Check for updates

Know Your Enemy: Successful Bioinformatic Approaches to Predict Functional RNA Structures in Viral RNAs

Chun Shen Lim and Chris M. Brown*

Department of Biochemistry, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand

Structured RNA elements may control virus replication, transcription and translation, and their distinct features are being exploited by novel antiviral strategies. Viral RNA elements continue to be discovered using combinations of experimental and computational analyses. However, the wealth of sequence data, notably from deep viral RNA sequencing, viromes, and metagenomes, necessitates computational approaches being used as an essential discovery tool. In this review, we describe practical approaches being used to discover functional RNA elements in viral genomes. In addition to success stories in new and emerging viruses, these approaches have revealed some surprising new features of well-studied viruses e.g., human immunodeficiency virus, hepatitis C virus, influenza, and dengue viruses. Some notable discoveries were facilitated by new comparative analyses of diverse viral genome alignments. Importantly, comparative approaches for finding RNA elements embedded in coding and non-coding regions differ. With the exponential growth of computer power we have progressed from stem-loop prediction on single sequences to cutting edge 3D prediction, and from command line to user friendly web interfaces. Despite these advances, many powerful, user friendly prediction tools and resources are underutilized by the virology community.

Keywords: bioinformatics, *cis*-regulatory elements, comparative genomics, non-coding RNAs, pseudoknots, RNA structure prediction, RNA viruses, structural motifs

INTRODUCTION

This review illustrates the key concepts and strategies used for prediction of RNA structural elements in RNA viral sequences. A range of RNA structure prediction software and relevant resources are available, but most are underutilized by virologists. Here the concepts and strength of these methods are introduced using examples of successful approaches in viruses, with the intention of bridging the gap. The roles of RNA elements in viral biology is illustrated using well-studied viruses, flaviviruses, influenza, and barley yellow dwarf virus (BYDV). We further review the structures and functions of well-characterized types of RNA elements with the emphasis on prediction approaches and their limitations.

There have been several excellent recent reviews on generally predicting RNA structures, in particular relating to integrating experimental data and on 3D predictions (Cantara et al., 2014; Achar and Sætrom, 2015; Weeks, 2015; Dawson and Bujnicki, 2016; Lorenz et al., 2016; Turner and Mathews, 2016). RNA 3D structure prediction methodology and incorporation of experimental

constraint is beyond the scope of this review, but we include examples where they have been utilized.

Concepts of RNA Structure Prediction

Stems involving G-C, A-U, and G-U canonical Watson-Crick base-pairs are the basis of most viral RNA structures, indeed the stem-loop is the basic building block (**Table 1**, **Figure 1**). These stems usually form an A-form helix structure, as the 2′-hydroxyl prevents the B-form helix found in DNA. RNA sections with unpaired bases may form structures such as loops or bulges (Bindewald et al., 2008; **Table 1**, **Figure 1**). An RNA secondary structure is more likely to be functional if it (i) has a low minimum free energy (MFE) that enables it to fold and base-pair, and/or (ii) is conserved during evolution with covarying stem base-pairs (compensatory base-pair changes). This RNA structural conservation is based on the concept that RNA stems can be conserved regardless of the base-pairs used (Akiyama et al., 2016; Rivas et al., 2017). Both of these features can form the basis of predicting new RNA elements (Xu and Mathews, 2016; Taylor and Hamilton, 2017) and can be integrated with experimental data. Notably, in addition to the fold with the lowest free energy, MFE suboptimal predictions are particularly useful in assessing possible alternative structures of RNA (e.g., pseudoknots; Theis et al., 2008) and long-distance base-pairs (Fricke and Marz, 2016; long-range interactions; **Table 1**).

The limits of current methodology means stems are usually predicted initially using only the canonical base-pairs. However, many non-canonical base-pairs and other structural elements are found in experimentally determined RNA structures (**Table 1**, **Figure 1**). About 40% of bases in known crystal and solution structures were either unpaired, or form non-canonical interactions (Stombaugh et al., 2009). Some of the more common non-Watson-Crick pairs in the RNA Basepair Catalog are U-U (about 10% as frequent as A-U pairs, 432 of 4,200) and A-G (about 2% as frequent as C-G, 191 of 9,316; Stombaugh et al., 2009). For example, (i) the base-triple in retroviral encapsidation signals (D'souza et al., 2004) and the base-triples in the pseudoknots of Beet western yellows virus (Su et al., 1999) and Sugarcane yellow leaf virus (Cornish et al., 2005; **Figure 1A**, and (ii) the kink-turn/k-turn (**Figure 1A**, **Table 1**) in the panhandle RNA structure of Influenza A virus that is inducible (Lee et al., 2016) and the A-minor k-turn in the encapsidation signal of Moloney murine leukemia virus (Miyazaki et al., 2010; **Table 1**, **Figure 1**).

RNA 3D structures can also be predicted directly from sequences. The accuracy of these prediction tools has improved in the past few years (Miao and Westhof, 2017), such as RNAComposer (Popenda et al., 2012; Antczak et al., 2016), 3dRNA (Zhao et al., 2012; Wang et al., 2017) and SimRNAweb (Magnus et al., 2016). Notably, SimRNAweb has accurately predicted a previously solved frameshifting RNA pseudoknot from beet western yellow virus (Egli et al., 2002).

These predictions can be tested experimentally. For example, to demonstrate that a predicted RNA structure exists and is functionally important, a wild type phenotype can be destroyed with mutations that disrupt the RNA structure (e.g., Fang et al., 2012; Chapman et al., 2014b). This phenotype may be restored by compensatory base-pair changes—changing the primary sequence where base-pairing is still allowed. However, primary sequence motifs and structures of loops and bulges may also have important roles (Bindewald et al., 2008).

Representations of RNA Structures

The conventional representation of an RNA structure is the 2D stem-loop diagram (**Figure 2**). However, the stem-loop diagram is not suitable to represent higher order interactions such as pseudoknotted interactions (**Figure 2C**, **Table 1**). In contrast, these tertiary interactions represented by the dot-bracket notation (Hofacker et al., 1994), and circular (Nussinov et al., 1978) and arc (Wattenberg, 2002) diagrams are easier to interpret (e.g., **Figures 2A,B,D**, respectively). These diagrams can be generated using VARNA, which requires dot-bracket notation as the input (Darty et al., 2009). Arc diagrams can also be created using R-CHIE, which is available as R package and web service (Lai et al., 2012).

However, these diagrams and dot-bracket notation normally represent one sequence at a time. Often common RNA elements are found in related sequences e.g., viral genotypes. Stockholm format is commonly used to represent the consensus RNA secondary structure of aligned sequences (**Figure 2E**; https://en.wikipedia.org/wiki/Stockholm_format). Stockholm format also stores some metadata e.g., the description of the aligned RNA sequences. R2R uses Stockholm file as the input to generate a novel stem-loop diagram of the consensus RNA secondary structure annotated with sequence conservation and covarying base-pairs (**Figure 2F**). Stockholm format and R2R stem-loop diagrams are both used by the Rfam database (see section on "Sources of Known RNA Structures"). Stockholm format file editors are available (Griffiths-Jones, 2005; Waterhouse et al., 2009).

Bioinformatic Tools

Many RNA structure prediction tools were initially released as command line software (Zuker, 1989). Biologists and virologists who are interested in using these would first learn the command line interface, this was and is a barrier for many researchers. However, where possible significant efforts have been made by developers to make their tools more readily available, as webservers (Backofen et al., 2017; Fallmann et al., 2017) or integrated graphic user interfaces (e.g., RNAStructure, or the Simple Sequence Editor, SSE; Simmonds, 2012; Bellaousov et al., 2013; Wang et al., 2017). For example, mfold which is the most cited RNA software in virology papers, was first released as a command line software in the late 1980s and became available as web interfaces in early 2000s (Zuker, 2003).

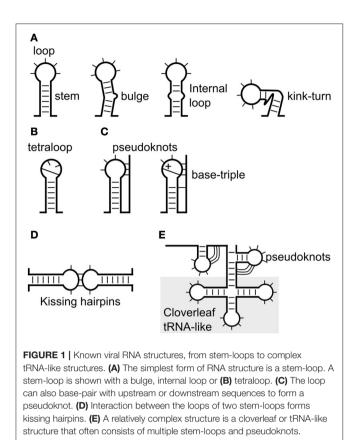In contrast to folding one sequence at a time [single-sequence methods e.g., mfold/UNAfold (Zuker, 2003), RNAfold (Gruber et al., 2008)], a new generation of software such as LoCARNA (Smith et al., 2010) and RNAz (Gruber et al., 2007) work on multiple sequences (comparative methods). This alleviates the need of predicting RNA structures from related virus sequences one at a time and comparing them manually. Different methods

TABLE 1 | Structural RNA elements, the most used prediction tools, and challenges for their prediction.

