RNA-mediated translation regulation in viral genomes: computational advances in the recognition of sequences and structures

Asmita Gupta and Manju Bansal

Abstract

RNA structures are widely distributed across all life forms. The global conformation of these structures is defined by a variety of constituent structural units such as helices, hairpin loops, kissing-loop motifs and pseudoknots, which often behave in a modular way. Their ubiquitous distribution is associated with a variety of functions in biological processes. The location of these structures in the genomes of RNA viruses is often coordinated with specific processes in the viral life cycle, where the presence of the structure acts as a checkpoint for deciding the eventual fate of the process. These structures have been found to adopt complex conformations and exert their effects by interacting with ribosomes, multiple host translation factors and small RNA molecules like miRNA. A number of such RNA structures have also been shown to regulate translation in viruses at the level of initiation, elongation or termination. The role of various computational studies in the preliminary identification of such sequences and/or structures and subsequent functional analysis has not been fully appreciated. This review aims to summarize the processes in which viral RNA structures have been found to play an active role in translational regulation, their global conformational features and the bioinformatics/computational tools available for the identification and prediction of these structures.

Key words: RNA viruses; non-canonical translation; bioinformatics algorithms; RNA secondary structure; RNA structure dynamics; viral bioinformatics

Introduction

In biological systems ranging from viruses to prokaryotes to eukaryotes, single-stranded RNA molecules have been shown to be ubiquitously distributed in cells. The sequence and structural features of these molecules cover a gamut of length and size in terms of nucleotides, molecular mass, cellular localization and biological functions. At one end of this spectrum lie small RNAs (~10–100 nucleotides) that include heteronuclear RNAs, small nuclear/nucleolar RNAs and group-I introns, tRNA and small regulatory RNAs like miRNA, siRNA and piRNAs [1]. Further along this spectrum are long non-coding RNAs (>200 nucleotides) like HOTAIR [2] that are believed to be involved in the regulation of gene expression and ribosomal RNAs that encompass larger sequence lengths, ~1400 and ~3000 nucleotides for 23S and 50S rRNAs (prokaryotic 70S ribosome), respectively. With the surge in high-throughput sequencing technologies, numerous small and long non-coding RNAs have been identified, and several reviews comprehensively discuss the functions of these classes of RNAs [3–5]. Finally, at the other end of this spectrum lie many RNA viral
genomes ranging in size from ~1.7 kb for hepatitis delta virus to ~30 kb for Coronaviridae.

All the RNAs from this staggering size range can, except siRNA-5s, miRNAs and certain IncRNAs, in principle, adopt a variety of folded structures, yet available scientific literature documents a limited number of cases where RNA chains that fold into stable, three-dimensional structures are involved with a biological process. These structures are composed of basic building blocks formed by simple helices and loops that associate with a number of ways to generate a variety of tertiary structures. Many viral genomic RNAs are reported to assume a highly complex, extensive and dynamic structural landscape, during different stages of their life cycle as illustrated from studies on dengue virus [6]. This landscape is studded with many RNA structural modules or units that play critical roles in regulating different stages of translation from initiation to termination. The past two decades have witnessed a steady progress in the studies involved in identifying the presence of such RNA structures in viruses. While a combination of chemical-probing techniques and high-throughput sequencing have yielded the entire ‘structuromes’ of viruses like HIV-1 [7], continuous efforts employing X-ray crystallography and nuclear magnetic resonance (NMR) have shed light on various relatively well-defined, smaller RNA structures that are located in specific regions of viral genomes, for example 5′-untranslated regions (UTRs), coding regions, overlapping genes, etc. [8].

While many viruses carry their own enzymatic machinery for autonomous genome replication, transcription, genome assembly and packaging, they completely lack the translational apparatus required for protein synthesis. Apart from the limitations posed in terms of their utter dependency on host translation machinery, viruses are also under tremendous selective pressure to optimize their coding capacity due to their small genome sizes. In order to evade the multiple immunity barriers of their host cells, viruses have evolved countermeasures of their own to disrupt the host’s antiviral mechanisms. As a result, they possess highly sophisticated strategies by which they bypass the normal translation process of their host to recruit the translation apparatus in alternative ways. Furthermore, the different types of mRNAs that are synthesized in viruses exhibit various features that determine the kind of alternative translation they are involved in. For example, the presence or absence of a 5′-cap or a cap-substituted structure in which, unconventional 5′-moeities cap viral mRNAs instead of m7G (7-methyl Guanosine) [9], presence or absence of poly-A tails at 3′-end and finally presence of a range of tertiary structures in either 5′-UTRs, intergenic or coding regions. These alternative translation schemes fall under an umbrella term ‘Non Canonical translation’ strategies. Which is why a wide range of functions involved in controlling and commandeering the host translation apparatus are encoded by viruses [10]. These alternative or non-canonical translation strategies are often mediated by RNA structures encoded within the viral genomes.

This essay is aimed towards reviewing the role of, which have been found to play a critical role in these non-canonical translation pathways. The following sections review the processes where RNA structures/sequences have been found to regulate translation at the stages of initiation, elongation or termination and (i) the complexity and dynamics of these structures, (ii) computational studies in which they were either identified, characterized or where simulation/modeling were performed for their structural analysis, (iii) databases that store the structural and sequence information pertaining to these processes, and prediction algorithms, wherever available. We then conclude by reviewing the ongoing research, to illustrate the techniques that are being developed to efficiently discover and predict novel RNA structures from viral genome sequences. The experimental characterization of these structures as well as those viral RNA structures that play roles other than regulating translation has been discussed in earlier reviews [11–13] and will not be discussed here.

