellow). Three initiation factors (magenta) are also important. Eukaryotic mRNAs (right) have a 5' 7-methyl-guanosine "cap" (red, 7mG) that is bound by initiation factor 4E (4E, yellow). Multiple initiation factors (blue and magenta) serve to recruit the 40S subunit (grey) and allow it to scan to the

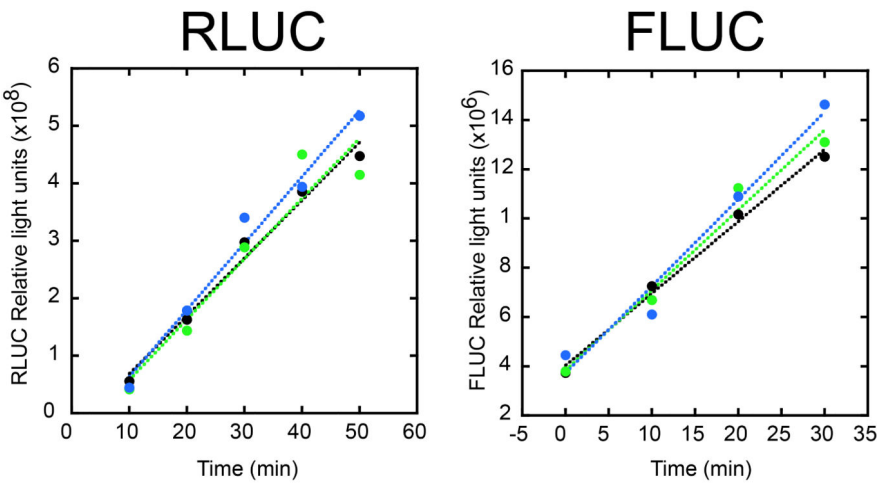
start codon. **b**, Left: cryo-electron microscopy (cryo-EM) reconstruction of an IGR IRES RNA (magenta) bound to a human 40S subunit (yellow)[8]. The compact structure occupies the tRNA-binding groove of the subunit. Right: cryo-EM reconstruction of an IGR IRES RNA (magenta) bound to a human 80S ribosome[8]. The 40S subunit is yellow and the 60S subunit is cyan. The IRES RNA occupies the conserved intersubunit space. **c**, Cartoon representation of the secondary structure of a type 1 IGR IRES RNA (the type to which PSIV belongs). This structure is found between two open reading frames within the viral RNA genome. The two independently folded domains (domain 1+2 and domain 3) are indicated with dashed grey ovals. The locations of two pseudoknot interactions critical for inducing the correct IRES folded structure, and thus for function (PK 1 and PK 2), are shown. **d**, The structured IRES studied here is found in the intergenic region of the viral genome (red). It was placed into a dual-luciferase (LUC) reporter construct (blue, *Renilla* LUC, RLUC; yellow, Firefly LUC, FLUC) and this was cloned into bacterial expression vector pET30a. This vector was used to transform *E. coli.* Induction of the culture leads to expression of the dual-LUC mRNA. Aliquots of the culture were harvested at defined time points and the amount of each LUC was measured. These data were used to determine the initial rate of LUC production (generally linear over the first 30-40 min post-induction) for each of the two reporters. The RLUC served as a consistent internal control for different bacterial cultures, clones, growth rates, etc.




**Extended Data Figure 2. Verification of independent quantifiable LUC production in bacteria**

**a**, An empty pET30a vector (no inserted LUC genes) shows negligible signal. **b**, Traces of LUC activity as a function of time are shown from a construct in which the RLUC gene was driven by the SDS and enhancer sequence from pET30a and FLUC was driven by an SDS only (Downstream SDS). The red octagon denotes stop codons. Both LUCs are generated, and RLUC production is higher, as expected. **c**, Removal of the SDS driving FLUC production (Downstream SDS_K/O) results in a loss of FLUC production, as expected. **d**, Insertion of the IRES from classical swine fever virus (CSFV) in position to drive initiation of FLUC results in negligible FLUC activity. In panels a-d, the y axis indicates relative light units (RLUs). Error bars represent one standard deviation from the mean from three biological replicates. Here and throughout this study, we observed different LUC versus time profiles with different constructs. For example, the RLUC traces for the Downstream SDS and Downstream SDS_K/O constructs are different, despite no change to the SDS driving RLUC production (one shows a decrease of RLUC in later time points, the other maintains RLUC levels). The reason for this effect is unknown, but it only appears ~60 minutes after induction. **e**, Despite differences in longer time-courses, LUC production was consistent and linear over the first 30-40 minutes post-induction. The RLUC and FLUC traces from the Downstream SDS and Downstream SDS_K/O constructs are shown. The consistency of these initial rates, before high levels of mRNA and reporter might build up and affect bacterial behavior, justified their use as a means to quantitate LUC production (Extended Data Fig. 3).

