LETTER

doi:10.1038/nature14219

Initiation of translation in bacteria by a structured eukaryotic IRES RNA

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

The central dogma of gene expression (DNA to RNA to 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]. However, the core structures and conformational dynamics of ribosomes that are responsible for the translation steps that take place after initiation are ancient and conserved across the domains of life[2]. We wanted to explore 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 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 transfer RNAs. 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 and provides an example of an RNA structure-based translation initiation signal capable of operating in two domains of life.

Bacteria cannot recognize the 'cap' on the 5′ end of eukaryotic messenger RNAs and eukaryotic ribosomes cannot use the Shine–Dalgarno sequence (SDS)[1] (Extended Data Fig. 1a). Although non-canonical mechanisms exist[3,4], 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[2]. 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 (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 the simultaneous measurement of initial rates of production of each protein (Extended Data Figs 2 and 3). We used this construct to test whether an IGR IRES RNA can drive translation in live bacteria. *Renilla* luciferase (RLUC) was placed to initiate translation from an SDS (and 'enhancer' sequence), and firefly luciferase (FLUC) was placed after a wild-type *Plautia stali* intestine virus (PSIV) IGR IRES. There was some production of both LUCs before induction (due to expected 'leaky expression'; Extended Data Fig. 4), but induction resulted in a marked increase in both reporters; the production

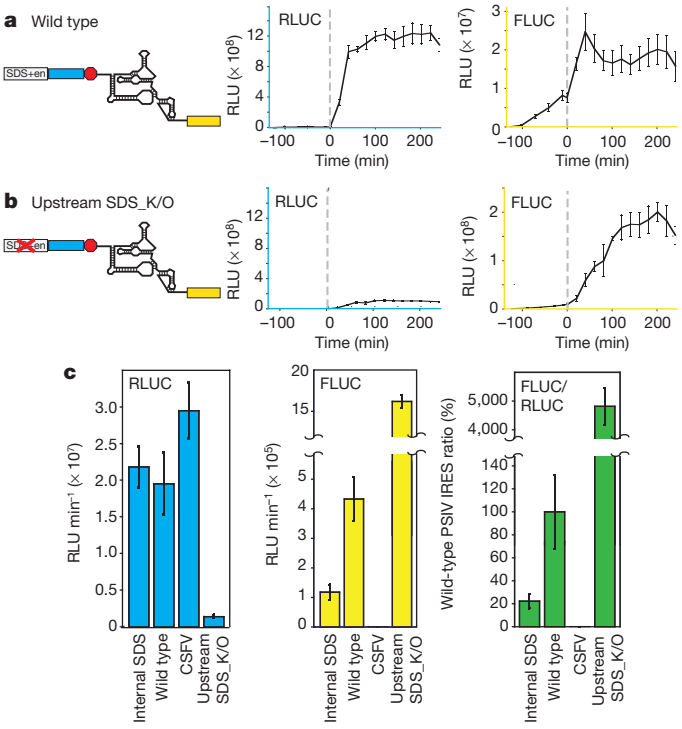


Figure 1 | Translation initiation assays in bacteria. **a**, Full-length wild-type IRES. Left, diagram of the construct. en, enhancer. Middle, graph shows relative light units (RLU) from the upstream RLUC as a function of time. Dashed grey line is $t = 0$, the point of induction. The trace is the average signal of at least three experiments, with error bars showing 1 standard deviation (s.d.) from the mean. Right, 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 1 s.d. from the mean from three biological replicates. See Extended Data Figs 2 and 5 for diagrams and raw traces of the Internal SDS and CSFV constructs.

[1]Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, Colorado 80045, USA. [2]Howard Hughes Medical Institute, University of Colorado Denver School of Medicine, Aurora, Colorado 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, California 95064, USA. †Present addresses: Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, USA (T.M.C.); Cocrystal Discovery, Inc., Mountain View, California 94043, USA (J.Z.); RNA Therapeutics Institute, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA (A.A.K.); Department of Reproductive Medicine, University of California at San Diego, La Jolla California 92093, USA (T.-D.M.P.).
*These authors contributed equally to this work.

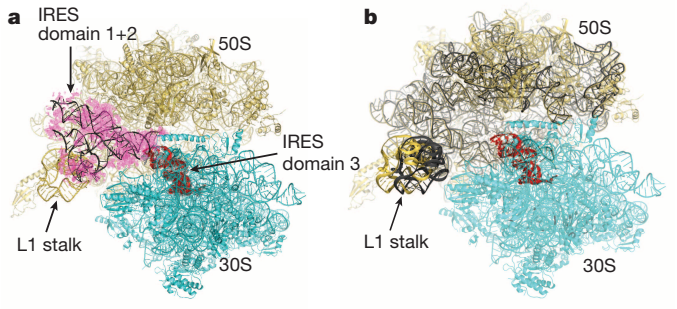


Figure 2 | IRES•70S ribosome structure. **a**, Crystal structure of a full-length PSIV IGR IRES bound to *Thermus thermophilus* 70S ribosomes. Cyan, small subunit; yellow, large subunit; red, PSIV IRES domain 3; grey, density corresponding to domain 3; magenta, unbiased difference $F_o - F_c$ density 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 ~15 Å.

of FLUC is consistent with translation beginning at the IRES (Fig. 1c and 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 more than tenfold (Fig. 1b; all raw LUC data in Extended Data Table 1a), which we attributed to decreased competition for ribosomes and to 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 wild-type 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 probably 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 domains 1 and 2 (domain 1+2) contact both subunits, whereas 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 and Extended Data Fig. 7); this may represent an initiation-state or translocated IRES. The density of domain 1+2 was weak but its location could be modelled using the crystal structure of unbound PSIV IGR IRES domain 1+2 (Fig. 2a)[26]. The location of domain 1+2 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 the 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 ~15 Å compared with 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 compact structure of the IGR IRES is essential for its function in eukaryotes[25,26], and the IRES•70S structure suggested that this is also true in bacteria. To test this, we disrupted two pseudoknots essential for the compact structure of the IRES, both individually (PK1_K/O, PK2_K/O) and together (PK1+PK2_K/O), and measured activity (Fig. 3a, b and Extended Data Fig. 8a)[10]. FLUC production decreased in all three mutants, with FLUC production in the double mutant at a level that could be accounted for by activity from 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-hydrolysable GTP analogue (GMPPNP)