Non-canonical translation regulation mechanisms in RNA viruses

Internal ribosome entry sites

Eukaryotic translation initiation requires recognition of a 5′-modified nucleotide cap (m′GpppN, N = A/C/U/G) of mRNA by pre-initiation complex machinery containing 40S ribosome and some initiation factors (eIF4F, formed by three polypeptides, viz. eIF4A, eIF4E and eIF4G). As the mRNAs of many positive-sense RNA viruses do not possess this cap structure, the canonical signatures for the recruitment of 40S subunit of ribosome are absent. An internal ribosome entry site (IRES) element is an RNA sequence present in either 5′-UTRs (example—Picomavirus family) or intergenic regions (example—Virgaviridae family) of some viral mRNAs that recruit the ribosome at an ‘internal’ position for translation, in a cap-independent manner (Figure 1). These sites were first identified in poliovirus and encephalomyocarditis virus mRNAs, which possessed unusually long 5′-UTR regions with multiple non-initiating AUG codons scattered throughout this sequence [14, 15]. These sequences were proposed to fold in order to adopt highly organized, complex RNA structures that were capable of recruiting ribosomes to initiation ‘AUG’ codons internally on mRNAs that lacked a conventional m′G 5′-cap. Since their discovery, various IRES elements have been identified in the genomes of other RNA virus families, the best characterized examples belonging to Dicistroviridae, Flaviviridae, Retroviridae [16] as well as cellular mRNAs, from both bioinformatics-based predictions and experiments [17]. Viral IRES elements are broadly divided into four groups depending on their secondary structure, involvement of eukaryotic translation factors and distance between IRES element and start codon. Group I IRES-RNA structures bind to the ribosome directly in absence of any protein factors and initiator methionyl-tRNAi [18–20]. Group II IRES-RNA structures bind to 40S subunit directly along with certain initiation factors like eIF2/3 and Met-tRNAi [21, 22]. Group III and IV IRES-RNA bind ribosomes in presence of additional IFs and proteins described as IRES trans-activating factors (ITAFs) [23, 24]. Most of the available structural information about known IRES-RNA structures comes from groups I and II as they directly bind to ribosomes. On the other hand, due to a gap in our understanding of the mechanism by which ITAFs interact with IRES sequence/structure motifs to recruit ribosome, not much is known about IRES belonging to groups III and IV.

Recent computational analyses on data from high-throughput IRES activity assays have attempted to bridge this gap, by employing supervised machine-learning algorithms [25]. In the structure from group I IRES belonging to viruses in Dicistroviridae family, three pseudoknots I, II and III and two conserved stem–loops IV and V were found to interact directly with the ribosome [26, 27]. The different stem-loop segments in full-length IRES structures are often brought together by long-range tertiary interactions like pseudoknots, kink-turns and A-minor interactions (Figure 2A, B) as identified by studies
RNA-mediated translation regulation

Figure 1. Schematic representation of non-canonical translation strategies employed by viruses and discussed in this review. (A) IRES: an extensive RNA structure in 5′-UTR composed of many hairpin stems recruits 40S ribosome and other translation factors. These factors assemble at the structure and scan the mRNA for the nearest AUG codon to start translation. (B) -1 ribosomal frameshifting: an RNA pseudoknot or a stem-loop present in the overlapping genomic region stalls the actively translating ribosome and induces a shift in the reading register, so that downstream translation can be resumed from a new reading frame. (C) Ribosomal shunting: an extensive RNA structure located between a short and a long ORF shunts the 40S subunit of the ribosome between the two ORFs. (D) Stop codon readthrough: RNA structure that is located between two ORFs prevents the reading of the cognate stop codon and releases factor binding so that translation can be continued. (E,F) Reinitiation and non-AUG initiation: in reinitiation, upstream sequence motifs (termination upstream ribosome binding site [TURBS]) interact with 40S subunit to reinitiate translation. In non-AUG initiation, the RNA sequence and/or structure stimulates reading of a near-cognate start codon by the ribosome.

on hepatitis C virus (HCV) and foot-and-mouth disease virus [28–31]. In 2013, a full-length structure of an IRES element from HCV at atomic resolution was presented [32], which combined data from small angle X-ray scattering and molecular dynamics simulation for generating a model structure for HCV IRES. The study used the data from available cryo-electronmicroscopy (cryo-EM) maps of ribosome-bound IRES structures to assemble and fit the atomic resolution structures available for its constituent modular units. The study illustrated that full-length HCV IRES is composed of many extended structured regions that are connected by flexible linkers, such as four-way junctions and pseudoknots [32]. Molecular dynamics simulations have also been used for studying the conformational plasticity of individual IRES domains, dynamics involved in the folding of separate modular units of IRES element from HCV and its dependence on Mg2+/Na+ ions [33, 34].

Viral genomes display high genome variation that is reflected in their 5′- and/or 3′-UTRs and also causes variation in the IRES elements. Such variation in the IRES elements for HCV genome has been documented in HCVIVdb [35] database, which could be used for deducing the effect of sequence variations on different structural domains. Database like IRESite [36] houses the complete list of experimentally and computationally annotated IRES sequences from a broad range of eukaryotic viral and cellular sources (Table 1). The IRES sequences have
Figure 2. Experimentally verified atomic resolution RNA structures. (A) HCV IRES secondary structure showing the modular architecture and corresponding tertiary structures of different units [secondary structure schematic adapted from (40)]. (B) Cryo-EM-derived structure of ribosome-bound cricket paralysis virus IRES. (C) Secondary and tertiary structure (1Z2J) of RNA stem-loop from HIV involved in −1 PRF. (D) Secondary structure of pseudoknot involved in programmed −1 ribosomal frameshift (−1 PRF) from mouse mammary tumor virus (MMTV) and its NMR structure (1RNK).

been found to be poorly conserved among different viral families, although sequences within a single family have shown similarity to a certain level [37]. The low sequence homology among different viral families poses a further bottleneck in in silico IRES prediction using BLAST searches [17]. However, the IRES elements do show similarity at the level of their secondary structure with many sequences adopting an extended structure composed of multiple stems and loops (Figure 2). The similarity found at the level of secondary structures can be used for comparative structural alignment of predicted secondary structures with known IRES elements and could provide a basis for de novo IRES prediction. This strategy is used by programs like IRSS [38] and online prediction server VIPS [39], which were developed to predict IRES, based on secondary structure prediction using minimum free energy methods [40, 41] and subsequent structure alignments using RNALFold [42] (Table 2). More recently, a new IRES prediction tool had been developed for searching these sites in eukaryotic and viral genomes based on support vector machines, called IRESPred [43], and employs ∼35 classifiers including predictions of secondary structure [44] and interaction probabilities between 5′-UTR sequences and proteins from the small ribosomal subunit [45] (Table 2). However, many of these programs that were developed into web servers like VIPS, IRSS and IRESPred are not maintained anymore. IRSS, however, provides useful downloadable Perl and R scripts for carrying out secondary structure prediction and comparative analysis. We have mentioned these programs, nonetheless, in order to highlight the chronological progress in development of approaches and algorithms that led to increase in our understanding of these elements.

Frameshifting

Frameshifting is a recoding mechanism, first described in Rous sarcoma alpha retrovirus where the gag-pol genes were found to be co-expressed from a single polycistronic transcript [46]. Frameshifting is widely used by retroviruses for efficient replication of their genomes and infection [47, 48]. Programmed ribosomal frameshifting (PRF) can occur when the ribosome is systematically guided to express open reading frames (ORFs) that are shifted by −1 (Figure 1), −2 [49] or +1 [50] reading register with respect to each other. It has been reported recently
that organisms other than viruses also employ frameshifting to express the overlapping ORFs [51–53]. –1 PRF process per se is coordinated by three key determinants: a 7 nucleotides long 'slippery-site', a 5–9 nucleotides long spacer sequence and an RNA structure that could be either a simple stem-loop or a pseudoknot. Several studies have reiterated the fact that an RNA pseudoknot is a more efficient stimulator of frameshifting than a structurally simple hairpin loop [54, 55]. According to Recode V2.0 database, a large number of viruses use an H-type pseudoknot for efficiently mediating –1 PRF process [56]. In single-stranded RNA viruses belonging to Coronaviridae (SARS-CoV) [57], Luteoviridae (sugarcane yellow leaf virus) [58] and Astroviridae [59] families, –1 ribosomal frameshift has been found to be modulated by RNA pseudoknots. The retroviruses (HIV-1, HTLV-2) [60, 61] display involvement of both pseudoknots as well as extended hairpins in the process [62].