a

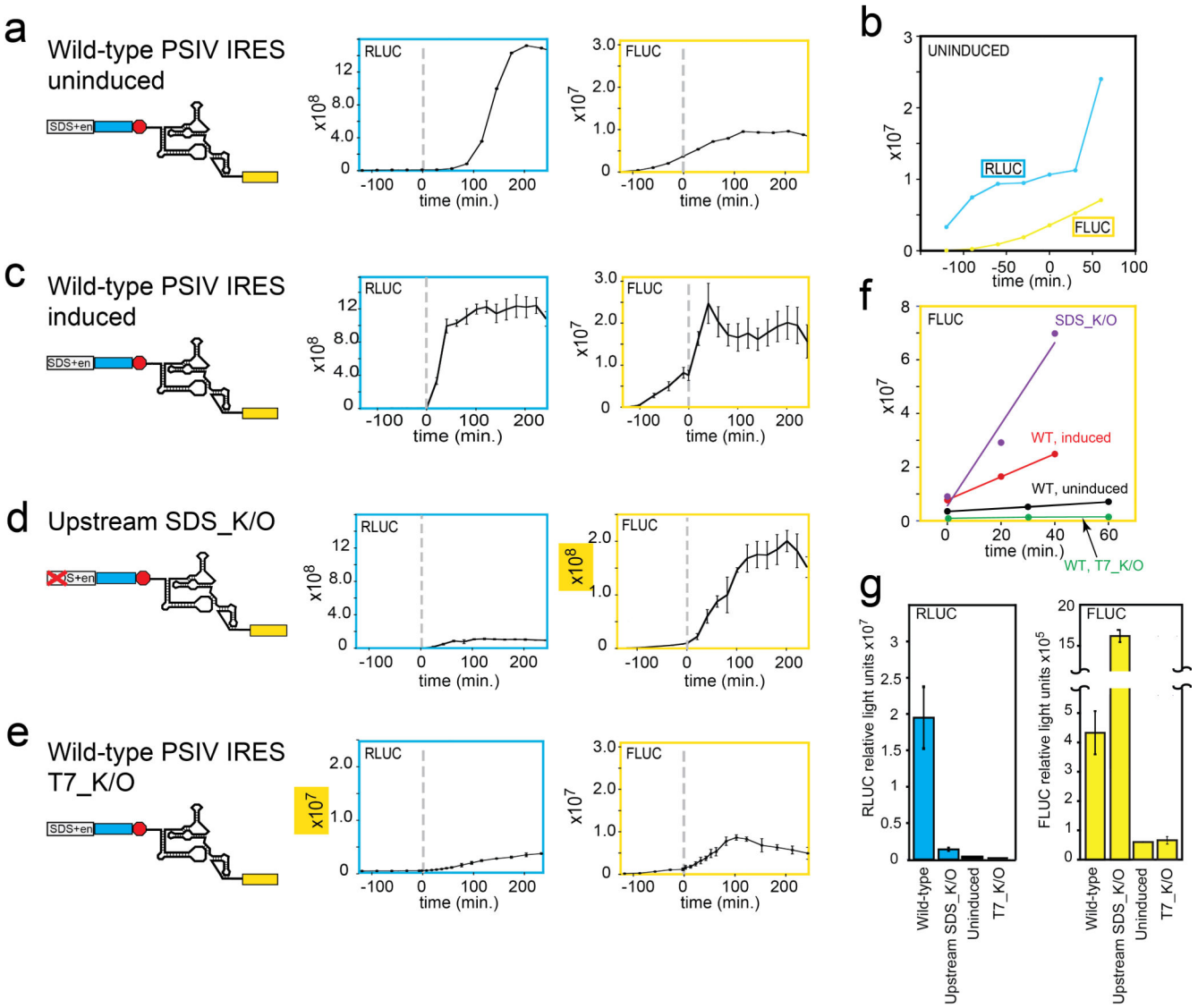



b

| Run | RLUC initial rate | FLUC initial rate | Ratio (FLUC/RLUC) |
|---|---|---|---|
| 1 | 9.14E+06 | 1.87E+05 | 2.05E-02 |
| 2 | 7.20E+06 | 1.30E+05 | 1.81E-02 |
| 3 | 8.49E+06 | 2.80E+05 | 3.30E-02 |
| 4 | 1.13E+07 | 2.92E+05 | 2.60E-02 |
| 5 | 1.37E+07 | 3.25E+05 | 2.37E-02 |
| 6 | 1.21E+07 | 3.53E+05 | 2.92E-02 |
| | | | |
| AVERAGE: | 1.03E+07 | 2.61E+05 | 2.51E-02 |
| STDEV: | 2.45E+06 | 8.52E+04 | 5.51E-03 |

**Extended Data Figure 3. Determination of IRES activity from initial rates of LUC production**

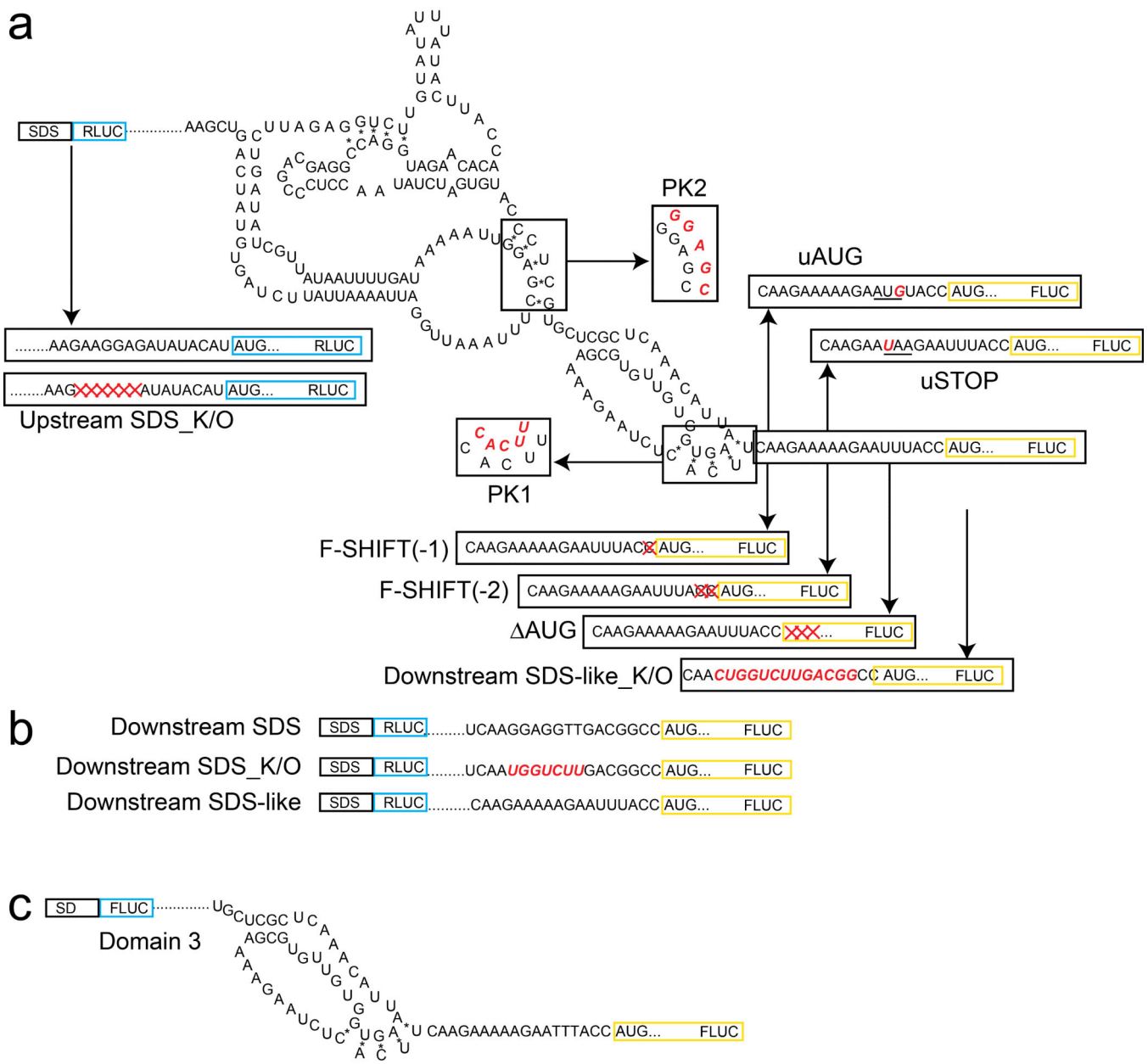
**a**, Representative graphs of RLUC and FLUC levels at early time points from three cultures of bacteria transformed with an IRES-containing bicistronic vector, induced with IPTG at time = 0. Data from the three cultures are shown as black, green, and blue points, and a linear fit is shown with a dashed line for each. The slopes of these fit lines were used as the initial rate of LUC production per minute. **b**, Representative table of data for one IRES construct. Data from six cultures is shown, with initial rates for RLUC and FLUC production in (Relative Light Units (RLUs)/minute). Throughout this manuscript, the average rate for each LUC is shown in blue (RLUC) and yellow (FLUC) bar graphs. The ratio of these rates was determined from each culture, and these were averaged and shown in green bar graphs.