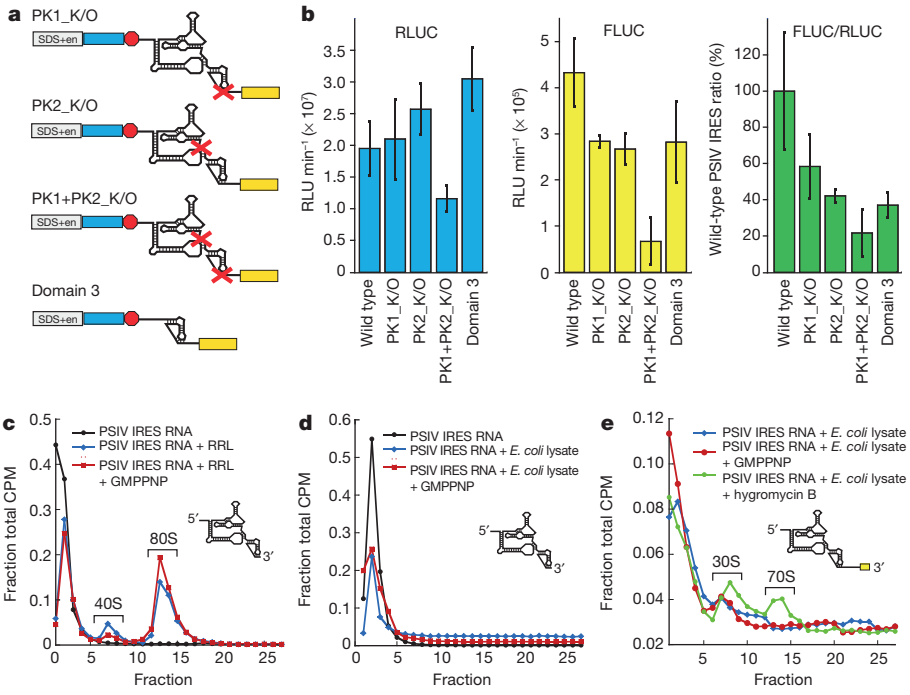


Figure 3 | Importance of IRES structure and ribosome binding. **a**, IRES constructs with structural domains disrupted or removed. en, enhancer. **b**, Rates of LUC production and LUC ratio. Error bars represent 1 s.d. from the mean from three biological replicates. **c**, Ribosome assembly assay with the PSIV IGR IRES in RRL, resolved on a sucrose gradient. Locations of complexes are indicated. CPM, counts per minute. **d**, As for **c**, but in *E. coli* lysate. **f**, As for **d**, but with an IRES RNA containing a downstream sequence to include the FLUC start codon. **c**–**e**, The addition of GMPPNP or hygromycin B is indicated. Data from one experiment are shown.

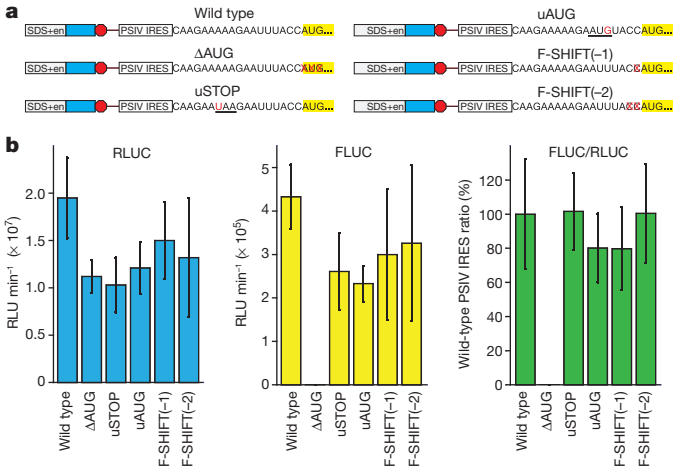


Figure 4 | Location of initiation on an IGR IRES in bacteria. **a**, Constructs designed to determine the location of initiation. For uAUG and uSTOP, the start and stop codons are underlined. en, enhancer. **b**, Rates of LUC production and LUC ratio from these constructs. Error bars represent 1 s.d. from the mean from three biological replicates.

(Fig. 3c). In contrast, 70S formation on the IRES in *Escherichia 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 the 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 whether this is true in bacteria, in which 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 and 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 that would be 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 the IRES studied here 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 structure-assisted initiation may be used in potentially all domains of life, driven by diverse RNAs perhaps possessing tRNA-like character or decoding groove binding capability.

Online Content Methods, along with any additional 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.

Received 29 January 2014; accepted 7 January 2015.
Published online 4 February 2015.

1. Hershey, J. W., Sonenberg, N. & Mathews, M. B. Principles of translational control: an overview. *Cold Spring Harb. Perspect. Biol.* **4,** a011528 (2012).
2. Melnikov, S. *et al.* One core, two shells: bacterial and eukaryotic ribosomes. *Nature Struct. Mol. Biol.* **19,** 560–567 (2012).
3. Malys, N. & McCarthy, J. E. Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* **68,** 991–1003 (2011).
4. Moll, I., Grill, S., Gualerzi, C. O. & Blasi, U. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Micr. obiol.* **43,** 239–246 (2002).
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.* **1,** 63–72 (1973).
6. Hertz, M. I. & Thompson, S. R. Mechanism of translation initiation by *Dicistroviridae* IGR IRESs. *Virology* **411,** 355–361 (2011).
7. Schüler, M. *et al.* Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nature Struct. Mol. Biol.* **13,** 1092–1096 (2006).
8. Spahn, C. M. *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* **118,** 465–475 (2004).
9. Pestova, T. V., Lomakin, I. B. & Hellen, C. U. Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. *EMBO Rep.* **5,** 906–913 (2004).
10. Jan, E. & Sarnow, P. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J. Mol. Biol.* **324,** 889–902 (2002).
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 Res.* **31,** 2434–2442 (2003).
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.* **66,** 285–292 (2001).
13. Thompson, S. R., Gulyas, K. D. & Sarnow, P. Internal initiation in *Saccharomyces cerevisiae* mediated by an initiator tRNA/eIF2-independent internal ribosome entry site element. *Proc. Natl Acad. Sci. USA* **98,** 12972–12977 (2001).
14. Wilson, J. E., Pestova, T. V., Hellen, C. U. & Sarnow, P. Initiation of protein synthesis from the A site of the ribosome. *Cell* **102,** 511–520 (2000).
15. Sasaki, J. & Nakashima, N. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc. Natl Acad. Sci. USA* **97,** 1512–1515 (2000).
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*. *J. Virol.* **73,** 1219–1226 (1999).
17. Costantino, D. A., Pfingsten, J. S., Rambo, R. P. & Kieft, J. S. tRNA–mRNA mimicry drives translation initiation from a viral IRES. *Nature Struct. Mol. Biol.* **15,** 57–64 (2008).
18. Jan, E., Kinzy, T. G. & Sarnow, P. Divergent tRNA-like element supports initiation, elongation, and termination of protein biosynthesis. *Proc. Natl Acad. Sci. USA* **100,** 15410–15415 (2003).
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. *Proc. Natl Acad. Sci. USA* **108,** 1839–1844 (2011).
20. Hellen, C. U. IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. *Biochim. Biophys. Acta* **1789,** 558–570 (2009).
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. *J. Biol. Chem.* **282,** 7770–7776 (2007).
22. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157,** 823–831 (2014).
23. Deniz, N., Lenarcic, E. M., Landry, D. M. & Thompson, S. R. Translation initiation factors are not required for *Dicistroviridae* IRES function *in vivo*. *RNA* **15,** 932–946 (2009).
24. Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A. & Willis, A. E. c-Myc 5′ untranslated region contains an internal ribosome entry segment. *Oncogene* **16,** 423–428 (1998).
25. Costantino, D. & Kieft, J. S. A preformed compact ribosome-binding domain in the cricket paralysis-like virus IRES RNAs. *RNA* **11,** 332–343 (2005).
26. Pfingsten, J. S., Costantino, D. A. & Kieft, J. S. Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. *Science* **314,** 1450–1454 (2006).
27. Koh, C. S., Brilot, A. F., Grigorieff, N. & Korostelev, A. A. Taura syndrome virus IRES initiates translation by binding its tRNA-mRNA-like structural element in the ribosomal decoding center. *Proc. Natl Acad. Sci. USA* **111,** 9139–9144 (2014).
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 Res.* **35,** 1514–1521 (2007).
29. Landry, D. M., Hertz, M. I. & Thompson, S. R. RPS25 is essential for translation initiation by the *Dicistroviridae* and hepatitis C viral IRESs. *Genes Dev.* **23,** 2753–2764 (2009).
30. Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334,** 1524–1529 (2011).

