Supplementary information S2 | Viral IRESs and systematic IRES discovery

Viral IRESs are ancient mRNA structures evolved to hijack cellular translation

Viral IRES elements are RNA motifs that can directly recruit the 40S small ribosomal subunit close to the start codon (reviewed in refs.\(^1\)–\(^4\)). This provides sustained translation of the viral genome by host ribosomes as eukaryotic cells inhibit cap-mediated translation initiation upon viral infection. Since the first discovery of IRES elements in poliovirus and encephalomyocarditis virus (EMCV) in 1988\(^5\)–\(^6\), many viral IRESs have been well characterized. Thanks to comparative analysis, there is a growing appreciation of the diversity of viral IRES structures that recruit the translation machinery to the 5’ UTR. Indeed, viral IRESs vary greatly in their requirements for eIFs. Based on these requirements, viral IRESs have been classified into four types that encompass their co-factor needs, RNA structure, location of the initiation codon relative to the IRES, and their activity in an *in vitro* translation system (reviewed in ref.\(^7\)) (see the figure, part a). Only the highly structured and compact Type IV IRESs, found in ssRNA+ Dicistroviridae such as cricket paralysis virus (CrPV), *Plautia stali* intestine virus (PSIV), or taura syndrome virus (TSV), do not require any host initiation factors or initiator methionyl tRNA\(_i\), but can directly recruit the 40S\(^9\) (see the figure, part b). One extreme example of this recruitment is the molecular mimicry of a tRNA-like structure by the IRESs of the Dicistroviridae, including CrPV and TSV. The pseudoknot I (PKI) of the intergenic region IRES of CrPV domain 3 mimics the tRNA anticodon stem-loop bound to a cognate mRNA codon\(^7,9\). PKI binds the 40S which is stabilized by elongation factor eEF2, occupies the aminoacyl (A)-site in the decoding center in the 80S very similar to tRNAs, and is translocated in the 80S ribosome\(^10\)–\(^13\). Thereby, the IRES RNA undergoes a structural change to a more stretched conformation\(^14\). While this structural mimicry appears to be limited to the small region of PKI in the decoding center as PKI globally does not look like a tRNA, the overall IRES structure interacts with the tRNA binding sites on both subunits. Surprisingly, this impressive structural feature of the CrPV IRES RNA is not only able to recruit and initiate translation by eukaryotic ribosomes, but is even able to promote translation in bacteria, albeit less efficiently\(^15\). In its crystal structure, the CrPV IRES occupies the tRNA space on both eukaryotic and prokaryotic ribosomes. This elegantly illustrates how IRESs in RNAs function as structure-driven signals for translation initiation spanning billions of years of evolution. Not only IRESs located in the 5’ UTR can mimic a tRNA shape. The turnip crinkle virus (TCV) IRES has a tRNA-like IRES structure, as suggested by in-line probing\(^16\), and resides in the 3’ UTR\(^16\). This IRES acts as a translational enhancer by binding the 60S from this locus which stimulates initiation upon mRNA circularization\(^17\).

Besides tRNA mimicry, base pairing of rRNA with viral IRES elements at the ribosome binding site is another mechanism to directly recruit the translation machinery, akin to the Shine-Dalgarino (SD) ribosomal binding sequence in bacterial and archaeal mRNAs. For example, the Type III hepatitis C virus (HCV) IRES directly base pairs with 3 nt in both helix 26 and expansion segment (ES) 7 of 18S rRNA\(^18,19\), as shown by hydroxyl radical structure probing and structural mutagenesis. The observation that ES7 is a central anchor point for the highly structured HCV IRES on the human 80S ribosome was confirmed by cryo-electron microscopy (cryo-EM) studies\(^20\) (see the figure, part a). This direct interaction induces conformational changes in the 40S subunit which positions it accurately at the initiation codon as shown by footprinting\(^21,22\). Moreover, variations in the RNA structure of HCV-like IRESs itself can dictate their behavior and may regulate their activity upon stress\(^23\). Importantly, initiation by the HCV IRES starts with 40S binding, but also requires binding to the eIF2 ternary complex and eIF3\(^21,22\) for start codon recognition. In comparison, EMCV IRES requires eIF4A, eIF4B, and eIF4G as co-factors. Together, viral
IRES elements further highlight the notion that the more compact and structured a viral IRES is, the lower its requirement for assisting RBPs and canonical translation initiation factors. Yet, despite the wealth of knowledge about the defined secondary and tertiary RNA structures of viral IRESs, their interactions with the ribosome are less well understood (reviewed in ref.10). After decades of work, we only have cryo-EM structures for CrPV10–13 and HCV20,25–27 IRESs on the ribosome, which show completely different modes of interaction (see the figure). In addition, besides in-depth analysis of viral IRES RNA structures, there are few structural or sequence motifs that are in common between the different classes of IRESs. This variability limits bioinformatics prediction of IRESs28–30, especially in the context of cellular genomes, which also harbor IRES elements.