**Extended Data Figure 4. Examination of leaky expression and cryptic promoter activity**

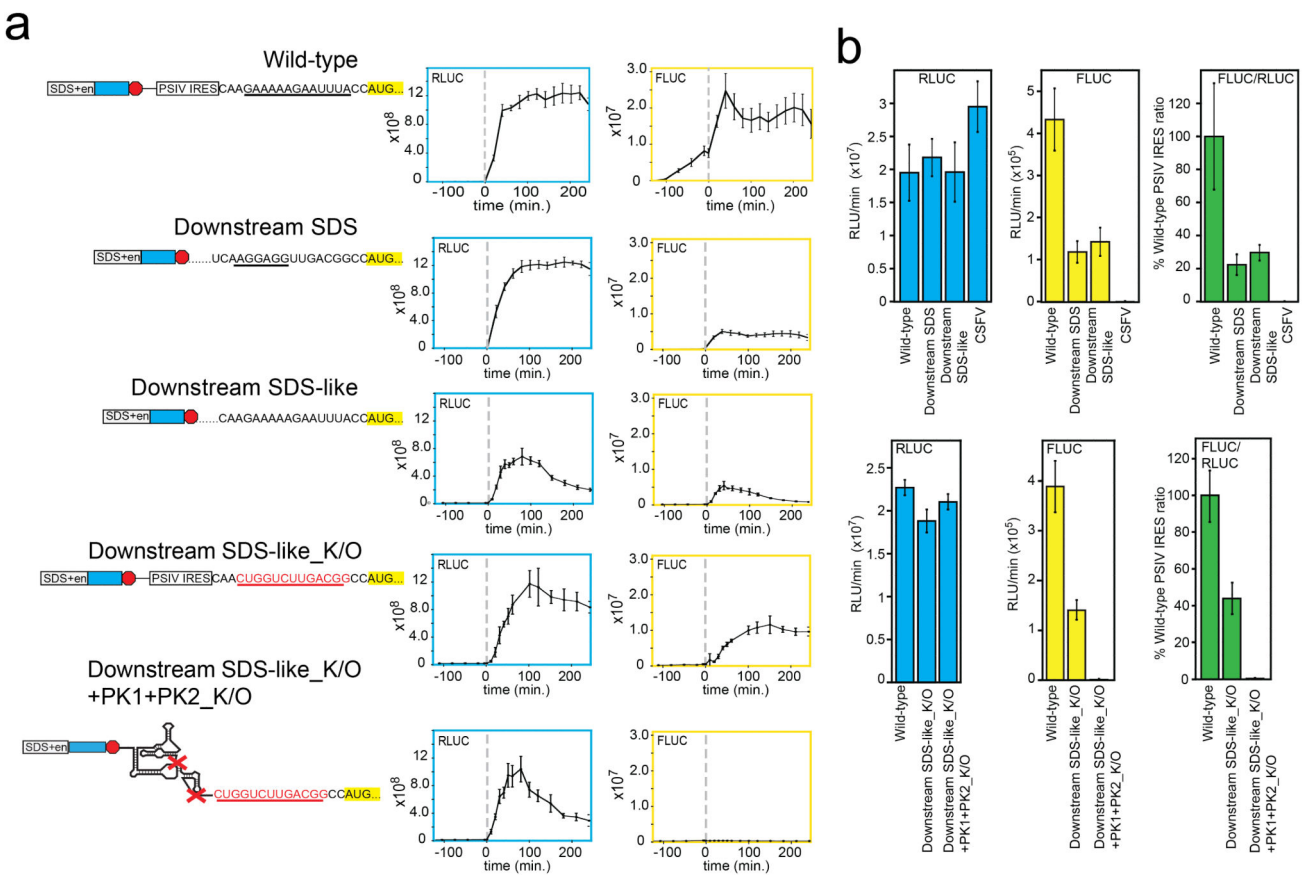
**a**, Traces of LUC production from the WT PSIV IRES-containing construct without induction with IPTG. Both RLUC and FLUC are produced due to "leaky expression" of mRNA, a common observation with pET30a bacterial expression vectors. In panels a-f, the y axis shows relative light units (RLUs). **b**, Examination of the early time-points of the traces from panel 1 show that both RLUC and FLUC are expressed to a low level without induction, and thus this leaky expression is not due to the IRES. **c**, Traces of WT PSIV IRES with IPTG induction at time = 0 (grey dashed line), showing the increase due to induction. **d**, Traces of a construct with the RLUC-driving SDS knocked out (Upstream SDS_K/O, same as in Fig. 1b), shown for comparison. **e**, To check for cryptic promoter activity due to transcription from a site other than the authentic T7 promoter, we cloned the full IRES-containing dual-LUC cassette into a pET30a vector in which the T7 promoter was mutated from 5'-TAATACGACTCACTATA-3' to 5'-TAATGGTGTCTGAATTC-3' (T7_K/O). Both RLUC and FLUC are produced to low levels, indicating some T7 promoter-

independent expression exists in this vector, but the initial rates of producing upon induction are very low (see panels f&g). **f**, Initial rates of production of FLUC from the T7_K/O (induced), WT (uninduced), WT (induced), and Upstream SDS_K/O (induced) constructs. Rates of FLUC production from the T7_K/O and uninduced WT are very low and not sufficient to account for apparent initiation from the IRES upon induction. This graph also illustrates the importance and utility of using the initial rates of LUC production for analysis, rather than the entire curve or an arbitrary later time point. **g**, Quantitated and graphed initial rate data for the four constructs in this figure. Error bars represent one standard deviation of the mean from three biological replicates, except the uninduced control, which was done once.