Acknowledgements We thank the members of the Kieft laboratory for insight and discussion and the staff at the Advanced Photon Source for their support. The original PSIV IGR IRES-containing plasmid was from N. Nakashima and the source of the luciferase genes was a plasmid from A. 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).

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.

Author Information Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under accession number 4XEJ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S.K. (jeffrey.kieft@ucdenver.edu).

METHODS

Plasmid construction. DNA containing the *Plautia stali* intestine virus (PSIV) IGR IRES (nucleotides 6000–6195) between the RLUC and FLUC coding sequences 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, nucleotide 6195) and the AUG start codon of the FLUC open reading frame (ORF).

Generation of mutants. Mutants were generated using several methods. First, 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.

Second, site-directed mutagenesis using the QuikChange (Agilent) mutagenesis strategy using appropriate primer pairs.

Third, 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 (bp) 5′ of the SpeI site and 12 bp 3′ of the NcoI site of the vector. Assembled constructs maintained both restriction sites.

Fourth, for the T7 knockout construct, a pET30a vector containing a mutated T7 promoter (TAAATGGTGTCTGAATTC) was synthesized (DNA 2.0) and DNA coding for the wild-type 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.

Fifth, 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 earlier 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 isopropyl-β-D-thiogalactoside (IPTG) (Gold Bio) and allowed to grow for 4 h. Samples (50 µl) were taken at 10–30 min 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 1× Passive Lysis Buffer (PLB, Promega). Twenty microlitres 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.

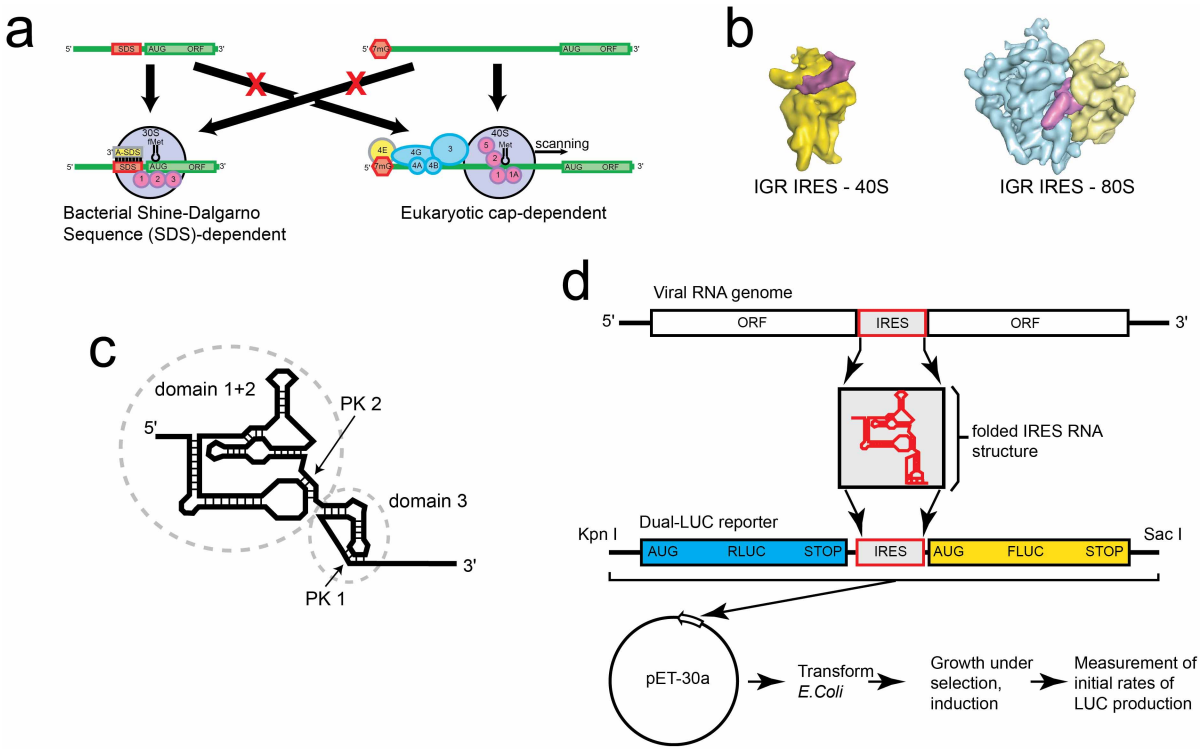
Determination of IRES activity. FLUC and RLUC activity (expressed as RLU) were graphed as a function of time for each culture using the program KaleidaGraph. The initial rate of FLUC and RLUC production was determined using the data from the first 30–40 min after induction. LUC production was generally linear over this time scale after a 5–10 min 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 deviation 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 wild-type PSIV IRES as the template and primers designed to amplify just the DNA of interest under the 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 high-performance liquid chromatography (HPLC). The first RNA used in assembly assays contained nucleotides 6000–6195 of the PSIV IGR IRES, and the second contained this same sequence, plus the sequence GAAAAAGAATTTACCATGGAAGACGCCAAAA ACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAAGATGGAACCG CTGGAGAGC downstream of the IRES.