RNA Structures	Most used prediction tools	Challenges
Stem-loop/hairpin. The helical stem consists of base-pairs. The loop consists of unpaired or non-canonical base-pairs (Zhang et al., 2011).	mfold/UNAFold (Zuker, 2003), RNAfold (Gruber et al., 2008; Lorenz et al., 2011), RNAStructure/Fold (Bellaousov et al., 2013).	Predictions normally only consider standard or canonical base-pairs C-G, U-A, and U-G. Single base-pairs ("lone pairs") are often excluded by default. Functionally important alternative structures depending on ligand binding need special consideration (e.g., riboswitches).
Bulge. A region of a helix where there are no canonical base-pairs at one strand (Zhang et al., 2011).	mfold/UNAFold, RNAfold, RNAStructure/Fold.	Predicted bulges may use non-canonical base-pairs e.g., U-U, A-G, kink-turn.
Internal loop. A region of a helix where there are no canonical base-pairs at both strands (Zhang et al., 2011).	mfold/UNAFold, RNAfold, RNAStructure/Fold.	Predicted internal loops may use non-canonical base-pairs e.g., U-U, A-G.
Tetraloop. A four-base terminal loop stabilized by intra-loop hydrogen bonds. This stabilizes the stem-loop structure. The GNRA loop is most common, where N represents any base and R represents either A or G (Zhang et al., 2011).	mfold/UNAfold, RNAfold (free energy bonus). RNAComposer (Popenda et al., 2012), 3dRNA (Zhao et al., 2012) (RNA 3D motif prediction).	Most 2D predictions do not predict the intraloop pair (e.g., the G-A pair of the GNRA loop). 3D predictions may predict them. Should be considered if a terminal four-base loop is predicted. Other types of loops e.g., tri-loop and anticodon like loops, can also be stabilized.
Pseudoknot. Bipartate structure in which the loop of one stem-loop base-pairs with a sequence outside of the stem-loop (Zhang et al., 2011).	PknotsRG (Janssen and Giegerich, 2015), DotKnot (Sperschneider and Datta, 2010), Pknots (Rivas and Eddy, 1999).	Not predicted by most 2D software. Alternative forms of pseudoknot are found.
Kink-turn/k-turn. A three nucleotide bulge in a helix followed by G-A and A-G pairs. Bends the helix (Petrov et al., 2013).	RNAComposer, 3dRNA.	Widespread but most software will not predict these due to non-canonical base-pairs. Requires 3D or homology based software which are yet to be integrated into the most used RNA structure prediction tools.
Junction. The point of connection between a number of different helices (Lilley et al., 1995; Bindewald et al., 2008; Zhang et al., 2011).	mfold/UNAFold, RNAfold, RNAStructure/Fold.	Junctions may be important for ligand binding but 3D structures are difficult to predict.
Base-triple. A group of three bases which interact by hydrogen-bonds that include edge-edge bonds.	RNAComposer, 3dRNA.	Common in 3D structures.
Kissing hairpins/kissing loops/kissing stem-loops. Base-pair interactions between the loops of two stem-loops (Zhang et al., 2011).	pAliKiss (Janssen and Giegerich, 2015).	Difficult to predict without prior knowledge.
tRNA-like or cloverleaf structures. Structures with a tRNA-like tertiary structure. In viruses pseudoknots are often located nearby.	Combination of stem-loop and pseudoknot prediction tools.	No specialized tools available to date.
Long-range intra-molecular interactions. Base-pair interactions over long distances. Arbitrarily defined as base-pairs over 100 bases apart.	mfold/UNAFold, LRIscan (Fricke and Marz, 2016), CovaRNA (Bindewald and Shapiro, 2013).	Difficult to predict without prior knowledge. Only two specialized tools available to date—CovaRNA and LRIscan. Only LRIscan is optimized for viral genomes and yet to be proven useful.
Inter-molecular interactions. Base-pair interactions between two RNA molecules e.g., two copies of a RNA genome.	RNAhybrid (Rehmsmeier et al., 2004), RNAaliduplex (Gruber et al., 2008; Lorenz et al., 2011), bifold (Mathews et al., 1999).	Difficult to predict without prior knowledge.

vary in whether they align or fold first or do both simultaneously (Gardner and Giegerich, 2004).

The current range of functional RNA structures and prediction tools may seem intimidating (http://en.wikipedia.org/wiki/List_of_RNA_structure_prediction_software). However, many RNA structure prediction tools and RNA-RNA interaction prediction tools have been compared for use in different applications (Gardner and Giegerich, 2004; Gardner et al., 2005; Puton et al., 2013; Umu and Gardner, 2016). In general, comparative methods are more accurate than the older single-sequence methods (Puton et al., 2013).

Many of these powerful applications have been underused by virologists. For example, Infernal (INFERence of RNA Alignment; Nawrocki and Eddy, 2013) and CMfinder (Yao et al., 2006) that are based on both sequence and RNA secondary structure conservation allow sensitive detection of homologous

FIGURE 1 | Known viral RNA structures, from stem-loops to complex tRNA-like structures. **(A)** The simplest form of RNA structure is a stem-loop. A stem-loop is shown with a bulge, internal loop or **(B)** tetraloop. **(C)** The loop can also base-pair with upstream or downstream sequences to form a pseudoknot. **(D)** Interaction between the loops of two stem-loops forms kissing hairpins. **(E)** A relatively complex structure is a cloverleaf or tRNA-like structure that often consists of multiple stem-loops and pseudoknots.

RNA structures. A list of software that has been cited in selected virology publications is available (http://bioanalysis.otago.ac.nz/Lim2017.htm). Notably, these are beginning to include newer webservers which predict RNA 2D and 3D structures with high confidence.

Current methods often provide a 2D and oversimplified view of a certain sequence forming a single RNA structure. This is incorrect particularly in viral RNAs, where structures need to be transiently formed and melted (Moss et al., 2012a; Zhu and Meyer, 2015). This one-to-one sequence and structure relationship is also not true in many RNA viruses because they may exist in a quasispecies state where sequence space is sampled by high levels of replication error (Holmes, 2010; Lauring and Andino, 2010; Marz et al., 2014). Conservation in RNA structures but not the primary sequences across rapidly evolving species being particularly striking, e.g., the HIV frameshift site is one of the most conserved parts of the genome (Mathew et al., 2015).

KNOW YOUR ENEMY

The starting point for RNA structure analysis is likely to be a complete (or partial) RNA genome (**Figure 3**). This could be a well-studied virus, or come from an outbreak of a new or emerging virus (e.g., SARS or Zika). The aim of these analyses is to further understand the biology of the viruses, and also to identify drug or vaccine targets.

Viral RNA elements have been identified as antiviral targets due to conservation of sequence and functions that are distinguishable from the host (Panjaworayan and Brown, 2011; Chen et al., 2014; Cardno et al., 2015; Le Grice, 2015; Hermann, 2016; Hilimire et al., 2017). For example, the internal ribosome entry sites (IRES) of hepatitis C virus (HCV) is targeted by benzimidazole (Dibrov et al., 2012). Another example is the HIV frameshift site, that has characteristics distinct from human frameshift sites (Cardno et al., 2015; Mathew et al., 2015; Hermann, 2018). In addition, double stranded viral RNA structures could be targeted by the host innate immune response, e.g., through Protein kinase RNA-activated (PKR) (Hartmann, 2017).
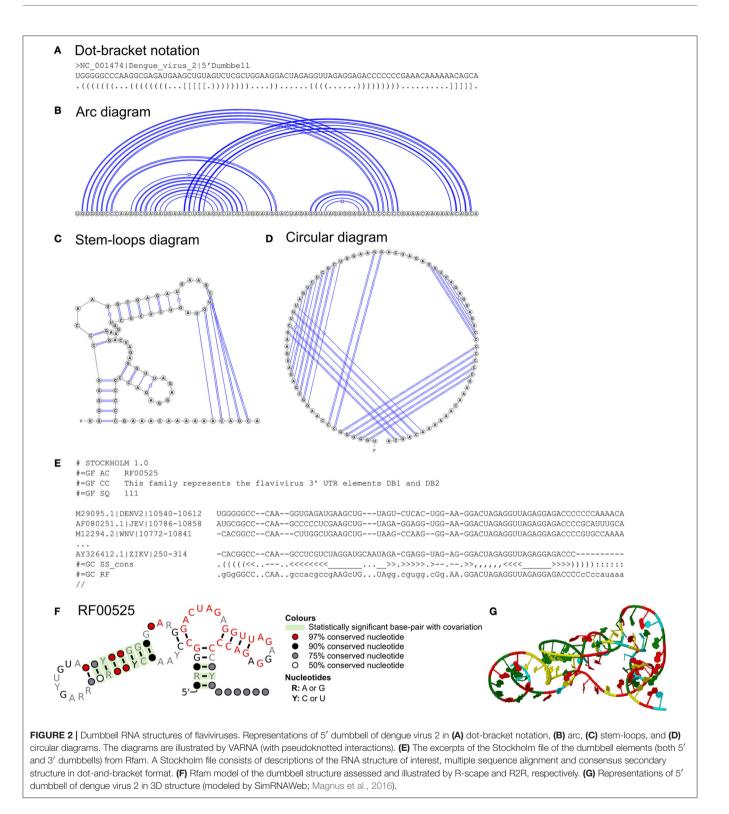
Targeting these viral specific features requires understanding of both the viral genomic sequence and its functional and sequence variation constraints—including gene structure and RNA *cis*-elements (Newburn and White, 2015; Hermann, 2018).

As a first step a similarity search on the NCBI/RefSeq database may not only allow identification of the virus, but also identify related viral sequences that could assist in predicting functional elements (**Figure 3**). Deep and accurate multiple sequence alignment is crucial in predicting likely RNA structures (Backofen et al., 2017; Fallmann et al., 2017). Specialized databases may also provide high quality sequence alignments to researchers, such as the LANL sequence databases for HIV, HCV, and hemorrhagic fever viruses (e.g., Ebola; Kuiken et al., 2012; Hatcher et al., 2017).