The H-type pseudoknots consist of two helical stems (named S1 and S2) interspersed by two or three loops (named L1, L2 and L3) [8]. Crystallographic and NMR studies have suggested that the two helical stems in these H-type pseudoknots are twisted and bent with respect to each other [63–65], as observed from the solution structure of mouse mammary tumor virus (MMTV) RNA pseudoknot where the two helical stems showed a bending angle of ~121° [66] (Figure 2D). This kink/hinge arises mainly due to a single nucleotide in loop 2 that intercalates itself between the helical stems, thus preventing co-axial stacking of the two helices. The loop and stems interact through multiple non-canonical hydrogen bonds involving the sugar edge of stem nucleotides and Watson–Crick face of loop bases, thus providing added mechanical strength to the structure [8, 67]. The structural features of frameshifting inducing RNA pseudoknots have been discovered mainly by NMR [66, 68], crystallographic [69] and molecular dynamics studies [70, 71], which also highlight the importance of ions and ion-coordinated water molecules in holding such a compact structure together. The physical significance of the presence of a kink in frameshifting inducing structures is important as it was observed that structurally simpler hairpin structures also displayed a kink between their helices [72, 73] although the helices in these hairpins were larger than the constituent helices of retroviral and luteoviral pseudoknots [72] (Figure 2C).

PRFdb [74] catalogs the programmed ribosomal frameshift signals filtered from Yeast Genome project and Mammalian Gene Collection [75]. As is the case with IRES elements, RNA structures (pseudoknots/hairpins) share little sequence similarity among different viral families, although within a sub-group the sequence signals do exhibit a high degree of sequence conservation, as was observed for retroviruses. The length of these structures vary from the tightly compact polerovirus and enamovirus pseudoknots (~30 nucleotides) to extended coronaviral structures (~200 nucleotides). This renders their in silico prediction extremely difficult. Recent studies have been devoted to search for existing PRF signals and develop predictive rules for identification of novel signals, not only in viruses but also across different domains of life, e.g. KnotInFrame [76] and FSFinder [77, 78]. These algorithms employ a set of rules in which a preliminary screening of putative frameshift sequences is done based on consensus slippery sequence motif (X XXY YYZ) followed by secondary structure prediction on filtered sequences using minimum free energy programs like pknotsRG [41] and RNAFold [40] (Tables 1 and 2). While some algorithms can predict stem–loop frameshift structures (FSFinder), others take into account the pseudoknot stimulators in –1 PRF (KnotInFrame). The accuracy of the prediction of novel –1 frameshift signals could be further augmented by the fact that overlapping regions from different RNA viral sub-groups show high sequence similarity, a feature that could be used to increase the reliability of secondary structure prediction (Table 2). Apart from these predictive programs, databases like FSDB [78] store all the experimental and predicted frameshift hotspots, and RECODE-2 [56] houses all the information related to translation recoding events and signals involved. Programmed frameshifting is also utilized by prokaryotic and mammalian genomes for translating multiple products from same set of ORFs. FSscan [79] predicts such +1 frameshift signals in Escherichia coli. A remarkable feature that has been observed in case of –1 PRF inducing RNA structures is that relatively smaller structures like RNA pseudoknots from MMTV and other viruses can act as ribosomal roadblocks and induce –1 PRF. This extraordinary ability has been attributed to not

### Table 1. List of existing tools and databases dedicated to storing information and prediction of sequence signals involved in non-canonical translation processes in viruses and other organisms

| Program          | Description                                                                 |
|------------------|-----------------------------------------------------------------------------|
| IRESite          | Database housing experimently annotated cellular and viral IRES elements, as well as in silico predictions freely available to scientific community for experimental verification [36]. URL—http://iresite.org/ |
| HCVIVdb          | Database of HCV IRES sequences with published natural or engineered mutations [35]. URL—http://hcvivdb.org/ |
| PRFdb            | Database of ribosomal frameshifting signals from eukaryotic database [41]. URL—http://prfdb.umd.edu/ |
| KnotInFrame      | Prediction of –1 PRF signals [42], uses pknotsRG as a background program for secondary structure prediction. URL—http://bibiserv.techfak.uni-bielefeld.de/knotinframe |
| FSFinder2        | Prediction of –1 and +1 ribosomal frameshifting sites in genomic and mRNA sequences [43, 44]. URL—http://wilab.inha.ac.kr/fsfinder2/ |
| FScan            | Prediction of +1 ribosomal frameshifting signals [45]. Python-based framework; detailed algorithm described in associated reference. Currently, there is no web server implementation. |
| FSDB             | Database of experimentally verified and predicted ribosomal frameshifting [46]. URL—http://wilab.inha.ac.kr/fsdb/ |
| Recode           | Database of experimentally known translation recoding events and signals [47]. URL—http://recode.ucc.ie |
just the thermodynamic and mechanical strength of these structures but also to their conformational plasticity, defined by their abilities to adopt alternate conformations during unfolding, as documented in recent RNA unfolding studies employing optical tweezers [80, 81].

Ribosome shunting
Ribosome shunting is a translational mechanism that comprises features of both 5’-cap-dependent translations but is partially independent of internal sequence scanning. It was first discovered in plant pararetroviruses in the family Caulimoviridae and subsequently in rice tungro bacilliform virus [RTBV, a cauliflower mosaic virus (CaMV)-related plant pararetrovirus] [82, 83]. It was found that the mRNA (also pregenomic RNA, pgRNA) leader sequences in the members of this family were unusually long. The long leader sequences folded into a large stem-loop structure and contained several small ORFs (sORFs) (Figure 1). Different chemical and enzymatic probing studies have shed light on the architecture of this structure. In CaMV, a sORF and a strong downstream hairpin constitute a minimal shunting element. In CaMV, the 40S subunit along with a complement of initiation factors assemble at 5’-cap of mRNA, scan for a short 60–70 nucleotides downstream till an AUG codon of the proximal sORF. All the scanning ribosomes assemble at this codon to form complete 80S subunit. After initiating translation of the first sORF and release of the newly formed short peptide, the 80S ribosomes disassemble. A fraction of the remaining 40S ribosome then shunts (bypasses) over the downstream 500 residue-long sequence that forms a pseudoknot, connecting central and terminal parts of leader and a hairpin dimer at high ionic strength [85]. Shunting is believed to expand the coding capacity of mRNA by directing the ribosomes to inter-

| Program            | Background algorithm      | Description                                                                 | Performance parameters |
|--------------------|--------------------------|-----------------------------------------------------------------------------|------------------------|
| VIPS* (IRES)       | RNALfold [48]            | For calculating local, thermodynamically stable RNA secondary structure, minimum free energy parameters [49, 50] | Accuracy—51.87%        |
|                    | RNA Align [51]           | Comparative secondary structure analysis                                      | Specificity—81.08%      |
|                    | pknotsRG [52]            | Prediction of pseudoknoted regions in the predicted IRES secondary structures | Sensitivity—23.28%      |
|                    |                          |                                                                             | Precision—55.69%        |

Advantages

• Predicts both viral and cellular IRES structures.
• Can predict IRES structures with pseudoknots.