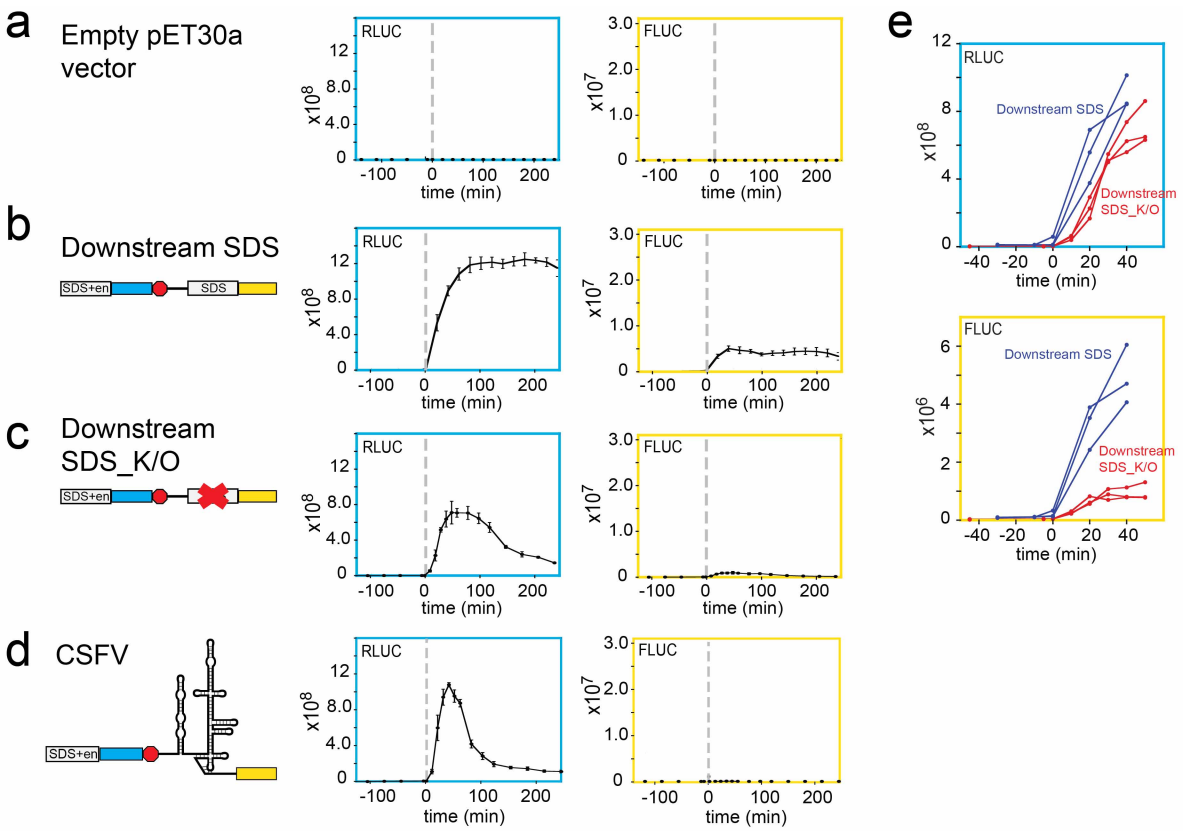
Ribosome assembly assay. RNA for use in assembly assays was 5′-end radiolabelled with P-32 using T4 polynucleotide kinase (NEB), purified by gel electrophoresis, and diluted to 1,000 CPM μl^{-1}. For the assays in RRL, 1 µl radiolabelled RNA was combined with 30 µl of lysate supplemented with amino acids. For the reaction with GMPPNP, 5 µl of a 20 mM stock of the analogue 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^{-1} stock was added to a final concentration of 2 mg ml^{-1}. 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. One microlitre of labelled 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 min, then 250 µl of ice-cold dilution buffer (40 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT)) was added and the reactions were immediately loaded on 15–30% sucrose gradients in dilution buffer. Gradients were centrifuged at 35,000 r.p.m. for 4 h 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]. 0.4 per cent 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 $F_o - F_c$ difference maps were calculated after rigid-body and simulated-annealing refinement was 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.

31. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
32. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
34. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).
35. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).

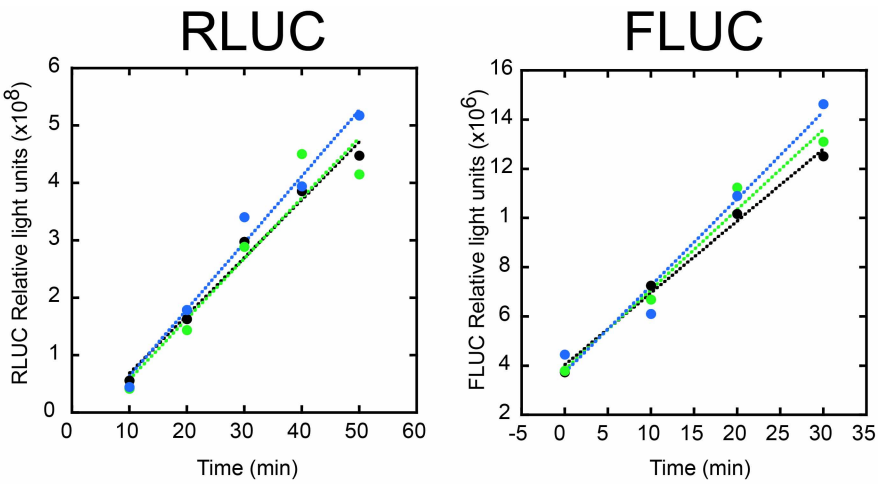


Extended Data Figure 1 | Canonical translation initiation signals, characteristics of IGR IRESs, and experimental design. **a**, Bacterial mRNAs (left) use a Shine–Dalgarno sequence (SDS; red) upstream of the AUG start codon and open reading frame (green) to recruit the 30S subunit (grey). The interaction is through the anti-SDS (A-SDS, yellow). Three initiation factors (magenta) are also important. Eukaryotic mRNAs (right) have a 5′ 7-methyl-guanosine 'cap' (red; 7 mG) 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, RLUC; yellow, FLUC) and this was cloned into bacterial expression vector pET30a. This vector was used to transform *Escherichia 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. RLUC served as a consistent internal control for different bacterial cultures, clones, growth rates, and so on.