A novel virus can be classified according to the International Committee on Taxonomy of Viruses (ICTV) (King et al., 2011). The viral biology can be inferred if its species is well-characterized using published literature, and facilitated by general databases e.g., the ViralZone knowledgebase (Hulo et al., 2011) and specialized parts of the sequence databases e.g., NCBI Virus Variation Resource (Hatcher et al., 2017), RefSeq (O'leary et al., 2016), and NCBI Viral Genomes Resource (Brister et al., 2015). ICTV and ViralZone are further discussed in the next section "Virus Biology and RNA Structures."

In conjunction with RNA structure analysis, potential coding sequences (CDS) can be predicted, in at least the three forward reading frames. This is an important step prior to prediction of RNA structures located in the coding sequence (Liu et al., 2009; Firth, 2014), for example frameshifting elements (Giedroc and Cornish, 2009). The beginnings and ends of these potential CDS are hotspots for RNA structures (Newburn and White, 2015).
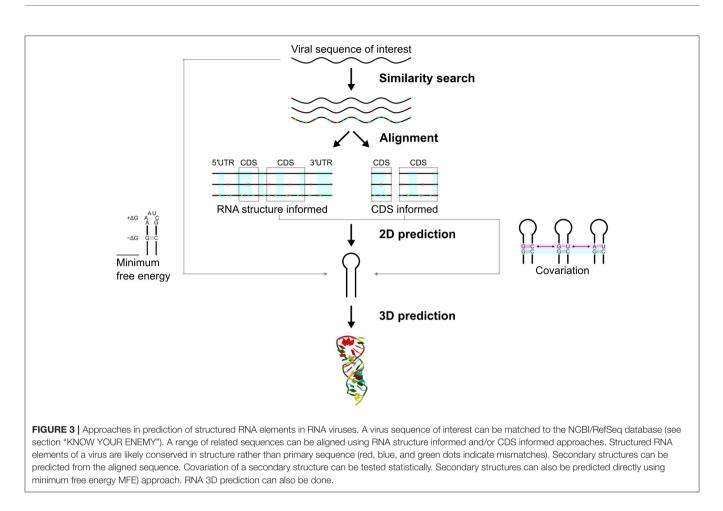
If possible, alignments should be made to assist in identifying likely CDS and RNA structures (Firth, 2014; **Figure 3**). Similar sequences may be found with blastn, although non-coding similarity may be missed unless the initial hit size (word size) is reduced from the default of 11 to the greatest sensitivity available: 7. Alternatively, more sensitive local similarity search programs based on Smith-Waterman algorithm such as FASTA (SSEARCH; Lipman and Pearson, 1985; Pearson and Lipman, 1988) and SWIPE (Rognes, 2011) may be used, but are slower than blastn. FASTA is available through EMBL-EBI tools (https://www.ebi.ac.uk/Tools; Mcwilliam et al., 2013). Creating alignments for detection of elements within CDS can be facilitated by searching with the encoded protein (e.g., tblastn

FIGURE 2 | Dumbbell RNA structures of flaviviruses. Representations of 5′ dumbbell of dengue virus 2 in **(A)** dot-bracket notation, **(B)** arc, **(C)** stem-loops, and **(D)** circular diagrams. The diagrams are illustrated by VARNA (with pseudoknotted interactions). **(E)** The excerpts of the Stockholm file of the dumbbell elements (both 5′ and 3′ dumbbells) from Rfam. A Stockholm file consists of descriptions of the RNA structure of interest, multiple sequence alignment and consensus secondary structure in dot-and-bracket format. **(F)** Rfam model of the dumbbell structure assessed and illustrated by R-scape and R2R, respectively. **(G)** Representations of 5′ dumbbell of dengue virus 2 in 3D structure (modeled by SimRNAWeb; Magnus et al., 2016).

and tblastx). This will give greater sensitivity than blastn searches.

If it is known that the RNA regions encode for proteins (CDS) and/or contain RNA structures, alignment algorithms that consider this should be used [e.g., webPRANK (Löytynoja and Goldman, 2010) or R-Coffee (Taly et al., 2011), respectively;

Figure 3]. RNA structures can also be detected in unaligned sequences, although these methods are more computationally intensive. Ideally, RNA primary sequence alignments should have dissimilarity of about 5–20% (Theis et al., 2015). Near identical aligned sequences may lack complexity that allows accurate RNA structure prediction and are not usually included in the

FIGURE 3 | Approaches in prediction of structured RNA elements in RNA viruses. A virus sequence of interest can be matched to the NCBI/RefSeq database (see section "KNOW YOUR ENEMY"). A range of related sequences can be aligned using RNA structure informed and/or CDS informed approaches. Structured RNA elements of a virus are likely conserved in structure rather than primary sequence (red, blue, and green dots indicate mismatches). Secondary structures can be predicted from the aligned sequence. Covariation of a secondary structure can be tested statistically. Secondary structures can also be predicted directly using minimum free energy MFE) approach. RNA 3D prediction can also be done.

prediction phase (see the success story on "RNA Structures in Coding Regions of Influenza A Virus"). However, the phenotype of a viable virus with a mutation in the structure may be informative (Kobayashi et al., 2016).
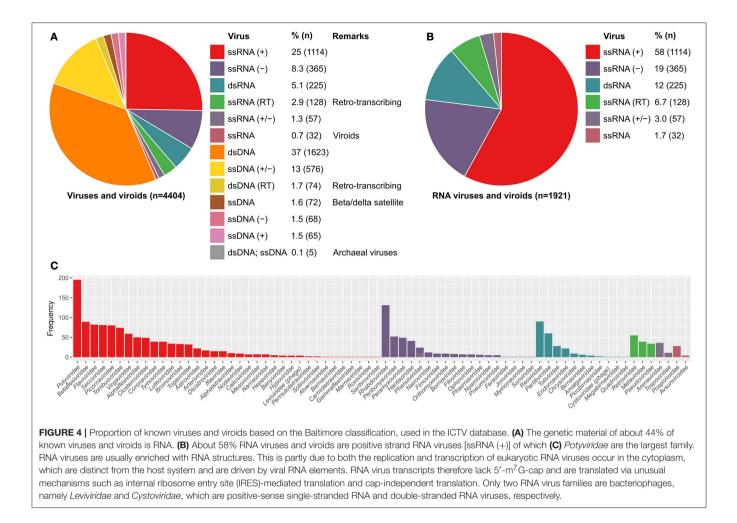
Virus Biology and RNA Structures

Most RNA structures play *cis*-regulatory roles in various stages of the virus life cycle. Therefore, the functions of RNA structures can partly be inferred from their locations (Newburn and White, 2015). The RNA structures located near the 5′ end are mostly involved in replication and initiation of translation, such as the dimer linkage structure (DLS) of retroviruses (Johnson and Telesnitsky, 2010) and IRES of *Picornaviridae*, *Flaviviridae* in particular HCV and *Discistroviridae*, respectively (see section on "Internal Ribosome Entry Sites (IRES)"; Lee et al., 2017). Overlapping CDS may indicate frameshifts which would then direct the search to specific primary features, and nearby stem-loops or pseudoknots (Miras et al., 2017; see sections on "KNOW YOUR ENEMY" and "pseudoknots"). Whereas, RNA structures located near the 3′ end are often important in nuclear export of viral RNAs, such as the Rev response element (RRE) of human immunodeficiency virus (HIV) (Groom et al., 2009) and in replication, processing, or RNA stability (Newburn and White, 2015). However, other elements e.g., *cis*-acting replication elements (CRE) can be found in various genomic locations.

For example, it is located at the 3′ end of HCV but the CDS of poliovirus (Tuplin et al., 2002; Dutkiewicz et al., 2016). Structured RNA elements in different locations of many viral genomes were reviewed in detail by Romero-López and Berzal-Herranz (2013); Brinton and Basu (2015); Newburn and White (2015); Nicholson and White (2015); Sagan et al. (2015); Madhugiri et al. (2016) and Fernández-Sanlés et al. (2017). For specific example of the functions and locations of RNA structures, see section on "RNA Structures in Barley Yellow Dwarf Virus (BYDV)."

Some guide to what structures to look for can also be obtained from the classification and biology of the virus of interest. ViralZone provides up-to-date information about viral biology, but it is protein and virus centered, rather than RNA structure focused (Hulo et al., 2011). As of June 2017, it documents the biology of 110 viral families, based on literature review, each entry is linked to Uniprot viral proteins. In ViralZone, summaries have been made under the section "Viral molecular biology: Transcription, replication, translation" (http://viralzone.expasy. org/915). This allows us to infer the viral molecular biology, which in turns provides some clues of what structural RNA elements to search for.