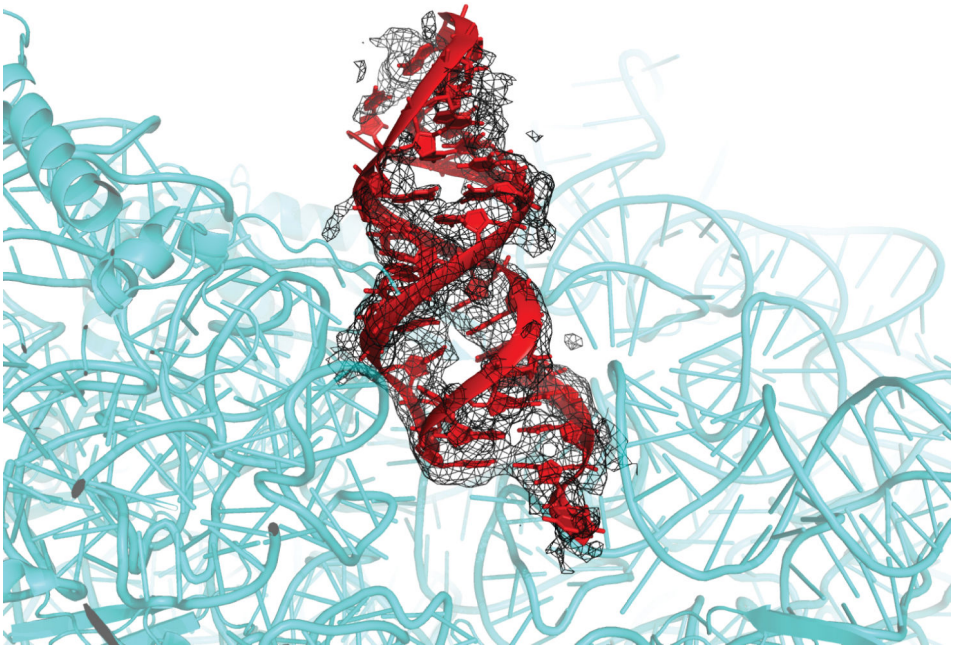

**Extended Data Figure 5. PSIV IGR IRES sequence, secondary structure, and design of mutants**

**a**, Secondary structure of the full-length IGR IRES from the *Plautia stali* intestine virus (PSIV). The specific changes that were introduced to generate the mutants and constructs described and tested in the main text are shown. For each, the altered region is boxed and the change is shown in red. For the uAUG and uSTOP constructs, the start and stop codons are underlined. RLUC and FLUC coding sequences are boxed cyan and yellow, respectively. **b**, Constructs without the IRES that contain various wild-type or mutant SDS and SDS-like sequences upstream of the FLUC ORF. **c**, Construct containing just domain 3 of the PSIV IGR IRES.