Limitations
Due to dependency of algorithm on sequence and structural features conservation, prediction of cellular IRES is poor, since cellular IRESes mostly lack any consensus sequence/structural features.

| Program            | Background algorithm      | Description                                                                 | Performance parameters |
|--------------------|--------------------------|-----------------------------------------------------------------------------|------------------------|
| IRESPred* (IRES)   | RNAFold [53]             | Support vector machine-based classifiers, RNAFold and RPISeq were used as back-end programs to compute classifying features | Accuracy—70.89%        |
|                    | RPISeq [54]              |                                                                             | Specificity—71.95%      |
|                    |                          |                                                                             | Sensitivity—69.84%      |
|                    |                          |                                                                             | Precision—71.35%        |

Advantages
Prediction scores were consistently better than VIPS, since algorithm is independent of intrinsic sequence conservation bias.

Limitations
Principal parameter used for prediction algorithm is the interaction between IRES sequence and 40S ribosome. This interaction is not conserved in cellular IRES and viral IRES from HCV and cricket paralysis virus. Hence, the algorithm is unable to predict IRES when they lack any of the features defined in the feature set of machine-learning algorithm.

| Program            | Background algorithm      | Description                                                                 | Performance parameters |
|--------------------|--------------------------|-----------------------------------------------------------------------------|------------------------|
| KnotInFrame (~1 PRF) | PknotsRG-fs             | Constraint-based folding of input sequence to enforce pseudoknot formation, modified from original pknotsRG program | Ranking the predictions based on differences in constrained and relaxed mfe values |

Advantages
Computationally efficient, scans complete genomes within few hours.

Limitations
Prediction accuracy is as good as the accuracy of thermodynamic parameters used for RNA secondary structure prediction.

| Program            | Description                                                                 | Performance parameters |
|--------------------|-----------------------------------------------------------------------------|------------------------|
| FSFinder           | The algorithm works by scanning for slippery sequence motif and estimating base-pairing possibility in the contextual region for presence of stimulatory signals. | Sensitivity (~1 FS)—88% |
|                    |                                                                             | Specificity (~1 FS)—97%|
|                    |                                                                             | Sensitivity (+1FS)—72%|
|                    |                                                                             | Specificity (+1 FS)—92%|

Advantages
Predicts both ~1 and +1 frameshift sequences.

Limitations
Prediction of +1 frameshifting has been tested on only two genes: protein chain release factor (prfB) and ornithine decarboxylase antizyme (oz2). Hence, prediction accuracy is limited for +1 frameshift.

*Indicates databases which are no longer being maintained.
nal coding regions by bypassing upstream sORFs. The shunting mechanism is employed by all genera of plant pararetroviruses, late adenovirus mRNA and animal viruses [86, 87] and is found to be conserved in all plant pararetroviruses, making it a widely used strategy of non-canonical translation.

Computational identification of a shunting element in rice tungro spherical virus, RTBV and CaMV revealed that the leader sequences were highly rich in GC content [88]. RNA secondary structure prediction using earliest dynamic programming algorithms like Mfold [89] (for thermodynamic parameters used to derive free energy values, see [90]) on the pgRNA leader sequences of these viruses suggested presence of a structure containing several stem-loops, which were found to be present in both the optimal and sub-optimal structures. The upper portion of these structures was capable of adopting various different conformations, as highlighted in recent review [86]. The shunt efficiency was determined by the stability of the helix located at the base of structure [88]. Unlike IRES and ribosomal frameshift signals, currently there are no in silico tools that can identify these structures.

Stop codon readthrough

Eukaryotic translation terminates at three stop codons, namely UAA, UAG and UGA, mainly due to recognition of these codons by class I release factors (eRF1) [91] and because of absence of any cognate tRNAs that might contain an appropriate anticodon sequence to recognize and pair with these codons [92]. The termination process not only depends on stop codon sequence but also on the proximal contextual sequence with some bases resulting in a significantly ‘leaky’ stop signal [93]. These leaky codons allow ‘read-through’ of the translating ribosome (Figure 1). In this process, the stop codon is recognized by near-cognate suppressor tRNA that allows translation to proceed and terminate at the next stop codon [92]. In many ssRNA viruses belonging to Luteovirus (Luteoviridae family), Alphavirus (Togaviridae family) and Tobamovirus (Virgaviridae family), programmed readthrough takes place, resulting in a C-terminal extension of the initial protein. This extension often encodes the viral RNA-dependent RNA polymerase or an extension to the coat protein [94–96]. Like programmed frameshifting, the efficiency of programmed readthrough is also modulated by features located in the 5′- and 3′-end of the suppressed stop codon, such as adenosine residues in the region immediately 5′ of the stop codon [97]. However, it is the presence of many sequence and structural features in the 3′-end of the stop codon that stimulates and enhances the readthrough efficiency. These signatures fall into three major classes, group I, II and III, out of which viral readthrough signals belonging to group II and III contain an RNA secondary structural element [96]. In type I motifs, readthrough across a UAG codon is stimulated by a six nucleotide motif having a consensus sequence of UAG_CAR_YYA (R = purine, Y = pyrimidine) and are mainly present in tobamoviruses to allow translation of polymerase [98, 99]. Type II motifs consist of a UGA stop codon followed by CGG or CUA and an extended stem-loop structure ~8 nucleotides downstream [100]. Type III motifs consist of a UAG stop codon and purine-rich octanucleotide spacer sequence followed by an RNA structure that has been found to be an RNA pseudoknot in cases of gammaretroviruses [98, 101]. Apart from these classes, a novel type of readthrough signature in RNA viruses was identified recently in ORF5 of potato leafroll virus (Luteoviridae family), which consisted of a C-rich region in the vicinity of coat protein stop codon and a distal RNA stem-loop structure ~640 nucleotides downstream of this ‘leaky’ stop codon. The presence of stem-loop structure was validated by SHAPE [102] and was found to be essential for readthrough protein translation.