Currently, the ICTV master species list (2016 v1.3) has the taxonomic classification of 4404 viruses and viroids, 44% of these are RNA viruses (**Figures 4A,B**). There are a total of 73

FIGURE 4 | Proportion of known viruses and viroids based on the Baltimore classification, used in the ICTV database. **(A)** The genetic material of about 44% of known viruses and viroids is RNA. **(B)** About 58% RNA viruses and viroids are positive strand RNA viruses [ssRNA (+)] of which **(C)** *Potyviridae* are the largest family. RNA viruses are usually enriched with RNA structures. This is partly due to both the replication and transcription of eukaryotic RNA viruses occur in the cytoplasm, which are distinct from the host system and are driven by viral RNA elements. RNA virus transcripts therefore lack 5'-m^{7}G-cap and are translated via unusual mechanisms such as internal ribosome entry site (IRES)-mediated translation and cap-independent translation. Only two RNA virus families are bacteriophages, namely *Leviviridae* and *Cystoviridae*, which are positive-sense single-stranded RNA and double-stranded RNA viruses, respectively.

RNA virus families. Notably, over half (58%) of the RNA viruses are positive-sense single-stranded RNA viruses. RNA viruses are often enriched with RNA structures. This is partly due to the replication and transcription of RNA viruses occurring in the cytoplasm, which are regulated by viral RNA elements. The genomes and transcripts of some RNA viruses lack the 5-m^{7}G (cap) requiring cap-independent translation (Simon and Miller, 2013). Indeed, some RNA viruses (e.g., picornaviruses) shutoff the host mRNA translation and use cap-independent translation such as IRES-mediated translation (Chase and Semler, 2012).
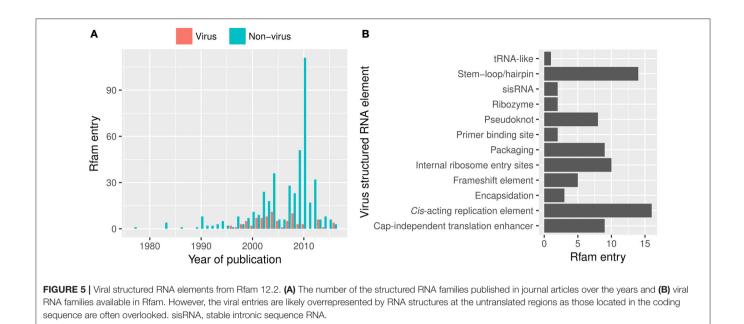
In contrast, over 99% of bacterial and archaeal viruses (bacteriophages) are DNA viruses (**Figure 4A**; ICTV master species list 2016 v1.3), although these may use RNA structures in their life cycles, notably as regulatory switches (Walsh and Mohr, 2011; Yang et al., 2014) and may have structured ncRNA (Hill et al., 2016). Only two RNA virus families infect bacteria, namely *Leviviridae* and *Cystoviridae*, which are positive-sense single-stranded RNA and double-stranded RNA viruses, respectively (**Figure 4C**). Several RNA bacteriophages are well-characterized such as MS2, Q, F1, and phi6. In particular, the 19-nucleotide MS2 packaging signal stem-loop of *E. coli* MS2 phage has been extensively studied. This high affinity MS2 packaging signal stem-loop is located at the ribosomal binding site of the replicase mRNA. Translation is inhibited upon the strong and specific

binding of MS2 capsid protein (Peabody, 1990; Lim and Peabody, 1994; Stockley et al., 1995; Johansson et al., 1998). Recent studies indicate that other RNA viral genomes may have multiple structured capsid protein binding sites (Patel et al., 2017). The properties of MS2 have been exploited for various novel applications such as pull-down, tethering proteins to RNAs, RNA affinity purification, and live cell imaging of RNAs and protein-RNA interactions (Bardwell and Wickens, 1990; Bertrand et al., 1998; Coller et al., 1998; Graveley and Maniatis, 1998; Rackham and Brown, 2004).

Sources of Known RNA Structures

An example of useful resource that is not frequently cited by virus research articles is Rfam, the database of RNA structure families (Nawrocki et al., 2014). It contains over 105 viral RNA structural elements from both DNA and RNA viruses (Rfam 12.2, release January 2017; **Figure 5**). The most common viral RNA elements in Rfam are those in 3'UTRs (e.g., CITEs), 5'UTR (e.g., IRES), and packaging elements [e.g., packaging elements ($n = 8$) and *cis*-replication elements ($n = 17$, CRE) or encapsidation elements].

Importantly, the Rfam database can be used to annotate a viral sequence by searching for known RNA families with simple online tools (Nawrocki et al., 2014). Alternatively, the

FIGURE 5 | Viral structured RNA elements from Rfam 12.2. **(A)** The number of the structured RNA families published in journal articles over the years and **(B)** viral RNA families available in Rfam. However, the viral entries are likely overrepresented by RNA structures at the untranslated regions as those located in the coding sequence are often overlooked. sisRNA, stable intronic sequence RNA.

roles of novel RNA structures can be inferred by comparing them to the existing RNA families (Eggenhofer et al., 2013). Once characterized researchers can submit new RNA families to Rfam. Automated resources and guides for building families are available (Eggenhofer et al., 2016). Building these models may be facilitated by using combinations of software (Chen X. S. and Brown, 2012) in particular the WAR webserver (Torarinsson and Lindgreen, 2008) then be published as family descriptions online and/or in RNA Biology (Gultyaev and Olsthoorn, 2010; Chen A. and Brown, 2012; Lim and Brown, 2016).

Experimentally determined three dimensional RNA structures and descriptors of common structural elements (e.g., kink-turns, **Table 1**) are found in the Nucleic Acid Database (NDB) and related databases (Coimbatore Narayanan et al., 2014). These elements can be automatically included in homology based 3D predictions (Antczak et al., 2016).

In addition to this general database, there are specialized databases containing particular structural or functional classes of elements, some of which are overrepresented in viruses. The database of pseudoknots (Pseudobase++) contains 252 virus records (accessed in June 2017). IRESite contains 44 viral IRES entries (June 2017; Mokrejs et al., 2010). Recode contains many viral recoding sites, in particular RNA elements stimulating frameshifting and readthrough (Bekaert et al., 2010).

VIRAL SUCCESS STORIES

To illustrate the key concepts of RNA structure prediction, in this section we review the approaches used to successfully study the RNA structures located in flaviviruses, influenza, and BYDV. Common types of RNA structures are illustrated in **Figure 1** and described in **Table 1**.

In choosing these examples we note that different concepts and approaches should be used in predicting the RNA structures located in the CDS in contrast to UTRs (**Figure 3**). RNA structures in the CDS have often been overlooked, and have only been discovered recently in some well-characterized viruses (see below). We will therefore review a successful story begun by several careful bioinformatics analyses of the CDS of the influenza A virus (Moss et al., 2011). As in experimental approaches, these examples show that independent approaches and tools have been required to accurately predict an RNA structure.

RNA Structures in the 3′UTRs of Flaviviruses

Flaviviruses are positive-sense single-stranded RNA viruses, e.g., the mosquito-borne Dengue and Zika viruses. The RNA structures of flaviviruses have recently been reviewed (Villordo et al., 2016; Fernández-Sanlés et al., 2017). The sequences and RNA structures of the 3′UTRs of flaviviruses have been studied over three decades. Earlier studies found that the 3′UTR sequences of flaviviruses are highly divergent immediately after the stop codon, but remarkably similar at the distal region of the 3′UTR (Mandl et al., 1993; Wallner et al., 1995; Poidinger et al., 1996). Earlier computational and biochemical studies also found that a long stable hairpin structure at the 3′UTRs of flaviviruses (3′-LSH) had a similar structure, but not sequence (Grange et al., 1985; Brinton et al., 1986; Hahn et al., 1987; Mandl et al., 1993; Wallner et al., 1995; Shi et al., 1996). Remarkably, the dumbbell RNA structures of the 3′UTRs of flaviviruses were first discovered by Proutski et al. and Rauscher et al. in 1997 using only computational approaches (Proutski et al., 1997; Rauscher et al., 1997) whereas Rauscher et al. used the Vienna RNA package and a comparative approach including multiple sequence alignment (Gruber et al., 2008; Lorenz et al.,

2011). Rauscher et al. found many covarying base-pairs in these structures, providing compelling evidence for RNA structure conservation. For example, there are 10 statistically significant covarying base-pairs in flavivirus dumbbell structures (RF00525; **Figures 2E,F**) currently annotated in Rfam (Nawrocki et al., 2014). Covarying base-pairs of RNA structures and the depth of aligned sequences can be statistically tested using R-scape (Rivas et al., 2017).

The RNA structures of flavivirus 3′UTRs were subsequently refined and proposed as H-type pseudoknots (tertiary structures) by Olsthoorn and Bol (2001) using mfold with suboptimal folding (e.g., 5′ dumbbell of dengue virus 2 flavivirus; **Figures 2A–D**). The structures of these flavivirus RNA elements were recently validated by SHAPE (Selective 2′-Hydroxyl Acylation analyzed by Primer Extension) chemical probing, mutation analysis and X-ray crystallography (Manzano et al., 2011; Chapman et al., 2014a,b; Villordo et al., 2015; Akiyama et al., 2016).

More importantly, many independent experiments have successfully uncovered their complex roles which have clinical implications. For example, deletion of the dengue virus 5′ dumbbell structure attenuates the virus, generating vaccine candidates that have been used for clinical testing (Whitehead et al., 2007). It is shown that assembly of the host RNA helicase DDX6 and other proteins at the dumbbell structures of dengue virus 2 is required for virus replication (Manzano et al., 2011; Ward et al., 2011). These 3′UTR structures also protect flaviviral subgenomic RNAs (sfRNAs) from the host Xrn1 5′-3′ exonuclease digestion (Pijlman et al., 2008). These sfRNAs are pathogenic and important in regulating viral life cycle (Manzano et al., 2011; Chapman et al., 2014b; Akiyama et al., 2016) and have been targeted by specific antiviral oligomers (Zhang et al., 2008).