Alternative modes of internal, structure-dependent translation initiation

Alternative internal initiation can also occur in a cap-dependent but scanning-free manner, via ribosome “shunting”, or cap-independent translation enhancer (CITE) elements. In ribosome shunting, the 40S is bound at the cap followed by discontinuous scanning, in which large 5’ UTR regions are bypassed to reach the initiation codon. Ribosome shunting has been reported for selective translation of several viral RNA genomes such as the adenovirus late mRNA 5’ UTR31,32, in which complex RNA structures of the tripartite leader facilitate shunting by complementarity to the 18S rRNA. Shunting in cellular mRNAs has been proposed for the human heat shock protein 70 (Hsp70) 5’ UTR32 and the structured 5’ UTR of beta-secretase 1 (BACE1) mRNA, an enzyme important in Alzheimer’s disease33. This mechanism is predominant when eIF4F is limited, for example upon infection or during heat stress, but it is not well understood and no common RNA motifs or structures have been associated with it so far. In contrast to shunting, CITEs reside mostly in the 3’ UTR (reviewed in ref.34). They exclusively recruit eIFs in a cap-independent fashion and deliver them to the 5’ end by long-distance base pairing with the 5’ UTR, bridging the 3’ and 5’ UTRs35–37. 3’CITEs are found in uncapped plant RNA viruses and in some cellular mRNAs such as human apoptotic protease activating factor 1 (Apaf-1) acting in apoptosis when eIF4E is limited38. CITEs require scanning from the uncapped 5’ end for initiation. However, many 3’CITEs function equally well when placed in the 5’ UTR. The histone H4 mRNA employs a different method for initiation39. This mRNA has a very short 5’ UTR and two structural elements in its coding region, one of which internally binds eIF4E cap-independently while the other sequesters the cap and thereby positions the ribosome onto the start codon. In fact, cryo-EM data suggest that the histone H4 mRNA directly base pairs with the 18S rRNA to tether the 40S ribosomal subunit40.

Systematic discovery of physiologically-relevant cellular IRESs

In contrast to stress-linked cellular IRESs, certain IRESs are active during and required for normal embryo development and are efficient in mediating cap-independent translation initiation in a physiological and tissue-specific manner. This finding provides the rationale for systematic approaches aimed at increasing the list of functional IRESs under physiological conditions. Indeed, recent unbiased methods systematically screened the transcriptome for novel IRES activity in vitro41 and in mammalian cells42. The in vivo screen measured IRES activity of a synthetic library of 174 nt long oligonucleotides in a bicistronic reporter assay, which places an RNA fragment between two reporter ORFs. While translation of the upstream ORF is cap-dependent, internal translation of the second reporter ORF is stimulated by the IRES in the intergenic region. This analysis led to the identification of putative cap-independent translation elements in hundreds of viral and human 5’ UTRs, as well as unexpected novel IRESs in 3’ UTRs42. Many of them appear to recruit the 40S ribosome by direct 18S rRNA base pairing. However,
perhaps because the relatively short fragment length of the library was not compatible with extensive RNA structures, the predominant novel IRESs reported from the screen were short linear poly(U) motifs or small highly structured elements that did not share any obvious common motifs. The importance of their local RNA structure, assisting co-factors or activity in their natural mRNA context remains to be tested. In fact, empirical and systematic discovery of potential novel IRES elements particularly demand meticulous independent validation of their activity in a set of carefully controlled experiments, as reviewed elsewhere\(^4^3\)\(^4^4\). Experimental limitations and caveats have indeed largely restricted the discovery and confirmation of new functional IRESs\(^4^5\)\(^-\)\(^4^8\). Despite these caveats, the established examples of cellular IRESs represent a means to fine-tune gene expression by internal ribosome recruitment.