The identification and characterization of readthrough stimulating RNA structure in viruses belonging to group II category involved a preliminary computational screening using BLAST [103], EMBOSS [104] and ClustalW to predict RNA secondary structures using sequence conservation and dynamic programming algorithms RNAfold [44] and PknRG [41]. The approach identified a phylogenetically conserved stem-loop structure [100] in alphaviruses, which was later verified experimentally by chemical probing [105]. The stem-loop was found to have 10–12 base pairs with a 1 nucleotide asymmetric bulge [106]. The presence of a classical H-type RNA pseudoknot in the 3′ region of the readthrough signal of murine leukemia virus (MuLV) and other retroviruses was also verified by chemical probing and computational studies that identified loop and stem regions, which were critical in stimulating the efficiency of readthrough [101, 106, 107]. RNA-folding algorithms further constituted a powerful and reliable tool to screen sequences from plant RNA viruses belonging to Virgaviridae family and were found to harbor potential phylogenetically conserved stem–loop structures involved in readthrough stimulation [100]. These RNA structures have been proposed to modulate readthrough in a number of possible ways that involve interference with release factors binding or unwinding by ribosomal helicases or by pausing the ribosome in a way akin to frameshift inducing pseudoknots [105].

Non-AUG initiation and reinitiation

Non-AUG initiation: In 1988, it was discovered in Sendai virus [108] and then in Moloney MuLV [109] that initiation is also possible at several alternative start codons other than the cognate AUG start codon but that are near cognate in sequence like ACG and CUG. It has been found that codons like ACG, CUG, AUU, AUU, AUC, etc. can initiate translation at ~2–30% of the levels obtained with an AUG start codon [110]. Initiation at a non-AUG codon depends strongly on the sequence context (A/G at −3 and a G at +4). This is significantly stimulated by the presence of an RNA structure that forms nearly 14 residues downstream of this codon [111] (Figure 1). This spacing appropriately positions the tertiary structure at the mRNA entry tunnel of the ribosome so that the potential non-AUG codon lies at the P-site of ribosome, while the ribosome is paused at the base of structure and unwinds it. Because of the relative inefficiency of a non-AUG codon initiation, some ribosomes eventually slide past this site until they encounter a cognate AUG or near-cognate codon.

Identification and characterization of an RNA structure involved in non-AUG initiation in dengue virus type 2 was done using Mfold [89] secondary structure prediction, followed by selecting a candidate minimum free energy structure obtained after a refining step from the ensemble, done using Vienna RNA package [112] and subsequent experimental validation [113]. Tertiary structural information regarding RNAs involved in the stimulation of the process is sparse, and in most known cases, presence of an optimal sequence context is sufficient. Hence, any bioinformatics prediction tool will have to rely heavily on scanning these features.
Reinitiation: Eukaryotic translation terminates when the ribosomal A-site encounters a stop codon. This is then followed by release of 60S subunit, aided by several release factors (eRF1 and eRF3, eRF1 being responsible for codon recognition). The dissociation of 60S subunit leaves a 40S/deacylated tRNA still bound to mRNA, which is dissociated following a sequence of events. However, in certain incidences, the 40S subunit remains associated within the message and reinitiates translation at a downstream AUG codon [114] (Figure 1). This happens when translation has terminated at very short (usually <30 codons) ORFs that can give rise to 40S subunits capable of scanning and reinitiating translation at a downstream AUG codon [115]. The chances of reinitiation depend on the length of upstream ORF and distance between the termination codon of this ORF and start codon of downstream ORF [116, 117]. Termination of translation of the upstream ORF is a mandatory requirement in order to distinguish this process from internal translation initiation by IRES.

As studied in viruses from Caliciviridae family, containing consecutive protein-coding ORFs, ORF1, ORF2 and ORF3, it was illustrated that reinitiation depends on an RNA sequence motif (UUGGA, along with proximal flanking nucleotides) [115, 118]. This sequence motif is located ~40–90 nucleotides upstream of ORF2 termination codon and forms base pair with loop region of helix 26 in 18S rRNA (Figure 1). These motifs are called termination upstream ribosome binding site (TURBS). The complex of TURBS and 18S rRNA has been shown to bind eIF3 and prevent 40S subunit from dissociating from the transcript.

Translation enhancers in the 3′-UTRs of viral genomes

Apart from these functions, certain RNA structures like 3′-cap-independent translation enhancers (3′-CITE) and tRNA-like structures (TLS) that have been found to be present in the 3′-UTRs of many positive strand RNA viruses. 3′-CITEs are found in the viruses from family Tombusviridae and genus Luteovirus. They enhance translation by recruiting multiple initiation factors or ribosomal subunits at 3′-end, to form a complex, which is subsequently brought near 5′-end to initiate translation [119]. Due to high structural diversity among different 3′-CITE structures, they have been classified into seven different classes on the basis of their structural features, types of initiation factors recruited and mode of interacting with the 5′-end. These aspects have been covered in a comprehensive way in recent review [120]. The high structural variability makes the bioinformatics prediction of 3′-CITEs extremely difficult. Furthermore, like RNA structures in shunting, the absence of experimental structural data from biophysical methods is a great hindrance in studying the conformational dynamics and folding landscape of these structures. However, studies proposing a tertiary structure model using programs like RNAComposer [121], generated from secondary structure predictions, have shed light on the putative conformational features of these structures.

Other class of such RNA structures is TLS, which are present in 3′-termini of many positive-strand RNA viruses across multiple genera. They functionally mimic tRNA molecules and also act as translation enhancers. They are involved in viral encapsidation, regulation of negative strand synthesis, and possess tRNA-like functional features such as ability to be aminoacylated and interaction with elongation factors (EF1A). The structural information available from chemical and enzymatic probing studies has been utilized in studying their dynamics using single-molecule Förster Resonance Energy Transfer (smFRET) studies [122] and can provide valuable preliminary data for setting up molecular simulation and modeling studies.

Non-canonical translation strategies employing RNA sequence motifs

In some plant RNA viruses under certain conditions, the majority of ribosomes do not initiate translation at the first AUG codon but continue scanning until they reach an alternative downstream start codon [123], in a process called leaky scanning. The efficiency of leaky scanning has been reported to depend on a sub-optimal sequence context of the initiator methionine AUG codon (defined from +1 to +3) defined by absence of a G at +4 and A/G at −3 position. Presence of these residues constitutes an optimal sequence context for canonical translation initiation to take place [124]. Leaky scanning can be stimulated when the AUG codon is located very close to 5′-end of the UTR, particularly when the length of UTR is <15 nucleotides [125]. When the two AUG codons are located in close vicinity to each other (<6 nt separation), the potential for leaky scanning increases [126]. As an example, in segment 6 of influenza virus B, leaky scanning involves alternating forward and backward movements about a downstream AUG codon that allows capturing of a proportion of scanning ribosomes in order to stack them in a position appropriate for initiating translation at the preceding AUG codon [127]. In vitro expression studies in plant plum pox virus (genus Potyvirus) have revealed that leaky scanning mainly depends on presence of initiator codons located in an optimal sequence context, and no RNA secondary structure has been found to regulate or stimulate the process. While programs like TITER [128] have been developed to predict alternative translation initiation sites (TISs) in mammalian genomes, no such dedicated tools are available for predicting alternative TISs in viral genomes.