RNA Structures in Coding Regions of Influenza A Virus

Influenza A virus is a zoonotic virus that infects a wide range of mammals and birds (Shi et al., 2014). It is a negative-sense single-stranded RNA virus that has an eight-segment genome. Moss et al. (2011) undertook a careful analysis of complete genomes of Influenza A strain H5N1 and H1N1 infecting human, avian and swine from NCBI Virus Variation Resource (Hatcher et al., 2017). This enabled them to discover many putative structured RNA elements located in the CDS of Influenza A virus (Moss et al., 2011).

To create multiple sequence alignment, Moss et al. first translated the CDS into protein sequences. The aligned protein sequences were then converted back to nucleotide sequences. They scanned the aligned CDS for putative RNA structures using RNAz (Gruber et al., 2007). They used sliding windows of 120-nucleotide, with 10-nucleotide steps. This allows rapid prediction of local RNA structures in the 120-nucleotide windows of the whole aligned sequence. They also detected synonymous (for the encoded protein) sites in the aligned CDS using SSE (Simmonds, 2012), that were constrained during evolution. These codon-based alignments detect synonymous constraints, possibly due to the presence of structured RNA elements. This is based on

the assumption that synonymous substitutions in a CDS are restricted by base-pairing required for RNA folding, but such constraints could also be due to primary sequence conservation in RNA (or DNA).

Alternatively, codon-based alignment could have been be done using webPRANK (Löytynoja and Goldman, 2010) or Codon Alignment (HCV sequence database; Kuiken et al., 2005). Significant synonymous constraint sites of aligned CDS can also be detected using FRESco (Sealfon et al., 2015) or synplot2 (Firth, 2014). Many automated alignments of viral genomes are available using codon based alignments in searches for conserved RNA structures or overlapping CDS (Hofacker et al., 2004; Firth and Brown, 2006; Firth, 2014).

Moss et al. (2011) predicted and refined the potential structured regions using RNAalifold (Bernhart et al., 2008), and Dynalign (Mathews and Turner, 2002). Pseudoknots were predicted using DotKnot (Sperschneider and Datta, 2010). Notably, a predicted pseudoknot located in the virus segment 2 genome was subsequently shown to be consistent with that of chemical probing results (Priore et al., 2015). The predicted RNA structures near the splice junctions of M1/M2 and NS1/NEP transcripts were also validated experimentally and/or found to be important for the virus viability (Moss et al., 2012b; Jiang et al., 2016).

To improve the power of detecting putative RNA structural elements, subsequent studies focused on specific genes/genome segments, namely HA (surface glycoprotein hemagglutinin; Gultyaev et al., 2016), M (Kobayashi et al., 2016), and NP (nucleoprotein; Gultyaev et al., 2014; Soszynska-Jozwiak et al., 2015) using deep multiple sequence alignment. Indeed, new structured RNA elements have been continuously discovered. For example, Kobayashi et al. (2016) analyzed 1884 sequences of M gene from 88 Influenza A virus subtypes. Similar to the Moss et al. (2011) approaches, they scanned the deeply aligned CDS for potential RNA structured regions and synonymous variations using SSE (Simmonds, 2012). Prediction of the RNA structured regions was based on UNAfold MFE algorithm implemented in SSE (Simmonds, 2012). They predicted RNA structures on the regions with both low MFE and synonymous substitution rate using RNAalifold. Remarkably, disrupting the base-pairs of the RNA structures located at the 5′ and 3′ ends of M gene using synonymous mutations reduced the infectivity and attenuated the virus, respectively (Kobayashi et al., 2016).

In sum, these studies highlight the strength of comparative approach in detection of RNA structures in the CDS. Different comparative methods used by these studies can be compared and combined to achieve better results. However, these powerful comparative approaches are underutilized by virologists.

RNA Structures in Barley Yellow Dwarf Virus (BYDV)

Luteoviruses including BYDV are important plant pathogens. BYDV infects barley, maize, oats, rice, and wheat, causing yellowing and dwarfing of the hosts (D'arcy and Domier, 2000). It is arguably the viral genome with the greatest range and diversity of RNA structures. The type member BYDV-PAV

has a 5.7 kb positive stranded RNA genome, and three coding and non-coding subgenomic RNAs (**Figure 6**). From the initial sequence of the genome and prediction of ORFs, it was likely that it would have require multiple non-canonical translation events to make key proteins—sgRNA expression, frameshift, readthrough, leaky-scanning, and cap-independent translation (Miller et al., 1988). Much careful experimental work indicated that these events require both structured and loosely-structured RNA elements (Miras et al., 2017). Studies on frameshifting and cap-independent translation in BYDV showed that both local and distant sequences are required for full activity (Miller et al., 2015). Some of these are used as illustrations in the following sections.

VIRAL RNA STRUCTURES

Stem-Loops/Hairpin

Most computational analysis will begin by predicting stem-loops but these can be built up into more complex structures (**Figure 1**, **Table 1**). Prediction accuracy is increased if an alignment of sequences that fold into the same structure is used (Gorodkin et al., 2014). When a stem-loop is predicted, attention should also be given to bulges, internal, and terminal loops (**Figure 1**). For example: terminal loops may form more stable structures e.g., tetraloops; and be sites of RNA or protein interaction; apparent bulges may form non-canonical pairs (e.g., A-G); and unpaired bases are more likely to form sites of interactions (Lozano et al., 2016). Modeling of loops is more difficult, but

can be done thermodynamically (Sloma and Mathews, 2016), by using similarity to known elements e.g., tetraloops, or known experimentally determined folds (Theis et al., 2015; Roll et al., 2016; Phan et al., 2017).

Pseudoknots

In some cases the terminal loop may form additional "pseudoknot" base-pairs (**Table 1**, **Figure 1C**, and **Figures 2A–D**). These are most easily visualized on arc and circular diagrams of the suboptimal RNA secondary structures (**Figures 2B,D**). Pseudoknots are found in specific parts of the viral genome involved in translation and replication (Brierley et al., 2007; Atkins et al., 2016), such as the domain IIIf of the HCV IRES (**Figure 7A**) and the dumbbell structures of dengue virus 2 (**Figure 5**), respectively.

The most common virus frameshift, is −1 frameshift, which often has a stimulatory pseudoknot(s). This frameshifting was first discovered in a retrovirus, Rous sarcoma virus (*gag/pol* frameshift; Jacks and Varmus, 1985). Many viral frameshifts have now been characterized [reviewed in detail by Atkins et al. (2016)]. Frameshifting elements consist of a slippery site, a spacer (∼5–9 nucleotides) and a frameshift stimulator (stem-loop or pseudoknot) [reviewed in detail by Giedroc and Cornish (2009) and Firth and Brierley (2012)]. However, frameshift stimulators are highly diverse (Chung et al., 2010). In some cases, non-canonical base-pairs e.g., base triples (Chen et al., 2017) and
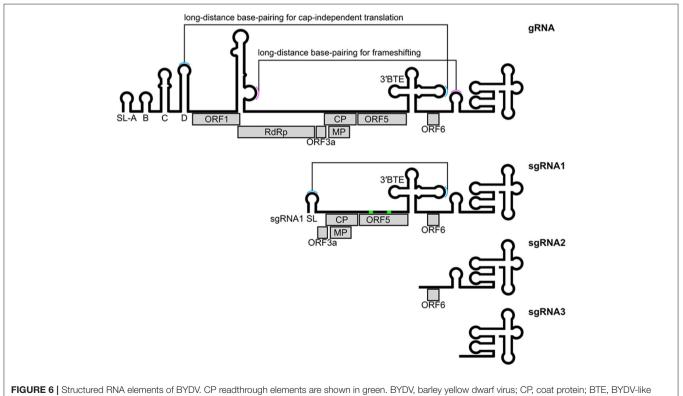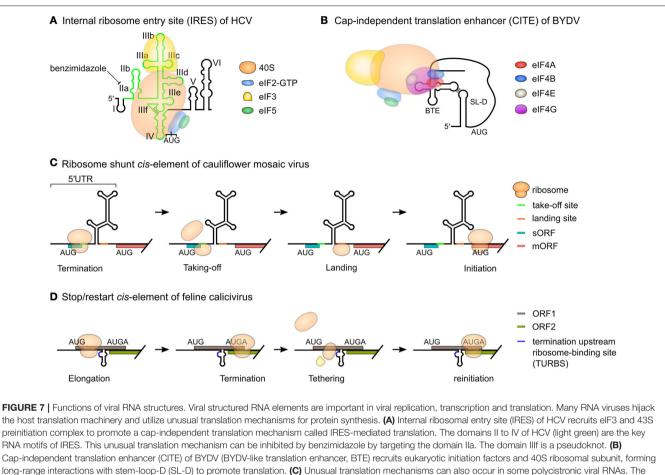

FIGURE 6 | Structured RNA elements of BYDV. CP readthrough elements are shown in green. BYDV, barley yellow dwarf virus; CP, coat protein; BTE, BYDV-like translation element; gRNA, genomic RNA; MP, movement protein; ORF, open reading frame; RdRp, RNA-dependent RNA polymerase; sgRNA, subgenomic RNA; SL, stem-loop.