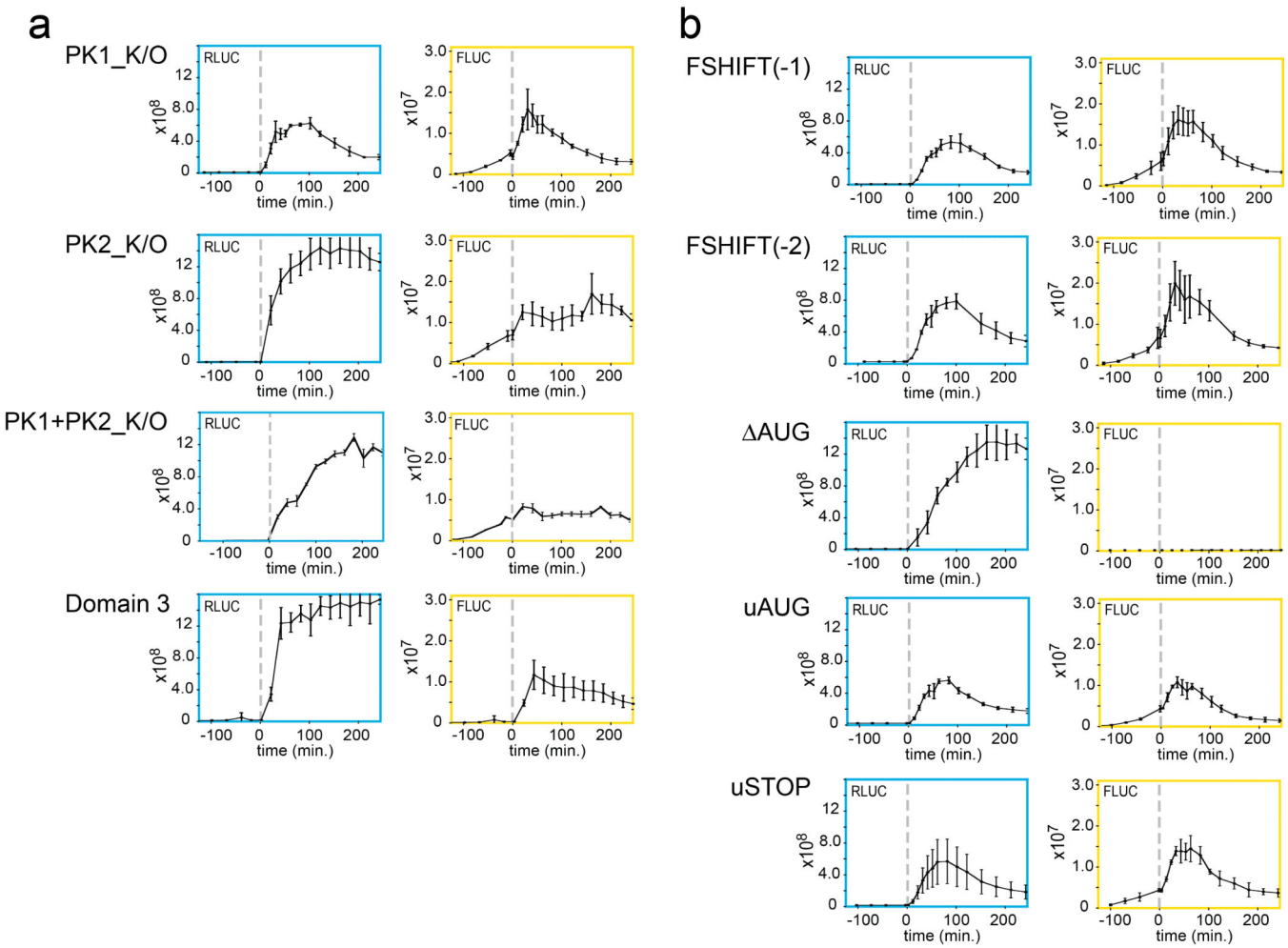

**Extended Data Figure 6. Contributions of region upstream of AUG to initiation activity**

**a**, Diagram of constructs tested and traces of FLUC and RLUC production. The y axis shows relative light units (RLUs). **b**, Quantitated initial rates from these constructs. Results from CSFV IRES (negative control) shown for comparison. “Downstream SDS” contains an SDS driving FLUC production (in place of the IRES), “Downstream SDS-like” contains the purine-rich sequence in place of the IRES and driving FLUC production. In “Downstream SDS-like_K/O”, this purine-rich sequence has been replaced by a pyrimidine-rich sequence. A PSIV IRES construct in which both pseudoknots are disrupted and the purine-rich SDS-like sequence just downstream of the IRES is mutated has essentially the same activity as the CSFV IRES (Downstream SDS-like_K/O+PK1+PK2_K/O). Error bars are one standard deviation from the mean of three biological replicates.



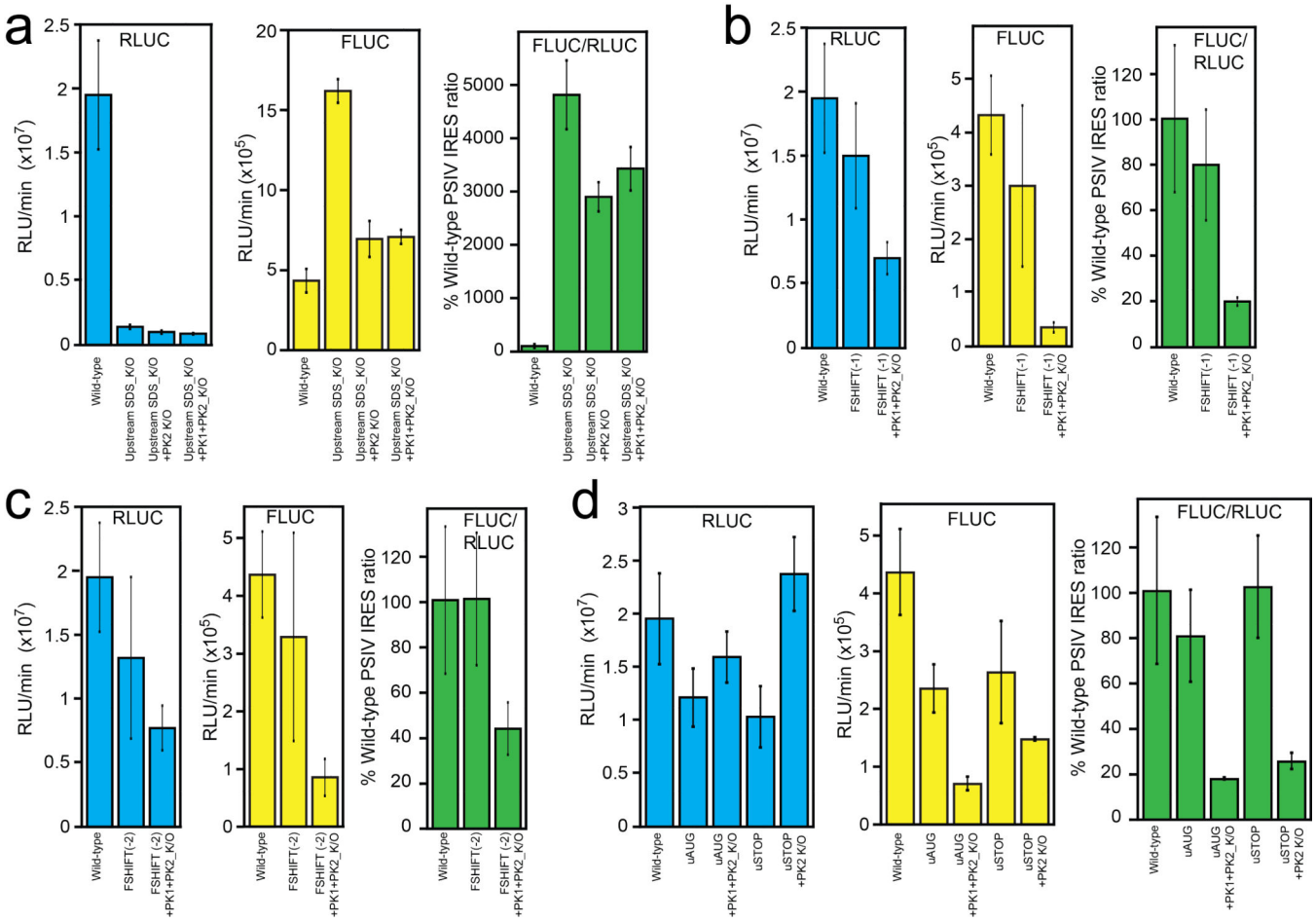
**Extended Data Fig. 7. Domain 3's position in the full-length PSIV IGR IRES-70S structure**

Crystal structure of a full-length PSIV IGR IRES bound to *T. thermophilus* 70S ribosomes. Cyan: small subunit; red: PSIV IRES domain 3; black: unbiased Fourier difference $F_{obs}$-$F_{calc}$ map for domain 3 in the P site of the small subunit. The large subunit and domains 1+2 are not shown.




**Extended Data Fig. 8. Luciferase activity time-courses for various constructs**

**a**, Time-course traces for constructs and bar graphs shown in Fig. 3. **b**, Time-course traces for constructs and bar graphs shown in Fig. 4. Error bars are one standard deviation from the mean of three biological replicates. In both panels, the y axis shows relative light units (RLUs).