Extended Data Figure 2 | Verification of independent quantifiable LUC production in bacteria. **a**, An empty pET30a vector (no inserted LUC reporter coding sequences) shows negligible signal. **b**, Traces of LUC activity as a function of time are shown from a construct in which the RLUC reporter 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 a position to drive initiation of FLUC results in negligible FLUC activity. **a**–**d**, The *y*-axis indicates relative light units (RLU). Error bars represent 1 s.d. 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 min after induction. **e**, Despite differences in longer time courses, LUC production was consistent and linear over the first 30–40 min 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 behaviour, justified their use as a means to quantitate LUC production (Extended Data Fig. 3).

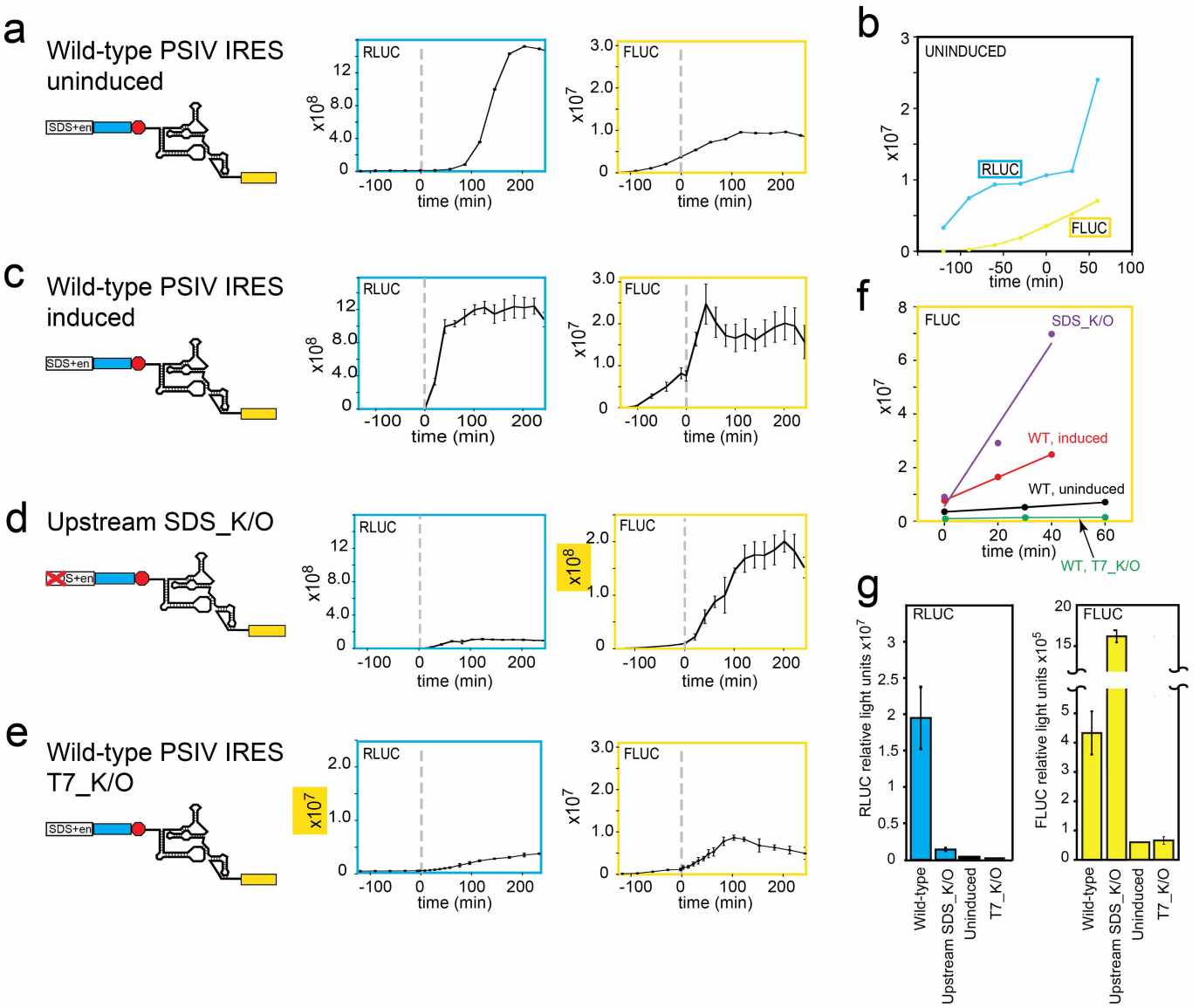
a



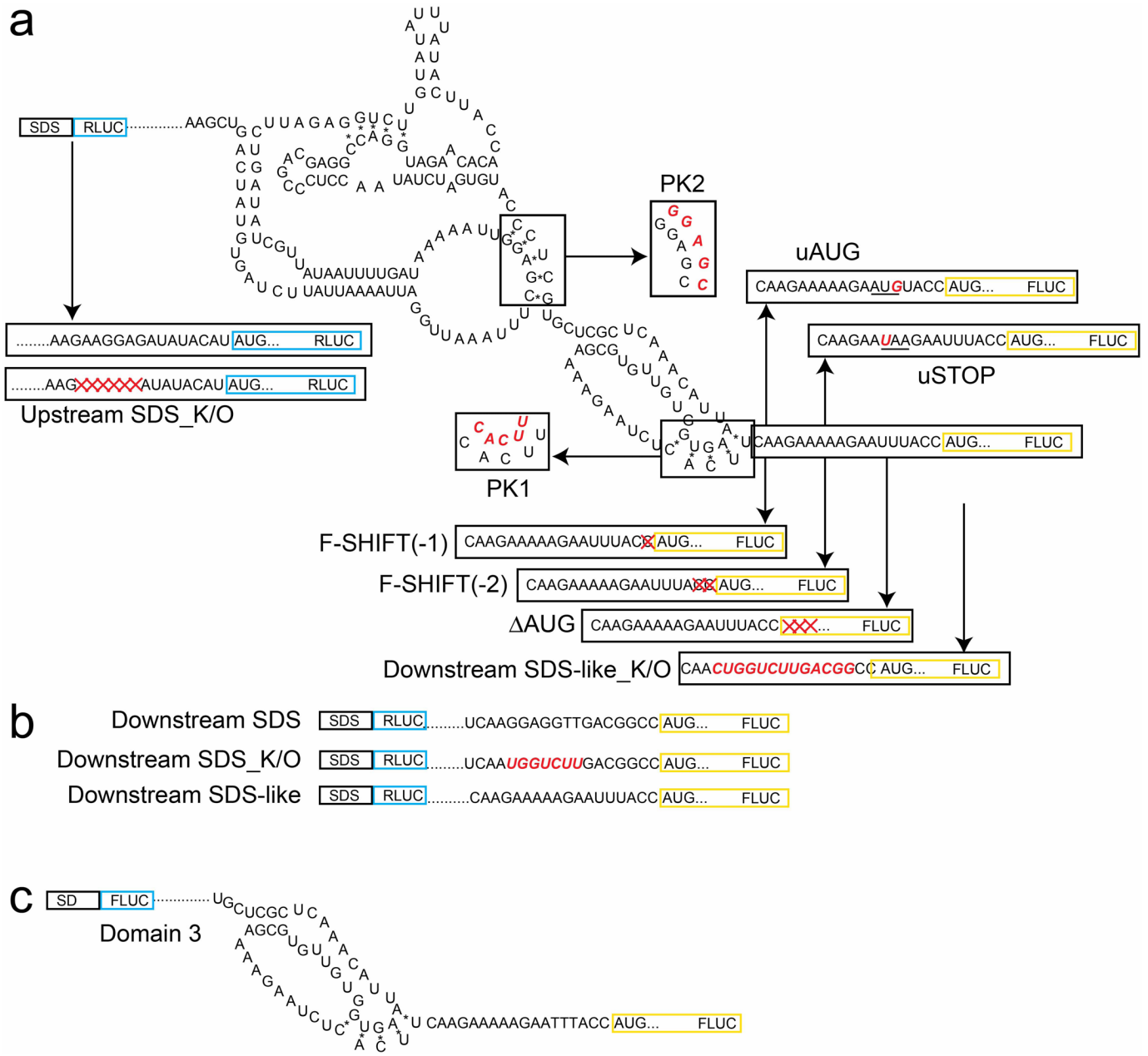
b

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
s.d.:	2.45E+06	8.52E+04	5.51E-03

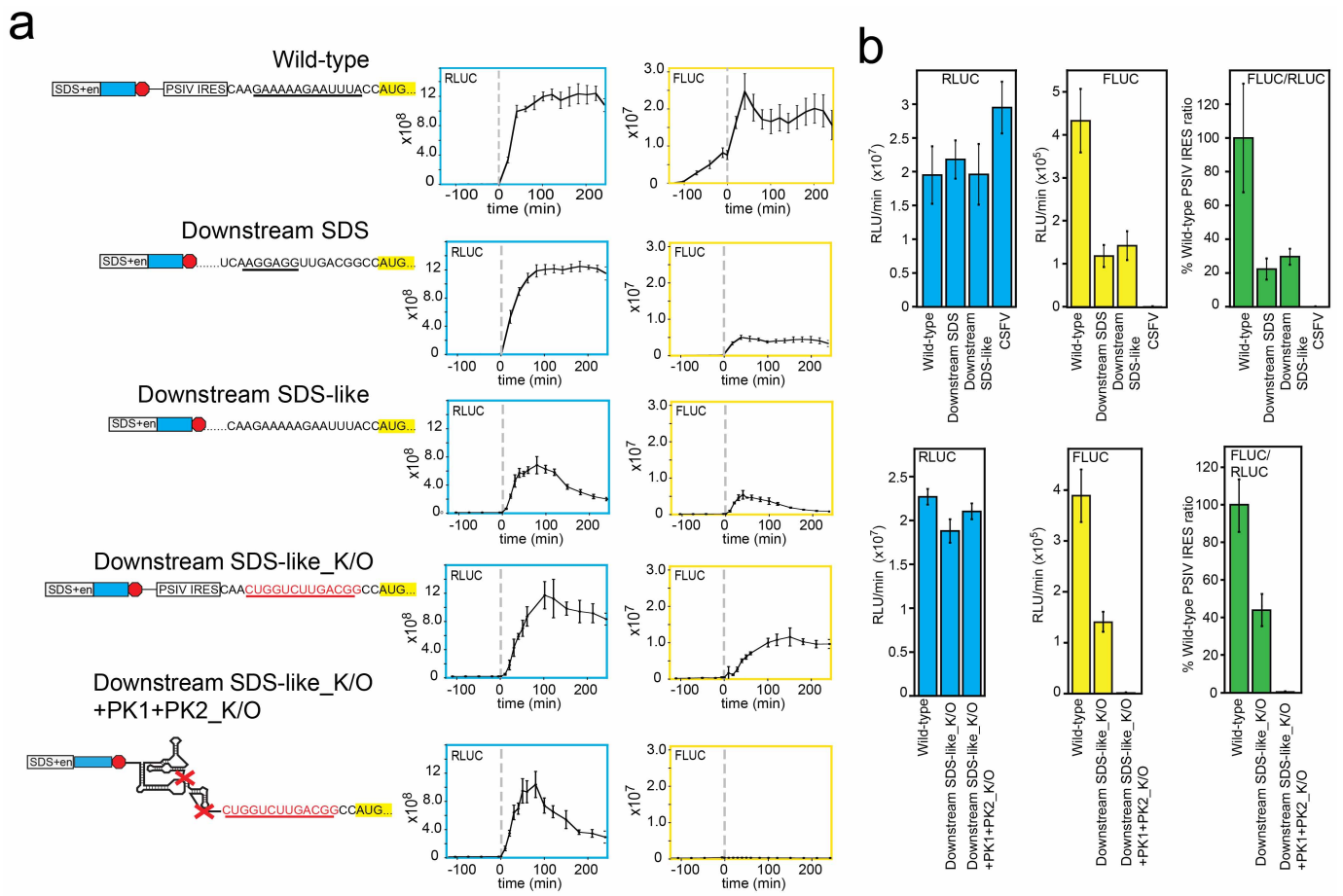
Extended Data Figure 3 | Determination of IRES activity from initial rates of LUC production. **a**, Representative graphs of RLUC and FLUC levels at early time points from three cultures of bacteria transformed with an IRES-containing bicistronic vector, induced with isopropyl-β-D-thiogalactoside (IPTG) 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 are shown, with initial rates for RLUC and FLUC production in RLU min^{-1}. 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 throughout the manuscript.