Insights from high-throughput RNA structure probing

Apart from IRES and ribosomal frameshifting, the current knowledge on the tertiary structural organization of RNA structures involved in ribosomal shunting, reinitiation and non-AUG initiation is acutely limited. This creates a large vacuum in the current understanding of RNA structure–function relationship as well as mechanisms by which these structures engage and recruit ribosome and various translation factors. Recent studies have highlighted the importance of whole-genome studies like Frag-Seq and PARS, accompanied by empirical structural modeling tools to come up with an accurate model of IRES elements [129] in Picornaviridae [37, 130]. Studies of a similar magnitude are still required for investigating the structures involved in −1/+1/−2/+2 ribosomal frameshift, stop-codon readthrough, shunting and reinitiation. Although getting a secondary structure map of entire genomes from high-throughput sequencing is now a routine process, inferring the long-range tertiary interactions from secondary structure components still remains a challenging task. Moreover, even the large-scale viral genomic maps, yielded by chemical probing data such as SHAPE, provide an incomplete picture, since protein induced dynamic structural changes are not taken into account in these analyses [131]. Recent approaches combining chemical probing with simultaneous native gel electrophoresis (in-gel SHAPE [132]) and high-throughput sequencing [132], ribosome profiling, fluorescence-activated cell sorting and viral deep sequencing [134] have paved
the way for developing deep-learning algorithms, capable of identifying sequence motifs and have provided fresh insights into the local, highly variable conformations of 5′-/3′-UTRs and coding regions [135]. Of the recent techniques developed to study RNA structures in vivo/in vitro in a high-throughput way, SHAPE technique has been tweaked extensively to generate different variants suitable for specific purposes, e.g. SHAPE-Map, SHAPE-Seq and aiSHAPE [133, 136]. Mod-Seq is another pipeline developed to study the in vivo and in vitro structures of long RNAs, e.g. ribosomes that have been combined with deep sequencing [137]. More recently, a state-of-the-art technique called LASER-Seq [138] has been developed to probe structures directly in cells, which exploits conformation-related parameters like solvent accessibility. A recent in vivo analysis of RNA conformational dynamics in Zika virus [139] has revealed the vast and fluid conformational landscape of entire RNA genomes. The observations from these studies combined with state-of-the-art computational RNA-folding algorithms could provide a stepping stone to predict RNA tertiary structure folding and long-range interactions that could vastly improve our current understanding of the mechanism by which various RNA structures modulate translation in viruses. There has been a steady rise in the development of bioinformatics pipelines along with aforementioned high-throughput structure probing algorithms; a few, but in no way comprehensive, examples being Mod-Seeker [137], StructureFold [140] and dStruct [141], along with excellent reviews describing these and other pipelines [142, 143].

Conclusion

Nearly all the non-canonical protein-coding strategies used by viruses have been found to be employed by eukaryotes as well. With increasing instances of such mechanisms being identified in yeast and mammals, efforts have redirected towards exploiting features of these mechanisms towards biotechnological and biomedical purposes. The ability to recruit the ribosome at an internal location, independently of cap-recognition and 5′-UTR scanning events, by IRES structures has serious biotechnological implications. The IRES elements have been used for expression of synthetic bi-cistronic [144] and multi-cistronic [145] constructs to express proteins of desire in a number of experimental setups. However, the lack of any global sequence conservation across different viral families poses technical challenges in designing effective sequence constructs for in vitro experiments. Recent computational studies employing RNA inverse-folding methods, which involve designing a sequence capable of assuming a desired reference structure, have proved to be valuable tools in de novo IRES predictions [146]. Algorithms developed for inverse folding are being developed, keeping specific scientific goals in perspective. They range mostly from heuristic methods, which focus on providing optimal sequence sets using user-specified constraints iteratively until a desired target structure is achieved [147], to probabilistic approaches like INFO-RNA [148] and methods employing statistical mechanical sampling schemes, reviewed in [149]. In addition to sequence limitations, the paucity of available structural information on a full-length, ribosome-free and bound IRES elements present another technical challenge for studying the conformational dynamics and folding landscapes as well as in setting up simulation and homology modeling studies.

However, despite the advances in computational power and development of algorithms for tertiary structure prediction/modeling, much of our current understanding is built on the information provided by computational studies that relied heavily on deciphering secondary structure information from sequences, a prime example being the studies by Tuplin et al. [150, 151] on entire HCV viral genomes to yield genome-scale-ordered RNA structures. The selection pressure on viruses to maintain the features in the regulatory regions, and suppression of synonymous site variability, is an often utilized feature in viral bioinformatics [135]. Hence, although the future lies in developing tertiary structure prediction, secondary structure prediction will still remain a gold standard for predicting novel structural motifs.

Key Points

- A range of non-canonical translation strategies employed by viruses involve RNA sequence/structure motifs; however, the experimental identification of such sequences and structures has remained an extremely challenging process to date.
- In almost every known example of non-canonical translation mechanism, computational studies have provided preliminary information on the consensus RNA sequences involved in regulation, hence providing a stepping stone for carrying out subsequent validation experiments.
- A bottleneck in any computational study involving viral RNA structures is posed by the high mutational rates in viruses, resulting in poor homology and identification of consensus sequence motifs. Although many alternative translation strategies have been identified in viruses, the detailed mechanism by which RNA structures regulate or stimulate these processes remains to be understood.
- With development of state-of-the-art computational platforms, significant progress has been made in ab initio RNA structure modeling and deep-learning algorithms, which have proved to be an invaluable tool to fill the lacuna in the knowledge base of the available structural information within RNA viruses.

Acknowledgements

M.B. is a recipient of the JC Bose fellowship from the Department of Science and Technology - Science and Engineering Research Board (DST-SERB) India. A.G. would like to acknowledge Aditya Kumar for manuscript reading and suggestions. The authors declare no conflict of interest.