FIGURE 7 | Functions of viral RNA structures. Viral structured RNA elements are important in viral replication, transcription and translation. Many RNA viruses hijack the host translation machinery and utilize unusual translation mechanisms for protein synthesis. **(A)** Internal ribosomal entry site (IRES) of HCV recruits eIF3 and 43S preinitiation complex to promote a cap-independent translation mechanism called IRES-mediated translation. The domains II to IV of HCV (light green) are the key RNA motifs of IRES. This unusual translation mechanism can be inhibited by benzimidazole by targeting the domain IIa. The domain IIIf is a pseudoknot. **(B)** Cap-independent translation enhancer (CITE) of BYDV (BYDV-like translation enhancer, BTE) recruits eukaryotic initiation factors and 40S ribosomal subunit, forming long-range interactions with stem-loop-D (SL-D) to promote translation. **(C)** Unusual translation mechanisms can also occur in some polycistronic viral RNAs. The 5′UTR of cauliflower mosaic virus is long and highly structured. The highly structured region contains multiple upstream AUGs. A highly structured 5′UTR with multiple upstream AUGs could inhibit translation of the main open reading frame (mORF) of a eukaryotic mRNA. Cauliflower mosaic virus overcome this problem with ribosome shunt *cis*-element. A ribosome first translate the small ORF (sORF) at the viral 5′UTR. During translation termination, the ribosome dissociates but the take-off site (the sequence surrounding the termination codon) induce ribosome shunting. This allows the ribosome to bypass the highly structured region of the 5′UTR, land on the landing site, followed by translation of the mORF. **(D)** Feline calicivirus contains two ORFs with a slightly overlapping sequence AUGA. A structured motif called stop/restart *cis*-element located upstream of AUGA permits effective reinitation and translation of the second ORF. A termination upstream ribosome-binding site (TURBS) located in the RNA structure allows tethering of 40S ribosomal subunit and eIF3. This promotes reinitiation of the second ORF.

long-range base-pairs may be required for −1 frameshifting (e.g., in BYDV, **Figure 6**; Miras et al., 2017).

In the prediction of −1 frameshifting RNA elements, a slippery site is usually searched for between the two frames (zero and −1). This has a consensus sequence of "X XXY YYZ," the zero frame codons are separated by spaces, X is an identical nucleotide, Y is either A or U, and Z is not G (Brierley et al., 2007). A ribosome first encounters the slowly decoded codons of the slippery site. The stalled ribosome then "slips" one nucleotide backward (−1 frame; XXX YYY Z) to resume translation elongation. In most cases, a frameshift stimulator downstream (e.g., RNA structure) of the slippery site is required for physiological frameshifting. Although such structure is dispensable in at least one virus, Hibiscus latent Singapore virus (Niu et al., 2014).

Pseudoknots are often predicted by visual inspection from studying the local secondary structures following these slippery sites. Software has also been used, for example, the H-type pseudoknot structure modulating −1 frameshifting in

Japanese Encephalitis virus was successfully predicted using PknotsRG (now part of pKiss; Janssen and Giegerich, 2015) and experimentally validated (Melian et al., 2009). This was confirmed in an independent study on a vaccine strain that harbors a synonymous mutation that abolishes the RNA structure (Sun et al., 2012). However, *de novo* pseudoknot computational prediction remains challenging, and current tools are <5% accurate (Leamy et al., 2016). This may be improved by including experimental data (Hajdin et al., 2013).

These ribosomal frameshifting sites can be predicted specifically using KnotInFrame (Theis et al., 2008). Others have used more general software, e.g., RNA Shapes Studio (Janssen and Giegerich, 2015) as was done for Zika virus, or combination of prediction programs as was used to predict a functional pseudoknot in West Nile virus (Moomau et al., 2016). Models in 3D can be built of pseudoknots e.g., using MC-Sym as was recently done for Venezuelan equine encephalitis virus (Kendra et al., 2017).

Kissing Hairpins

Kissing hairpins (also known as kissing-loops or kissing stem-loops) are formed from the base-pairing between the loop of two stem-loops (**Table 1**, **Figure 1**). Many kissing hairpins are related to virus replication or transcription (You and Rice, 2008; Ganser and Al-Hashimi, 2016).

The first viral kissing hairpins were discovered in enteroviruses (plus strand viruses), namely poliovirus and coxsackievirus B3 (Pilipenko et al., 1992). These structures are located at the 3′UTR of an enterovirus genomic RNA and required for synthesis of the viral negative strand RNA template (Dutkiewicz et al., 2016). These kissing hairpins are formed by base-pairing of two adjacent stem-loops which are known as X and Y motifs. Interestingly, the primary sequence of these motifs are conserved only in certain enterovirus subgroups but the Y motif variants were shown to be interchangeable between poliovirus and coxsackievirus B3 (Zoll et al., 2009).

Another well-studied example is the retroviral dimerization initiation sites (DIS). This structure is involved in dimerization of virus genomic RNAs, which is a critical step in retroviral replication (Paillart et al., 2004). The "kissing" begins at the DIS of two virus genomic RNAs prior to encapsidation (Mailler et al., 2016).

DotKnot (Sperschneider and Datta, 2010), pKnots (Rivas and Eddy, 1999), pKiss (Theis et al., 2010), and pAliKiss (Janssen and Giegerich, 2015) could also be used to predict such kissing hairpins. However, these are limited to predicting intramolecular kissing interactions.

Cloverleaf/tRNA-Like Structures

A tRNA-like structure harbors a four-way junction—three stem-loops (a cloverleaf), and in viral structures may also contain additional pseudoknots (**Figure 1**, **Table 1**). In enteroviruses, a cloverleaf structure known as *ori*L is involved in viral replication (Prostova et al., 2017). It is located at the 5′ leader of the plus strand genomic RNAs (Dutkiewicz et al., 2016). The cellular PCBP [poly(rC)-binding protein] and viral protein 3CD^pro binds to two different stem-loops of *ori*L, forming a replication complex. Other proteins could also bind to *ori*L (Prostova et al., 2017). A cloverleaf structure is also formed in the negative strand RNA template, the kissing interaction of the hairpin loops, within the cloverleaf structure is required for viral genomic RNA synthesis (Melchers et al., 1997).

Many positive strand plant viruses have such tRNA-like structures in the 3′UTRs of genomic RNAs (Dreher, 2010). Most viral tRNA-like structures are aminoacylated (e.g., by Val, His, or Tyr), mimicking cellular tRNAs to regulate translation. A recent study proposed that these tRNA-like structures can also act as mobile elements in plant by promoting transport of viral transcripts via phloem sap (Zhang et al., 2016). These tRNA-like structures are amenable to both modeling and experimental 3D determination. For example, the tRNA-like structure of Tobacco mosaic virus was recently solved by X-ray crystallography (Colussi et al., 2014), this tRNA-like structure has multiple additional upstream pseudoknots. Indeed, viral tRNA-like structures discovered to date have variable sequence, length, and structures (Dreher, 2010).

Long-Range Intra-Molecular Interactions

The elements considered above form mainly local structures. These local structures may form in nascent RNA (Meyer, 2017) or be stabilized by protein or RNA binding. Local structures can be predicted using appropriate windows of sequence (e.g., 80–200 bases) and it is also practical to analyse local alignments of similar lengths (Lange et al., 2012).

Long-range interactions from over a few hundred bases to >26 kb do occur in RNA viruses, but are difficult to predict accurately. Challenges include: there are many possible interactions; likely complex structures (e.g., pseudoknots); structures will form co-transcriptionally limiting interactions; and small molecules, proteins, RNAs and complexes (e.g., ribosomes) will bind and affect folding (Lai and Meyer, 2016; Napthine et al., 2016; Sun et al., 2017).

Methods based on MFE when applied to long RNAs (e.g., mfold on a viral genome) will tend to predict large structures with a large number of long-distance interactions—which should be viewed with caution. Indeed, experimentally determined structures of full length genomes show more local than long-range interactions e.g., HIV RNA has many local structures (Watts et al., 2009) but only five long-distance interactions (Fricke and Marz, 2016).

Specific tools have been developed to predict long-range interactions, e.g., LRIscan, with 14 of 16 known long-distance interactions confirmed and plausible candidates from other viruses predicted (Fricke and Marz, 2016).

Efficient frameshifting, in addition to the local frameshifting elements (e.g., pseudoknots), may require long-range interactions (Nicholson and White, 2014). These have been well-characterized in BYDV (Paul et al., 2001; Barry and Miller, 2002) and red clover necrotic mosaic virus (RCNMV; Tajima et al., 2011) and involve long-range kissing interactions (**Figure 6**).

Inter-Molecular Interactions

Some viral RNAs also form structures with other RNAs, both viral and cellular. Notable examples are viral RNA dimerization elements (see subsection "Kissing Hairpins"), co-packaging elements, or interactions during translation with the rRNA in the ribosome (Deforges et al., 2015; Angulo et al., 2016). Co-packaging of multiple segments of RNA may suggest that inter-molecular interactions occur, for example in RCNMV the loop of origin of assembly stem-loop on RNA2 interacts with on RNA1 (Newburn and White, 2015). Specific software e.g., RNAhybrid can be used to predict such inter-molecular interactions (Rehmsmeier et al., 2004).