**Figure** | **Viral IRESs are classic examples of RNA structures that mediate cap-independent translation initiation**

IRES elements in viral genomes are categorized into four types according to their features such as requirement for stimulating co-factors such as eIFs. Structurally related viral IRESs use distinct mechanisms to mediate internal ribosomal recruitment to an mRNA. 

- **Type III HCV-like IRES** (e.g. HCV, CSFV) 
  - no eIF1, 1A, 4A, 4B, 4F
  - non-structural proteins
  - structural proteins

- **Type IV Dicistroviridae IGR IRES** (e.g. CrPV, TSV, PSIV) 
  - no eIFs or tRNA
  - non-structural proteins
  - structural proteins

  - VPg
  - CrPV IRES
  - non-AUG start codon
  - CrPV domain 3
  - PKI
  - CrPV IRES
  - 40S-IGR IRES complex
  - non-AUG start codon
  - CrPV IRES
  - non-structural proteins
  - tRNA mimicry
  - mRNA channel
  - AUG start codon
  - 40S complex
  - 48S complex
  - eIF3

  - HCV IRES
  - 5' pseudoknot
  - eS27
  - eIF3
  - 3' mRNA channel
  - AUG start codon
  - 40S-IRES complex
  - 40S
  - HCV IRES
  - 60S

  - PSIV IGR IRES
  - Domain 3
  - PKI
  - CrPV IRES
  - 40S
  - 60S
  - 40S-IRES complex
  - non-structural proteins
  - tRNA mimicry
  - mRNA channel
  - AUG start codon
  - 40S complex
  - 48S complex
  - eIF3

  - HCV IRES
  - 5' pseudoknot
  - eS27
  - eIF3
  - 3' mRNA channel
  - AUG start codon
  - 40S-IRES complex
  - 40S
  - HCV IRES
  - 60S

HCV IRES-ribosome complex confirms that the IRES only contacts the 40S subunit, including interactions of rRNA and ribosomal proteins with the IRES as indicated with arrows on the RNA structure model. b The type IV IRESs (~200 nt in length) of the Dicistroviridae genomes include the intergenic region (IGR) IRESs of the cricket paralysis virus (CrPV), taura syndrome virus (TSV) and Plautia stali intestine virus (PSIV). A genome-linked viral protein (VPg, red circle) is covalently linked to the 5' end instead of the cap which is incompatible with cap-dependent initiation. Non-structural proteins are translated by a 5’ IRES whereas structural proteins are translated by the IGR IRES. The PSIV IRES RNA structure includes three pseudoknots (PK). These IRESs require no co-factors or initiator Met-tRNA, for ribosome recruitment but rather mimic the shape of a tRNA for direct 40S binding and translocation inside the decoding center of the 80S ribosome. This is exemplified for the domain 3 (pseudoknot 1, PKI) of the CrPV IGR IRES that mimics the tRNA-mRNA interaction (red, blue regions) and docks into the P-site of the 40S ribosome (ribbon representation from crystal structures reproduced with permission from ref.7). The cryo-EM structure of the CrPV IRES RNA (ribbon representation with PKI in green) and bound to the 80S ribosome was reproduced with permission from ref.11. Type I and II viral IRES (not shown) require more co-factors and are functionally related to cap-dependent initiation.

Part a is from REF. 20, Macmillan Publishers Limited. Part b is reproduced with permission from REF. 7, Elsevier; and REF. 11, Elsevier.

References:
18. Malygin, A. A., Kossinova, O. A., Shatsky, I. N. & Karpova, G. G. HCV IRES interacts with the 18S rRNA to activate