References

1. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet 2009;10:94–108.
2. Hung T, Chang HY. Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol 2010;7:582–5.
3. Morris KV, Mattick JS. The rise of regulatory RNA. Nat Rev Genet 2014;15:423–37.
4. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Curr Opin Cell Biol 2009;21:416–25.
5. Brosnan CA, Voinnet O. The long and the short of noncoding RNAs. Curr Opin Cell Biol 2009;21:416–25.
6. Iglesias NG, Gamarnik AV. Dynamic RNA structures in the dengue virus genome. RNA Biol 2011;8:249–57.
24. King HA, Cobbold LC, Willis AE. The role of IRES trans-acting factors in regulating translation initiation. Biochem Soc Trans 2010;38:1581–6.

25. Gritsenko AA, Weingarten-Gabbay S, Elias-Kirma S, et al. Sequence features of viral and human Internal Ribosome Entry Sites predictive of their activity. PLoS Comput Biol 2017;13:e1005734.

26. Kanamori Y, Nakashima N. A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation. RNA 2001;7:266–74.

27. Jan E, Sarnow P. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. J Mol Biol 2002;324:889–902.

28. Kieft JS, Zhou K, Jubin R, et al. The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. J Mol Biol 1999;292:513–29.

29. Boehringer D, Thermann R, Ostareck-Lederer A, et al. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. Structure 2005;13:1695–706.

30. Lukavsky PJ. Structure and function of HCV IRES domains. Virus Res 2009;139:166–71.

31. Lafuente E, Ramos R, Martinez-Salas E. Long-range RNA–RNA interactions between distant regions of the hepatitis C virus internal ribosome entry site element. J Gen Virol 2002;83:1113–21.

32. Pérard J, Leyrat C, Baudin F, et al. Structure of the full-length HCV IRES in solution. Nat Commun 2013;4:1612.

33. Golebiowski J, Antonczak S, Di-Giorgio A, et al. Molecular dynamics simulation of hepatitis C virus IRES IIId domain: structural behavior, electrostatic and energetic analysis. J Mol Model 2004;10:60–8.

34. Branilovic MG, Tomic S. Computational study of the structural plasticity and the ligand binding affinity of the IRES subdomain Ila. Mol Biosyst 2014;10:3272–9.

35. Floden EW, Khawaja A, Vopalensky V, et al. HCVIVdb: the hepatitis-C IRES variation database. BMC Microbiol 2016;16:187.

36. Mokrejš M, Masek T, Vopalensky V, et al. IRESite—a tool for the examination of viral and cellular internal ribosome entry sites. Nucleic Acids Res 2010;38:D131–6.

37. Martinez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, et al. Picornavirus IRES elements: RNA structure and host protein interactions. Virus Res 2015;206:62–73.

38. Wu TY, Hsieh CC, Hong JJ, et al. IRSS: a web-based tool for automatic layout and analysis of IRES secondary structure prediction and searching system in silico. BMC Bioinformatics 2009;10:160.

39. Hong JJ, Wu TY, Chang TY, et al. Viral IRES prediction system—a web server for prediction of the IRES secondary structure in silico. PLoS One 2013;8:e79288.

40. Hofacker IL, Fontana W, Stadler PF, et al. RNA structures. Monatshefte für Chemie/Chemical Monthly 1994;125:167–88.

41. Reeder J, Steffen P, Giegerich R. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. Nucleic Acids Res 2007;35:W240–4.

42. Jiang T, Lin G, Ma B, et al. A general edit distance between RNA structures. J Comput Biol 2002;9:371–88.

43. Kolekar P, Pataskar A, Kulkarni-Kale U, et al. IRESpred: web server for prediction of cellular and viral internal ribosome entry site (IRES). Sci Rep 2016;6:27436.

44. Hofacker IL. Vienna RNA secondary structure server. Nucleic Acids Res 2003;31:3429–31.

45. Muppipula UK, Honavar VG, Dobbs D. Predicting RNA–protein interactions using only sequence information. BMC Bioinformatics 2011;12:489.