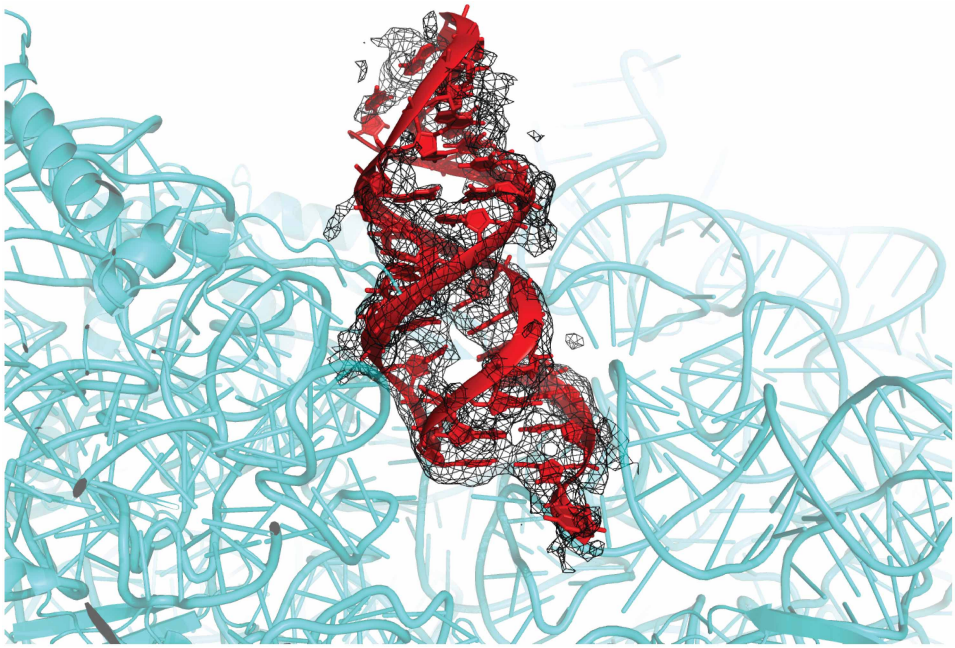
Extended Data Figure 4 | Examination of leaky expression and cryptic promoter activity. **a**, Traces of LUC production from the wild-type PSIV IRES-containing construct without induction with IPTG. Both RLUC and FLUC are produced due to 'leaky expression' of mRNA, a common observation with pET30a bacterial expression vectors. **a**–**f**, The *y*-axis shows RLU. **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 wild-type 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 **f** and **g**). **f**, Initial rates of production of FLUC from the T7_K/O (induced), wild-type (uninduced), wild-type (induced), and Upstream SDS_K/O (induced) constructs. Rates of FLUC production from the T7_K/O and uninduced wild type 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 1 s.d. of the mean from three biological replicates, except the uninduced control, which was done once.



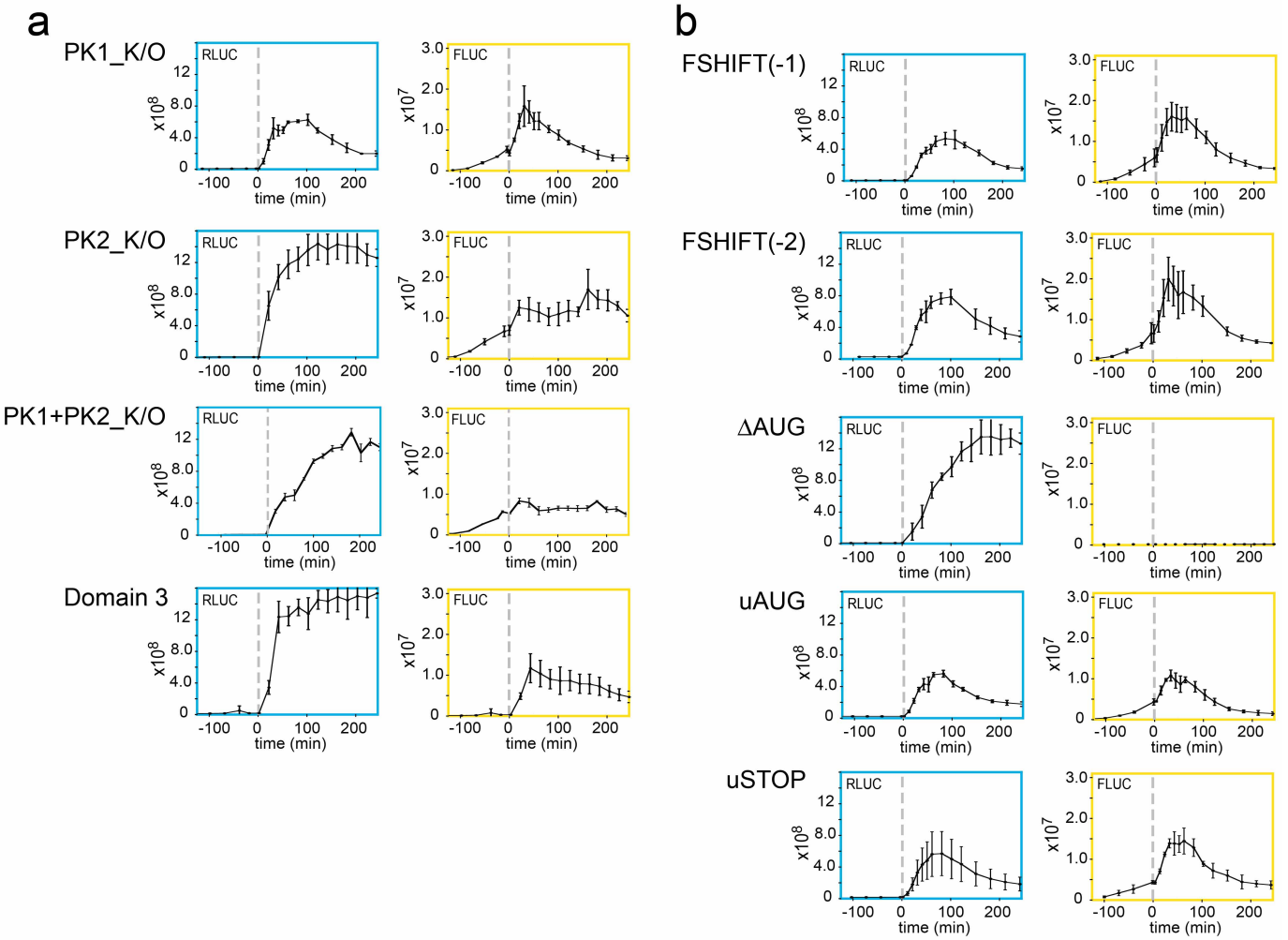
Extended Data Figure 5 | PSIV IGR IRES sequence, secondary structure, and design of mutants. **a**, Secondary structure of the full-length IGR IRES from the PSIV. The specific changes that were introduced to generate the mutants and constructs described and tested in the main text are shown. For each, 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 open reading frame. **c**, Construct containing just domain 3 of the PSIV IGR IRES.