FUNCTIONS OF VIRAL RNA STRUCTURES

The roles of many structured RNA elements of viruses have been studied in detail. Some examples are discussed in this section.

Internal Ribosome Entry Sites (IRES)

Viral RNAs are not always capped, this means that they have evolved specific mechanisms to enhance cap-independent translation. The RNAs of many viruses contain large structured

IRES, to promote this. Well-characterized examples are found in picornaviruses and HCV (Lee et al., 2017; **Figure 7A**). The IRES recruits ribosomes near or directly to the translation initiation codons of viral mRNAs, bypassing the need for the cap-binding complex. This allows the virus to manipulate the host translation machinery by inhibition or proteolytic cleavage of host eukaryotic initiation factors (eIFs). Translation of viral mRNAs is possible even during the host translation shutoff (Lee et al., 2017).

However, viral IRES are one of the most challenging structural elements to predict and characterize. This is because IRES are complex and diverse, often consisting of multiple stem-loops and/or pseudoknots (Dreher, 2010; Lozano et al., 2016). Limited progress has been made in development of automated pipelines for IRES prediction, however, two specialized webservers are available, namely VIPS (Hong et al., 2013) and IRESPred (Kolekar et al., 2016). Both VIPS and IRESPred predict IRES based on known IRES sequences and structures. In particular, IRESPred looks for the binding sequence motifs of small subunit ribosomal proteins. Alternatively, a combination of tools, in particular BLAST, Pfold, Centroid Fold, mfold, and pKiss have also proven to be useful in IRES prediction (Asnani et al., 2015, 2016).

Ribosome Shunt *cis*-Elements

Ribosome shunting consists of a series of unusual translation events (**Figure 7C**). A ribosome first initiates at a small ORF (sORF) and terminates right before a large RNA structure. The large ribosomal subunit dissociate but the small subunit bypasses the RNA structure, docks on a landing site and resumes scanning. The ribosome can then reinitiate even at a non-AUG codon. The shunt elements were discovered in DNA viruses, first in the pararetrovirus cauliflower mosaic virus (Fütterer et al., 1990, 1993), then a retrovirus, prototype foamy virus (Schepetilnikov et al., 2009), and a plant RNA virus, rice tungro spherical virus (Pooggin et al., 2012).

To predict the *cis*-elements driving ribosome shunting, several key characteristics of these elements have been taken into account. These elements are located in a long, highly structured 5′UTR of the virus genomic RNA that has multiple upstream AUGs. These 5′UTR features appear to inhibit translation of the main ORF(s). The sORF(s) precede the large RNA structure is involved, whereas the following upstream AUGs are folded up in a large RNA structure. This RNA structure has a stable base-pairing at the stem base. The shunt take-off site (sequence around the sORF termination codon) and landing site are expected to be conserved between closely related viruses or co-evolved viruses. For example, the shunt *cis*-elements are remarkably similar between a pair of co-evolved viruses, a RNA picorna-like virus, rice tungro spherical virus and a DNA pararetrovirus, rice tungro bacilliform virus (Pooggin et al., 2012).

Cap-Independent Translation Enhancers (CITEs)

Cap-independent translation may be stimulated by local RNA structures but surprisingly in some cases also by long-distance base-pairing. This long-distance base-pairing has been well-characterized in several plant viral RNAs (Miras et al., 2017). For example, BYDV has a CITE located at the 3′ end (**Figures 6, 7B**). This element interacts with a stem-loop located at the 5′UTR (long-range kissing interactions) to promote cap-independent translation (Miller et al., 2015). Other viral genera also use long-distance base-pairing or interaction with rRNA (Deforges et al., 2015). Published models of two such complex structures have been made using RNA2D3D (Mccormack et al., 2008) and MC-Sym (Wang et al., 2011; Newburn and White, 2015).

Stop/Restart *cis*-Elements

Some viruses use unusual mechanisms to reinitiate after translation of a long CDS. These stop/restart or termination-reinitiation mechanisms were initially found in Caliciviruses (**Figure 7D**) and then Influenza B viruses (reviewed in detail by Powell, 2010). These mechanisms allow effective translation of both the first and second ORFs of a viral mRNA, producing two distinct functional proteins (Zinoviev et al., 2015). These mechanisms require several *cis*-regulatory elements that can be partially structured, and these may interact with other RNAs e.g., the 18S rRNA.

These mechanisms are distinct from the mechanisms utilizing upstream ORFs (uORFs), or programmed ribosomal frameshifting (Miras et al., 2017). In eukaryotic mRNAs, including viral ones, uORFs are commonly found to repress translation of the mORFs (Hellens et al., 2016; Zong et al., 2017). These regulatory uORFs are usually short and therefore producing only small peptides (Hellens et al., 2016; Starck et al., 2016). Whereas, in ribosomal frameshifting, only one protein is produced with the use of two overlapping ORFs (Atkins et al., 2016).

The stop-start *cis*-elements in Caliciviruses and Influenza B viruses are found between −84 and the start codon of the second ORF (Powell, 2010; Zinoviev et al., 2015). These *cis*-elements consist of a termination upstream ribosome-binding site (TURBS) and a stop/restart site. TURBS consists of a motif 1 (18S rRNA complementary site), and motif 2 and 2* (likely base-pairing and structured). Motif 1 is loosely structural to allow tethering of small ribosomal subunit for reinitiation whereas motif 2 and 2* could form an RNA structure that enhances translation of the second ORF (Lee et al., 2017).

Recently, stop/restart *cis*-elements were found in helminthosporium victoriae virus 190S. These elements consist of a H-type pseudoknot and an A**UGA** stop/restart site (start and stop codons are underlined and bolded, respectively; Li et al., 2015). This pseudoknot was successfully predicted using HPknotter (Huang et al., 2005). Disruption of the tertiary base-pairs abolishes translation of the second ORF (Li et al., 2015).

To predict stop/restart *cis*-elements, one could first look for slightly overlapping ORFs with a stop/restart site. However, these *cis*-elements also enable translation of a synthetic, non-overlapping second ORF effectively within a range of 40 nucleotides downstream of the first ORF (Ahmadian et al., 2000; Napthine et al., 2009; Zinoviev et al., 2015). Motif 1 (18S rRNA complementary site) is likely present between −84 and the start codon of the second ORF. A RNA structure may also found within the region. However, suboptimal RNA structures could also be present (Napthine et al., 2009).

CHALLENGES AND LIMITATIONS

This review has presented examples where virology research has been enhanced by the appropriate use of bioinformatic methods for RNA structure prediction. These examples highlight how computer predictions were used in conjunction with experimental tools for functional studies. Some of computational tools and resources are generally applicable to RNA structure prediction whereas others are specific to virology. Additional prediction tools are continually becoming available (Backofen et al., 2017; Miao and Westhof, 2017; Miao et al., 2017).

However, some challenges remain for the application of newer RNA structure tools in virology (**Table 1**). Some of these are being addressed by user friendly suites and tools becoming available as noted throughout this review and listed in the companion website (http://bioanalysis.otago.ac.nz/ Lim2017.htm). In addition, specialized workshops and training may facilitate the use of these RNA tools e.g., The EMBO Practical Course on Computational RNA Biology course material available online (https://bibiserv.cebitec.uni-bielefeld.de/EMBO-RNACourse/).

AUTHOR CONTRIBUTIONS

Both authors have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Achar, A., and Sætrom, P. (2015). RNA motif discovery: a computational overview. *Biol. Direct* 10, 61. doi: 10.1186/s13062-015-0090-5

Ahmadian, G., Randhawa, J. S., and Easton, A. J. (2000). Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. *EMBO J.* 19, 2681–2689. doi: 10.1093/emboj/19. 11.2681

Akiyama, B. M., Laurence, H. M., Massey, A. R., Costantino, D. A., Xie, X., Yang, Y., et al. (2016). Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science* 354, 1148–1152. doi: 10.1126/science.aah3963

Angulo, J., Ulryck, N., Deforges, J., Chamond, N., Lopez-Lastra, M., Masquida, B., et al. (2016). LOOP IIId of the HCV IRES is essential for the structural rearrangement of the 40S-HCV IRES complex. *Nucleic Acids Res.* 44, 1309–1325. doi: 10.1093/nar/gkv1325

Antczak, M., Popenda, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., et al. (2016). New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochim. Pol.* 63, 737–744. doi: 10.18388/abp.2016_1329

Asnani, M., Kumar, P., and Hellen, C. U. T. (2015). Widespread distribution and structural diversity of Type IV IRESs in members of Picornaviridae. *Virology* 478, 61–74. doi: 10.1016/j.virol.2015.02.016

Asnani, M., Pestova, T. V., and Hellen, C. U. T. (2016). Initiation on the divergent Type I cadicivirus IRES: factor requirements and interactions with the translation apparatus. *Nucleic Acids Res.* 44, 3390–3407. doi: 10.1093/nar/gkw074

Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., and Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* 44, 7007–7078. doi: 10.1093/nar/gkw530

Backofen, R., Engelhardt, J., Erxleben, A., Fallmann, J., Grüning, B., Ohler, U., et al. (2017). RNA-bioinformatics: tools, services and databases for the analysis of RNA-based regulation. *J. Biotechnol.* 261, 76–84. doi: 10.1016/j.jbiotec.2017.05.019