**Extended Data Fig. 9. Quantitated data for various constructs in the context of the PK1+PK2_K/O mutation**

**a**, Combination of knocking out the RLUC SDS (Upstream SDS_K/O) with the PK2_K/O or PK1+PK2_K/O. Initial rates of RLUC are greatly diminished. Rates of FLUC are lower, but less diminished than RLUC. This is most likely attributable to the decreased competition for ribosomes and the presence of the SDS-like sequence upstream of the FLUC ORF and not to robust initiation on the IRES. **b**, The PK1+PK2_K/O dramatically reduced initial rate of FLUC production on the IRES with the FSHIFT(−1) mutation. **c**, The PK1+PK2_K/O dramatically reduced initial rate of FLUC production on the IRES with the FSHIFT(−2) mutation. **d**, The PK1+PK2_K/O dramatically reduced initial rate of FLUC production on the IRES with the uSTOP and uAUG mutations. Error bars are one standard deviation from the mean from three biological replicates.

## Extended Data Table 1a

Initial rates of RLUC and FLUC for all constructs tested and crystallographic data collection, phasing and refinement statistics. Raw values are shown for all contructs tested. All values are the mean of three independent experiments ± one standard deviation from the mean, except for the uninduced control that was done once.

| Construct or Condition Tested | FLUC (RLU/min; × $10^5$) | RLUC (RLU/min; × $10^7$) |
|---|---|---|
| WT, uninduced | 0.59 | 0.022 |
| WT | 4.33 ± 0.73 | 1.95 ± 0.42 |
| WT, 17_K/O | 0.65 ± 0.12 | 0.001 ± 0.0002 |
| Upstream SDS_K/O | 16.2 ± 0.73 | 0.14 ± 0.016 |
| CSFV | 0.0011 ± .0001 | 2.95 ± 0.38 |
| Downstream SDS | 1.18 ± 0.26 | 2.18 ± 0.28 |
| PK1_K/O | 2.84 ± 0.13 | 2.10 ± 0.63 |
| PK2 K/O | 2.67 ± 0.33 | 2.57 ± 0.40 |
| Domain 3 | 2.82 ± 0.87 | 3.05 ± 0.21 |
| PK1+PK2 K/O | 0.66 ± 0.51 | 1.16 ± 0.21 |
| ΔAUG | 0.002 ± 0.0003 | 1.12 ± 0.17 |
| uAUG | 2.33 ± 0.41 | 1.21 ± 0.27 |
| uSTOP | 2.61 ± 0.88 | 1.03 ± 0.29 |
| FSHIFT(−1) | 3.00 ± 1.51 | 1.50 ± 0.41 |
| FSHIFT(−2) | 3.26 ± 1.79 | 1.32 ± 0.63 |
| Downstream SDS-like | 1.42 ± 0.34 | 1.96 ± 0.45 |
| Downstream SDS-like_K/O | 1.40 ± 0.19 | 1.88 ± 0.14 |
| Downstream SDS-like_K/O +PK1+PK2 K/O | 0.0013 ± 0.0012 | 2.10 ± 0.17 |
| SDS_K/O+PK2_K/O | 6.95 ± 1.13 | 0.097 ± 0.01 |
| SDS_K/O+PK1 +PK2_K/O | 7.08 ± 0.44 | 0.084 ± 0.005 |
| FSHIFT(−1)+PK1+PK2_K/O | 0.35 ± 0.092 | 0.70 ± 0.12 |
| FSHIFT(−2)+PK1+PK2_K/O | 0.84 ± 0.32 | 0.72 ± 0.17 |
| uAUG+PK1+PK2_K/O | 0.69 ± 0.11 | 1.59 ± 0.24 |
| uST0P+PK1+PK2_K/O | 1.46 ± 0.03 | 2.37 ± 0.35 |

## Extended Data Table 1b

Crystallographic statistics.

| | |
|---|---|
| **Data collection** | |
| Space group | $P2_12_12_1$ |
| Cell dimensions | |

| | |
|---|---|
| *a, b, c* (Å) | 209.05, 447.22, 608.96 |
| α, β, γ (°) | 90, 90, 90 |
| Resolution (Å) | 60-3.8 (3.8 – 4.0)* |
| $R_{meas}$ # | 0.2 (1.6) |
| *CC(1/2)*## | 99.5 (41.6) |
| *I/ σI* | 8.28 (1.2) |
| Completeness (%) | 99.9 (99.8) |
| Redundancy | 4.8 (3.4) |
| **Refinement** | |
| Resolution (Å) | 60 – 3.8 |
| No. reflections | 555,726 |
| $R_{work}$ / $R_{free}$ | 0.246/0.284 |
| No. atoms | 287428 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.004 |
| Bond angles (°) | 0.702 |

*Values in parentheses are for highest-resolution shell.

#Rmeas is Rmeas as reported by XDS31.

##CC(1/2) is the percentage of correlation between intensities from random half-datasets as defined by Karplus and Diederichs[35].

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the members of the Kieft Lab for insight and discussion and the staffs at the Advanced Photon Source for their support. The original PSIV IGR IRES-containing plasmid was from Nobuhiko Nakashima and the source of the luciferase genes was a plasmid from Anne Willis. This work was supported by Grants GM-17129 and GM-59140 from the National Institutes of Health (NIH) and MCB-723300 from the National Science Foundation (to H.F.N.), Grant GM-103105 from the NIH (to A.A.K.), and Grants GM-97333 and GM-81346 from the NIH (to J.S.K.). J.S.K. is an Early Career Scientist of the Howard Hughes Medical Institute. T-D.M.P. was an American Heart Association Predoctoral Scholar (10PRE260143).

## References

1. Hershey JW, Sonenberg N, Mathews MB. Principles of translational control: an overview. Cold Spring Harbor perspectives in biology. 2012; 4
2. Moll I, Grill S, Gualerzi CO, Blasi U. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. Molecular microbiology. 2002; 43:239–246. [PubMed: 11849551]
3. Malys N, McCarthy JE. Translation initiation: variations in the mechanism can be anticipated. Cellular and molecular life sciences : CMLS. 2011; 68:991–1003. [PubMed: 21076851]
4. Melnikov S, et al. One core, two shells: bacterial and eukaryotic ribosomes. Nature structural & molecular biology. 2012; 19:560–567.
5. Berthelot F, Bogdanovsky D, Schapira G, Gros F. Interchangeability of factors and tRNA's in bacterial and eukaryotic translation initiation systems. Mol Cell Biochem. 1973; 1:63–72. [PubMed: 4610351]