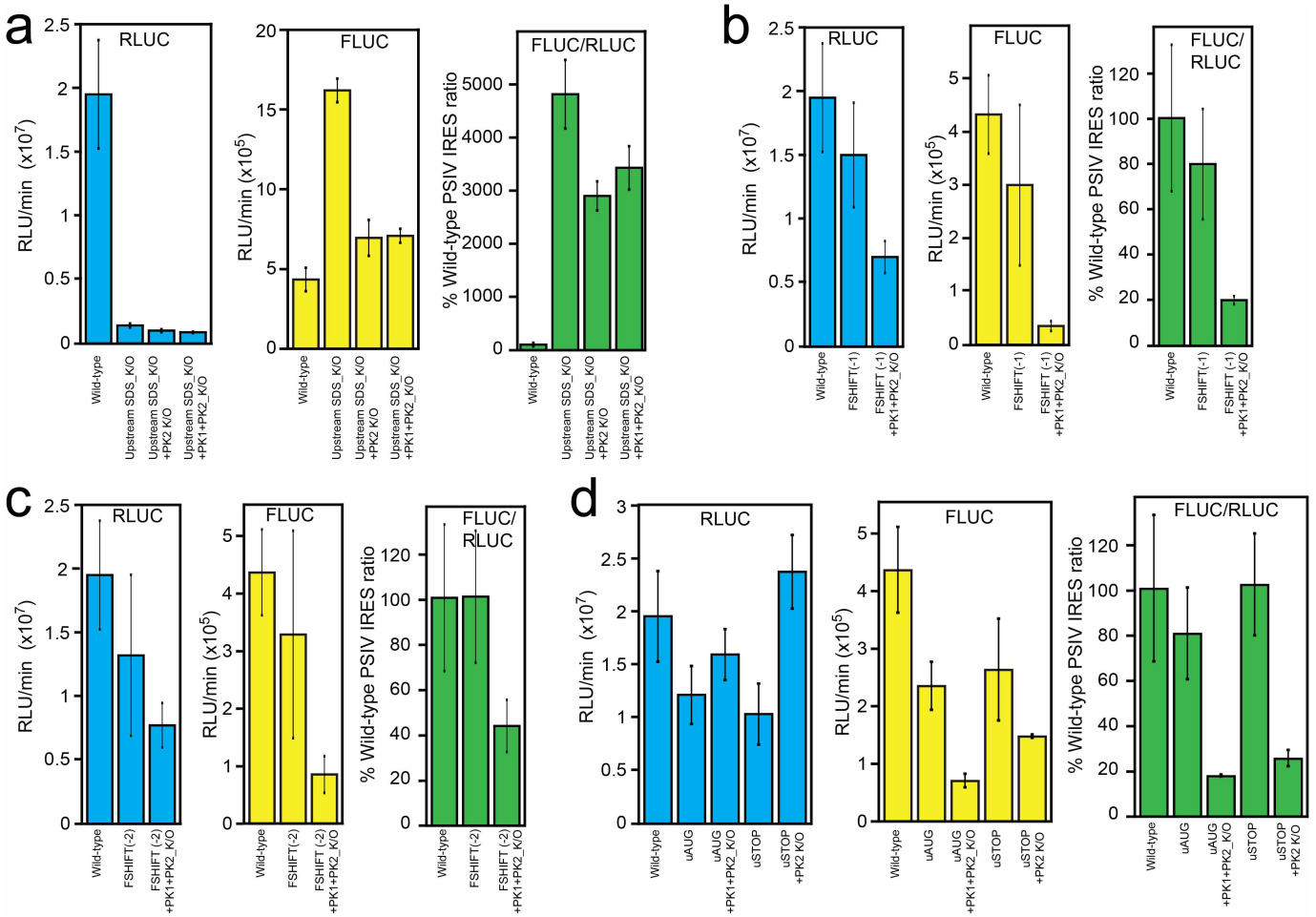
Extended Data Figure 6 | Contributions of region upstream of AUG to initiation activity. **a**, Diagram of constructs tested and traces of FLUC and RLUC production. The *y*-axis shows RLU. **b**, Quantitated initial rates from these constructs. Results from CSFV IRES (negative control) shown for comparison. '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 1 s.d. from the mean of three biological replicates.



Extended Data Figure 7 | The position of domain 3 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 $F_o - F_c$ map for domain 3 in the P site of the small subunit. The large subunit and domain 1+2 are not shown.



Extended Data Figure 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 1 s.d. from the mean of three biological replicates. **a**, **b**, The *y*-axis shows RLU.



Extended Data Figure 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 probably attributable to the decreased competition for ribosomes and the presence of the SDS-like sequence upstream of the FLUC open reading frame and not to robust initiation on the IRES. **b**, The PK1+PK2_K/O dramatically reduced the initial rate of FLUC production on the IRES with the F-SHIFT(−1) mutation. **c**, The PK1+PK2_K/O dramatically reduced the initial rate of FLUC production on the IRES with the F-SHIFT(−2) mutation. **d**, The PK1+PK2_K/O dramatically reduced the initial rate of FLUC production on the IRES with the uSTOP and uAUG mutations. Error bars are 1 s.d. from the mean from three biological replicates.

Extended Data Table 1 | Initial rates of RLUC and FLUC for all constructs tested and crystallographic data collection, phasing and refinement statistics

a

Construct or Condition Tested	FLUC (RLU/min; x 10^5)	RLUC (RLU/min; x 10^7)
T, uninduced	0.59	0.022
WT	4.33 ± 0.73	1.95 ± 0.42
WT, T7_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
Upstream SDS_K/O +PK2_K/O	6.95 ± 1.13	0.097 ± 0.01
Upstream 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
uSTOP+PK1+PK2_K/O	1.46 ± 0.03	2.37 ± 0.35

b

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)
$I / \sigma I$	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

a, Raw values are shown for all constructs tested. All values are the mean of three independent experiments ± 1 s.d. from the mean, except for the uninduced control that was done once. WT, wild type.
b, Crystallographic statistics.
* Values in parentheses are for highest-resolution shell.
† R_{meas} is R_{meas} as reported by XDS[31].
‡ CC(1/2) is the percentage of correlation between intensities from random half-data sets as defined previously[35].