46. Jacks T, Madhani HD, Masiarz FR, et al. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell 1988;55:447–58.

47. Maia IG, Seron K, Haenni AL, et al. Gene expression from viral RNA genomes. Plant Mol Biol 1996;32:367–91.
RNA-mediated translation regulation
87. Yueh A, Schneider RJ. Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. Genes Dev 2000;14:414–21.
88. Pooggin MM, Rajeswaran R, Schepetilnikov MV, et al. Short ORF-dependent ribosome shunting operates in an RNA picorna-like virus and a DNA pararetrovirus that cause rice tungro disease. PLoS Pathog 2012;8:e1002568.
89. Zuker M. Mold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003;31:3406–15.
90. Mathews DH, Sabina J, Zuker M, et al. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 1999;288:911–40.
91. Brown A, Shao S, Murray J, et al. Structural basis for stop codon recognition in eukaryotes. Nature 2015;524:493–9.
92. Bertram G, Innes S, Minella O, et al. Endless possibilities: translation termination and stop codon recognition. Microbiology 2001;147:255–69.
93. McCaughan KK, Brown CM, Dalphin ME, et al. Translational termination efficiency in mammals is influenced by the base following the stop codon. Proc Natl Acad Sci USA 1995;92:5431–5.
94. Brault V, van den Heuvel JF, Verbeek M, et al. Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. EMBO J 1995;14:650–9.
95. Adams MJ, Antoniw JF, Mullins JC. Plant virus transmission by plasmopodisorphid fungi is associated with distinctive transmembrane regions of virus-encoded proteins. Arch Virol 2001;146:1139–53.
96. Beier H, Grimm M. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res 2001;29:4767–82.
97. Tork S, Hatin I, Rouset JP, et al. The major 5′ determinant in stop codon read-through involves two adjacent adenines. Nucleic Acids Res 2004;32:415–21.
98. Firth AE, Brierley I. Non-canonical translation in RNA viruses. J Gen Virol 2012;93:1385–409.
99. Skuzeski JM, Nichols LM, Gesteland RF, et al. The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. J Mol Biol 1991;218:365–73.
100. Firth AE, Wills NM, Gesteland RF, et al. Stimulation of stop codon readthrough: frequent presence of an extended 3′ RNA structural element. Nucleic Acids Res 2011;39:6679–91.
101. Alam SL, Wills NM, Ingram JA, et al. Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukemia virus. J Mol Biol 1999;288:837–52.
102. Xu Y, Ju HJ, DeBlasio S, et al. A stem–loop structure in potato leafroll virus open reading frame 5 (ORF5) is essential for readthrough translation of the coat protein ORF stop codon 700 bases upstream. J Virol 2018;92:e01544–17.
103. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol 1990;215:403–10.
104. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet 2000;16:276–7.
105. Naphthine S, Yek C, Powell ML, et al. Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. RNA 2012;18:241–52.
106. Wills NM, Gesteland RF, Atkins JF. Pseudoknot-dependent read-through of retroviral gag termination codons: importance of sequences in the spacer and loop 2. EMBO J 1994;13:4137–44.
107. Wills NM, Gesteland RF, Atkins JF. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. Proc Natl Acad Sci U S A 1991;88:6991–5.
108. Curran J, Kolakofsky D. Ribosomal initiation from an AGC codon in the Sendai virus P/C mRNA. EMBO J 1988;7:245–51.
109. Prats AC, De Billy G, Wang P, et al. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. J Mol Biol 1989;205:363–72.
110. Touriol C, Bornes S, Bonnal S, et al. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biol Cell 2003;95:169–78.
111. Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl Acad Sci USA 1990;87:3301–5.
112. Lorenz R, Wolfinger MT, Tanzer A, et al. Predicting RNA secondary structures from sequence and probing data. Methods 2016;103:86–98.
113. Clyde K, Harris E. RNA secondary structure in the coding region of dengue virus type 2 directs translation start codon selection and is required for viral replication. J Virol 2006;80:2170–82.
114. Jackson RJ, Hellen CU, Pestova TV. Termination and post-termination events in eukaryotic translation. Adv Protein Chem Struct Biol 2012;86:45–93.
115. Powell ML. Translational termination–reinitiation in RNA viruses. Biochem Soc Trans 2010;38:1558–64.
116. Guo LH, Sun L, Chiba S, et al. Coupled termination/reinitiation for translation of the downstream open reading frame B of the prototypic hypovirus CHV1-EF713. Nucleic Acids Res 2009;37:3645–59.
117. Meyers G. Characterization of the sequence element directing translation reinitiation in RNA of the calcivirus rabbit hemorrhagic disease virus. J Virol 2007;81:9623–32.
118. Luttermann C, Meyers G. A bipartite sequence motif induces translation reinitiation in feline calcivirus RNA. J Biol Chem 2007;282:7056–65.
119. Simon AE, Miller WA. 3′ cap-independent translation enhancers of plant viruses. Annu Rev Microbiol 2013;67:21–42.
120. Truniger V, Miras M, Aranda MA. Structural and functional diversity of plant virus 3′-cap-independent translation enhancers (3′-CITEs). Front Plant Sci 2017;8:2047.
121. Biesiada M, Purzycka KJ, Szachniuk M, et al. Automated RNA 3D structure prediction with RNAComposer. Methods Mol Biol 2016;1490:199–215.
122. Vieweger M, Holmstrom ED, Nesbitt DJ. Single-molecule FRET reveals three conformations for the TLS domain of brome mosaic virus genome. Biophys J 2015;109:2625–36.
123. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. Gene 2002;299:1–34.
124. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 1986;44:283–92.
125. Kozak M. A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. Gene Expr 1991;1:111–5.
126. Williams MA, Lamb RA. Effect of mutations and deletions in a bicistronic mRNA on the synthesis of influenza B virus NB and NA glycoproteins. J Virol 1989;63:28–35.
127. Dinesh-Kumar SP, Miller WA. Control of start codon choice on a plant viral RNA encoding overlapping genes. Plant Cell 1993;5:679–92.

128. Zhang S, Hu H, Jiang T, et al. TITER: predicting translation initiation sites by deep learning. Bioinformatics 2017;33:i234–42.

129. Lozano G, Fernandez N, Martinez-Salas E. Modeling three-dimensional structural motifs of viral IRES. J Mol Biol 2016;428:767–76.

130. Witwer C, Rauscher S, Hofacker IL, et al. Conserved RNA secondary structures in Picornaviridae genomes. Nucleic Acids Res 2001;29:5079–89.

131. Wilkinson KA, Fernandez N, Martinez-Salas E. Modeling three-dimensional structural motifs of viral IRES. J Mol Biol 2016;428:767–76.

132. Witwer C, Rauscher S, Hofacker IL, et al. Conserved RNA secondary structures in Picornaviridae genomes. Nucleic Acids Res 2001;29:5079–89.

133. Lozano G, Fernandez N, Martinez-Salas E. Modeling three-dimensional structural motifs of viral IRES. J Mol Biol 2016;428:767–76.

134. Witwer C, Rauscher S, Hofacker IL, et al. Conserved RNA secondary structures in Picornaviridae genomes. Nucleic Acids Res 2001;29:5079–89.

135. Wilkinson KA, Fernandez N, Martinez-Salas E. Modeling three-dimensional structural motifs of viral IRES. J Mol Biol 2016;428:767–76.

136. Witwer C, Rauscher S, Hofacker IL, et al. Conserved RNA secondary structures in Picornaviridae genomes. Nucleic Acids Res 2001;29:5079–89.

137. Wilkins KA, Merino EJ, Weeks KM. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc 2006;1:1610–6.

138. Kenyon JC, Prestwood LJ, Le Grice SF, et al. In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. Nucleic Acids Res 2013;11:959–65.

139. Ciuffi A, Rato S, Telenti A. Single-cell genomics for virology. Viruses 2016;8:123.

140. Tang Y, Bouvier E, Kwok CK, et al. StructureFold: genome-wide RNA secondary structure mapping and reconstruction in vivo. Bioinformatics 2015;31:2668–75.

141. Choudhary K, Lai YH, Tran EJ, et al. dStruct: identifying differentially reactive regions from RNA structure probing data. Genome Biol 2019;20:40.

142. Ge P, Zhang S. Computational analysis of RNA structures with chemical probing data. Methods 2015;79-80:60–6.

143. Incarnato D, Neri F, Anselmi F, et al. RNA structure framework: automated transcriptome-wide reconstruction of RNA secondary structures from high-throughput structure probing data. Bioinformatics 2016;32:459–61.

144. Gurtu V, Yan G, Zhang G. IRES bicistronic expression vectors for efficient creation of stable mammalian cell lines. Biochem Biophys Res Commun 1996;229:295–8.

145. Bouabe H, Fassler R, Heesemann J. Improvement of reporter activity by IRES-mediated polycistronic reporter system. Nucleic Acids Res 2008;36:628.

146. Dotu I, Lozano G, Clote P, et al. Using RNA inverse folding to identify IRES-like structural subdomains. RNA Biol 2013;10:1842–52.

147. Garcia-Martín JA, Clote P, Dotu I. RNAiFOLD: a constraint programming algorithm for RNA inverse folding and molecular design. J Bioinform Comput Biol 2013;11:1350001.

148. Busch A, Backofen R. INFO-RNA—a fast approach to inverse RNA folding. Bioinformatics 2006;22:1823–31.

149. Churkin A, Retwitzer MD, Reinharz V, et al. Design of RNAs: comparing programs for inverse RNA folding. Brief Bioinform 2018;19:350–8.

150. Tuplin A, Evans DJ, Simmonds P. Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. J Gen Virol 2004;85:3037–47.

151. Tuplin A, Wood J, Evans DJ, et al. Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. RNA 2002;8:824–41.