6. Hertz MI, Thompson SR. Mechanism of translation initiation by Dicistroviridae IGR IRESs. Virology. 2011; 411:355–361. [PubMed: 21284991]
7. Schuler M, et al. Structure of the ribosome-bound cricket paralysis virus IRES RNA. Nature structural & molecular biology. 2006; 13:1092–1096.
8. Spahn CM, et al. Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. Cell. 2004; 118:465–475. [PubMed: 15315759]
9. Pestova TV, Lomakin IB, Hellen CU. Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. EMBO reports. 2004; 5:906–913. [PubMed: 15332113]
10. Jan E, Sarnow P. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. Journal of molecular biology. 2002; 324:889–902. [PubMed: 12470947]
11. Nishiyama T, et al. Structural elements in the internal ribosome entry site of Plautia stali intestine virus responsible for binding with ribosomes. Nucleic acids research. 2003; 31:2434–2442. [PubMed: 12711689]
12. Jan E, et al. Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. Cold Spring Harb Symp Quant Biol. 2001; 66:285–292. [PubMed: 12762030]
13. Thompson SR, Gulyas KD, Sarnow P. Internal initiation in Saccharomyces cerevisiae mediated by an initiator tRNA/eIF2-independent internal ribosome entry site element. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:12972–12977. [PubMed: 11687653]
14. Wilson JE, Pestova TV, Hellen CU, Sarnow P. Initiation of protein synthesis from the A site of the ribosome. Cell. 2000; 102:511–520. [PubMed: 10966112]
15. Sasaki J, Nakashima N. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:1512–1515. [PubMed: 10660678]
16. Sasaki J, Nakashima N. Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus in vitro. Journal of virology. 1999; 73:1219–1226. [PubMed: 9882324]
17. Costantino DA, Pfingsten JS, Rambo RP, Kieft JS. tRNA-mRNA mimicry drives translation initiation from a viral IRES. Nature structural & molecular biology. 2008; 15:57–64.
18. Jan E, Kinzy TG, Sarnow P. Divergent tRNA-like element supports initiation, elongation, and termination of protein biosynthesis. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100:15410–15415. [PubMed: 14673072]
19. Zhu J, et al. Crystal structures of complexes containing domains from two viral internal ribosome entry site (IRES) RNAs bound to the 70S ribosome. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:1839–1844. [PubMed: 21245352]
20. Hellen CU. IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. Biochimica et biophysica acta. 2009; 1789:558–570. [PubMed: 19539793]
21. Yamamoto H, Nakashima N, Ikeda Y, Uchiumi T. Binding mode of the first aminoacyl-tRNA in translation initiation mediated by Plautia stali intestine virus internal ribosome entry site. The Journal of biological chemistry. 2007; 282:7770–7776. [PubMed: 17209036]
22. Fernandez IS, Bai XC, Murshudov G, Scheres SH, Ramakrishnan V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. Cell. 2014; 157:823–831. [PubMed: 24792965]
23. Deniz N, Lenarcic EM, Landry DM, Thompson SR. Translation initiation factors are not required for Dicistroviridae IRES function in vivo. RNA. 2009; 15:932–946. [PubMed: 19299549]
24. Stoneley M, Paulin FE, Le Quesne JP, Chappell SA, Willis AE. C-Myc 5' untranslated region contains an internal ribosome entry segment. Oncogene. 1998; 16:423–428. [PubMed: 9467968]
25. Costantino DA, Kieft JS. A preformed compact ribosome-binding domain in the cricket paralysis-like virus IRES RNAs. RNA. 2005; 11:332–343. [PubMed: 15701733]

26. Pfingsten JS, Costantino DA, Kieft JS. Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. Science. 2006; 314:1450–1454. [PubMed: 17124290]
27. Koh CS, Brilot AF, Grigorieff N, Korostelev AA. Taura syndrome virus IRES initiates translation by binding its tRNA-mRNA-like structural element in the ribosomal decoding center. Proceedings of the National Academy of Sciences of the United States of America. 111:9139–9144. [PubMed: 24927574]
28. Nishiyama T, Yamamoto H, Uchiumi T, Nakashima N. Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. Nucleic acids research. 2007; 35:1514–1521. [PubMed: 17287295]
29. Landry DM, Hertz MI, Thompson SR. RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. Genes & development. 2009; 23:2753–2764. [PubMed: 19952110]
30. Ben-Shem A, et al. The structure of the eukaryotic ribosome at 3.0 A resolution. Science. 2011; 334:1524–1529. [PubMed: 22096102]
31. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. 1993; 26:795–800.
32. Adams PD, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002; 58:1948–1954. [PubMed: 12393927]
33. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
34. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific: 2002.
35. Karplus PA, Diederichs K. Linking crystallographic model and data quality. Science. 2012; 336:1030–1033. [PubMed: 22628654]

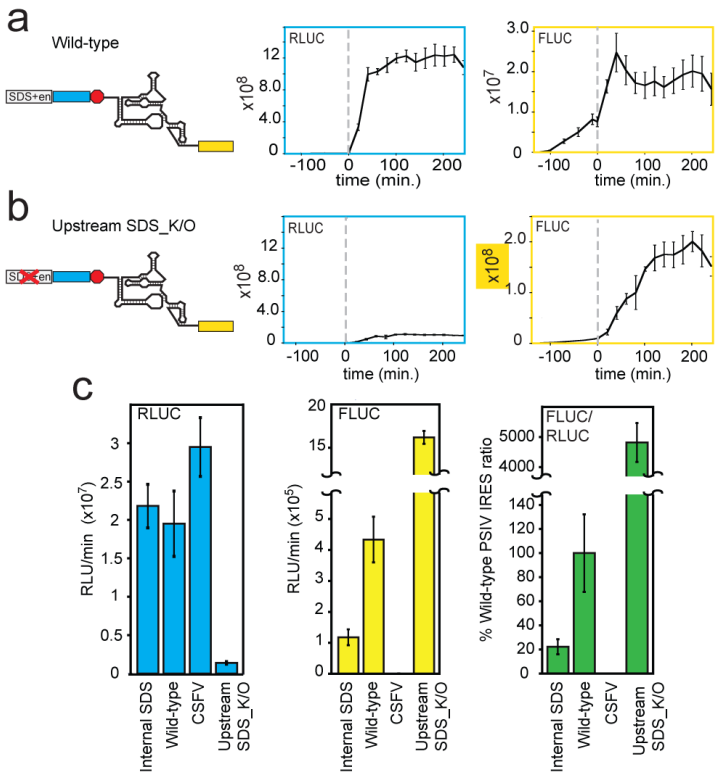



**Figure 1. Translation initiation assays in bacteria**

**a**, Full-length WT IRES. Left: diagram of the construct. The left (cyan outline) graph shows relative light units (RLU) from the upstream RLUC as a function of time. Dashed grey line is t=0, the point of induction. The trace is the average signal of at least three experiments, with error bars showing one standard deviation from the mean. The right (yellow outline) graph shows FLUC expression from the IRES. **b**, Diagram and traces from the Upstream SDS_K/O mutant. Note the change in scale of the y-axis for FLUC. **c.** Initial rates of RLUC and FLUC production, and the FLUC/RLUC ratio for the indicated constructs. Error bars represent one standard deviation from the mean from three biological replicates. See Extended Data Figs. 2&5 for diagrams and raw traces of the Internal SDS and CSFV constructs.

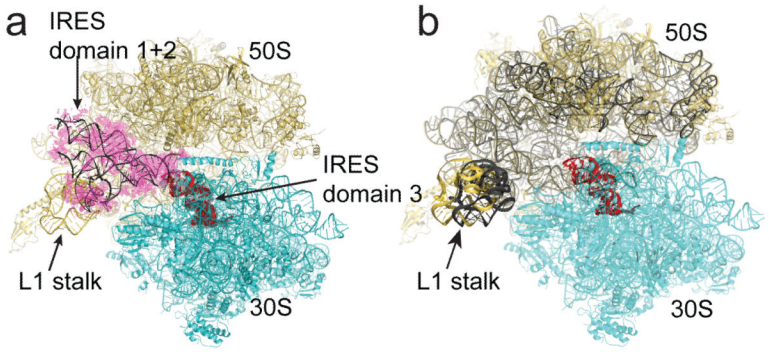



**Figure 2. IRES-70S ribosome structure**

**a**, Crystal structure of a full-length PSIV IGR IRES bound to *T. thermophilus* 70S ribosomes. Cyan: small subunit; yellow: large subunit; red: PSIV IRES domain 3; grey: density corresponding to domain 3; magenta: unbiased difference $F_{obs}$-$F_{calc}$ density corresponding to domain 1+2, with the crystal structure of PSIV IGR IRES domain 1+2 (black) docked as a rigid body[26]. **b**, Superimposition of crystal structures of the PSIV IGR IRES•70S ribosome complex (this work) and the 70S ribosome; yellow: IRES-bound 50S subunit. Domain 1+2 shifts the L1 stalk relative to its position in tRNA-bound complexes by ~15Å.

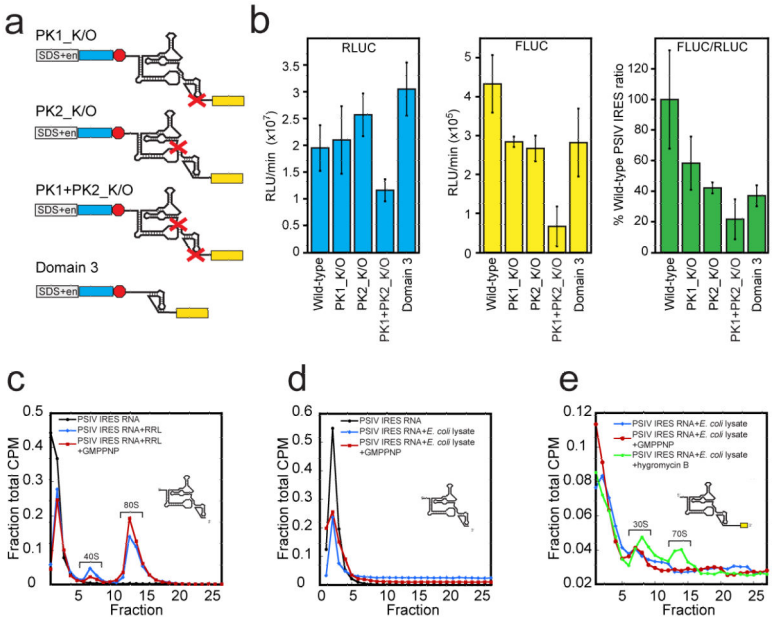



**Figure 3. Importance of IRES structure and ribosome binding**

**a**, IRES constructs with structural domains disrupted or removed. **b**, Rates of LUC production and LUC ratio. Error bars represent one standard deviation from the mean from three biological replicates. **c**, Ribosome assembly assay with the PSIV IGR IRES in RRL, resolved on a sucrose gradient. Locations of complexes are indicated. CPM, counts per minute. **d**, Same as panel c, but in *E. coli* lysate. **f**, Same as panel d, but with an IRES RNA containing downstream sequence to include the FLUC start codon. In all panels, the addition of GMPPNP or hygromycin B is indicated. Data from one experiment is shown.

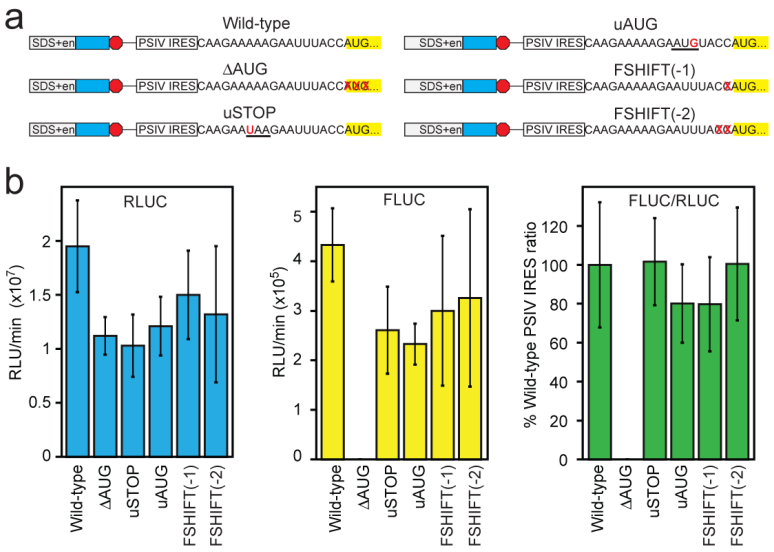



**Figure 4. Location of initiation on an IGR IRES in bacteria**

**a**, Constructs designed to determine the location of initiation. For uAUG and uSTOP, the start and stop codons are underlined. **b**, Rates of LUC production and LUC ratio from these constructs. Error bars represent one standard deviation from the mean from three biological replicates.