Bardwell, V. J., and Wickens, M. (1990). Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res.* 18, 6587–6594. doi: 10.1093/nar/18.22.6587

Barry, J. K., and Miller, W. A. (2002). A−1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11133–11138. doi: 10.1073/pnas.162223099

Bekaert, M., Firth, A. E., Zhang, Y., Gladyshev, V. N., Atkins, J. F., and Baranov, P. V. (2010). Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Res.* 38, D69–D74. doi: 10.1093/nar/gkp788

Bellaousov, S., Reuter, J. S., Seetin, M. G., and Mathews, D. H. (2013). RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.* 41, W471–W474. doi: 10.1093/nar/gkt290

Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R., and Stadler, P. F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9:474. doi: 10.1186/1471-2105-9-474

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., and Long, R. M. (1998). Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* 2, 437–445. doi: 10.1016/S1097-2765(00)80143-4

Bindewald, E., and Shapiro, B. A. (2013). Computational detection of abundant long-range nucleotide covariation in Drosophila genomes. *RNA* 19, 1171–1182. doi: 10.1261/rna.037630.112

Bindewald, E., Hayes, R., Yingling, Y. G., Kasprzak, W., and Shapiro, B. A. (2008). RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. *Nucleic Acids Res.* 36, D392–D397. doi: 10.1093/nar/gkm842

Brierley, I., Pennell, S., and Gilbert, R. J. C. (2007). Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat. Rev. Microbiol.* 5, 598–610. doi: 10.1038/nrmicro1704

Brinton, M. A., and Basu, M. (2015). Functions of the 3′ and 5′ genome RNA regions of members of the genus Flavivirus. *Virus Res.* 206, 108–119. doi: 10.1016/j.virusres.2015.02.006

Brinton, M. A., Fernandez, A. V., and Dispoto, J. H. (1986). The 3′-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology* 153, 113–121. doi: 10.1016/0042-6822(86)90012-7

Brister, J. R., Ako-Adjei, D., Bao, Y., and Blinkova, O. (2015). NCBI viral genomes resource. *Nucleic Acids Res.* 43, D571–D577. doi: 10.1093/nar/gku1207

Cantara, W. A., Olson, E. D., and Forsyth, K. M. (2014). Progress and outlook in structural biology of large viral RNAs. *Virus Res.* 193, 24–38. doi: 10.1016/j.virusres.2014.06.007

Cardno, T. S., Shimaki, Y., Sleebs, B. E., Lackovic, K., Parisot, J. P., Moss, R. M., et al. (2015). HIV-1 and human PEG10 frameshift elements are functionally distinct and distinguished by novel small molecule modulators. *PLoS ONE* 10:e0139036. doi: 10.1371/journal.pone.0139036

Chapman, E. G., Costantino, D. A., Rabe, J. L., Moon, S. L., Wilusz, J., Nix, J. C., et al. (2014a). The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* 344, 307–310. doi: 10.1126/science.1250897

Chapman, E. G., Moon, S. L., Wilusz, J., and Kieft, J. S. (2014b). RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *Elife* 3:e01892. doi: 10.7554/eLife.01892

Chase, A. J., and Semler, B. L. (2012). Viral subversion of host functions for picornavirus translation and RNA replication. *Future Virol.* 7, 179–191. doi: 10.2217/fvl.12.2

Chen, A., and Brown, C. (2012). Distinct families of cis-acting RNA replication elements epsilon from hepatitis B viruses. *RNA Biol.* 9, 130–136. doi: 10.4161/rna.18649

Chen, A., Panjaworayan T-Thienprasert, N., and Brown, C. M. (2014). Prospects for inhibiting the post-transcriptional regulation of gene expression in hepatitis B virus. *World J. Gastroenterol.* 20, 7993–8004. doi: 10.3748/wjg.v20.i25.7993

Chen, X. S., and Brown, C. M. (2012). Computational identification of new structured cis-regulatory elements in the 3′-untranslated region of human protein coding genes. *Nucleic Acids Res.* 40, 8862–8873. doi: 10.1093/nar/gks684

Chen, Y.-T., Chang, K.-C., Hu, H.-T., Chen, Y.-L., Lin, Y.-H., Hsu, C.-F., et al. (2017). Coordination among tertiary base pairs results in an efficient frameshift-stimulating RNA pseudoknot. *Nucleic Acids Res.* 45, 6011–6022. doi: 10.1093/nar/gkx134

Chung, B. Y., Firth, A. E., and Atkins, J. F. (2010). Frameshifting in Alphaviruses: a diversity of 3′ stimulatory structures. *J. Mol. Biol.* 397, 448–456. doi: 10.1016/j.jmb.2010.01.044

Coimbatore Narayanan, B., Westbrook, J., Ghosh, S., Petrov, A. I., Sweeney, B., Zirbel, C. L., et al. (2014). The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Res.* 42, D114–D122. doi: 10.1093/nar/gkt980

Coller, J. M., Gray, N. K., and Wickens, M. P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes Dev.* 12, 3226–3235. doi: 10.1101/gad.12.20.3226

Colussi, T. M., Costantino, D. A., Hammond, J. A., Ruehle, G. M., Nix, J. C., and Kieft, J. S. (2014). The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA. *Nature* 511, 366–369. doi: 10.1038/nature13378

Cornish, P. V., Hennig, M., and Giedroc, D. P. (2005). A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated−1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12694–12699. doi: 10.1073/pnas.0506166102

D'arcy, C. J., and Domier, L. L. (2000). Barley yellow dwarf. *PHI*. doi: 10.1094/PHI-I-2000-1103-01. Available online at: http://www.apsnet.org/edcenter/intropp/lessons/viruses/Pages/BarleyYelDwarf.aspx

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25, 1974–1975. doi: 10.1093/bioinformatics/btp250

Dawson, W. K., and Bujnicki, J. M. (2016). Computational modeling of RNA 3D structures and interactions. *Curr. Opin. Struct. Biol.* 37, 22–28. doi: 10.1016/j.sbi.2015.11.007

Deforges, J., Locker, N., and Sargueil, B. (2015). mRNAs that specifically interact with eukaryotic ribosomal subunits. *Biochimie* 114, 48–57. doi: 10.1016/j.biochi.2014.12.008

Dibrov, S. M., Ding, K., Brunn, N. D., Parker, M. A., Bergdahl, B. M., Wyles, D. L., et al. (2012). Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5223–5228. doi: 10.1073/pnas.1118699109

Dreher, T. W. (2010). Viral tRNAs and tRNA-like structures. *Wiley Interdiscip. Rev. RNA* 1, 402–414. doi: 10.1002/wrna.42

D'souza, V., Dey, A., Habib, D., and Summers, M. F. (2004). NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* 337, 427–442. doi: 10.1016/j.jmb.2004.01.037

Dutkiewicz, M., Stachowiak, A., Swiatkowska, A., and Ciesiołka, J. (2016). Structure and function of RNA elements present in enteroviral genomes. *Acta Biochim. Pol.* 63, 623–630. doi: 10.18388/abp.2016_1337

Eggenhofer, F., Hofacker, I. L., and Höner Zu Siederdissen, C. (2013). CMCompare webserver: comparing RNA families via covariance models. *Nucleic Acids Res.* 41, W499–W503. doi: 10.1093/nar/gkt329

Eggenhofer, F., Hofacker, I. L., and Höner Zu Siederdissen, C. (2016). RNAlien - Unsupervised RNA family model construction. *Nucleic Acids Res.* 44, 8433–8441. doi: 10.1093/nar/gkw558

Egli, M., Minasov, G., Su, L., and Rich, A. (2002). Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4302–4307. doi: 10.1073/pnas.062055599

Fallmann, J., Will, S., Engelhardt, J., Grüning, B., Backofen, R., and Stadler, P. F. (2017). Recent advances in RNA folding. *J. Biotechnol.* 261, 97–104. doi: 10.1016/j.jbiotec.2017.07.007

Fang, Y., Treffers, E. E., Li, Y., Tas, A., Sun, Z., van der Meer, Y., et al. (2012). Efficient−2 frameshifting by mammalian ribosomes to synthesize an

additional arterivirus protein. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2920–E2928. doi: 10.1073/pnas.1211145109

Fernández-Sanlés, A., Ríos-Marco, P., Romero-López, C., and Berzal-Herranz, A. (2017). Functional information stored in the conserved structural RNA domains of flavivirus genomes. *Front. Microbiol.* 8:546. doi: 10.3389/fmicb.2017.00546

Firth, A. E. (2014). Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic Acids Res.* 42, 12425–12439. doi: 10.1093/nar/gku981

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

Firth, A. E., and Brown, C. M. (2006). Detecting overlapping coding sequences in virus genomes. *BMC Bioinformatics* 7:75. doi: 10.1186/1471-2105-7-75

Fricke, M., and Marz, M. (2016). Prediction of conserved long-range RNA-RNA interactions in full viral genomes. *Bioinformatics* 32, 2928–2935. doi: 10.1093/bioinformatics/btw323

Fütterer, J., Gordon, K., Sanfaçon, H., Bonneville, J. M., and Hohn, T. (1990). Positive and negative control of translation by the leader sequence of cauliflower mosaic virus pregenomic 35S RNA. *EMBO J.* 9, 1697–1707.

Fütterer, J., Kiss-László, Z., and Hohn, T. (1993). Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* 73, 789–802. doi: 10.1016/0092-8674(93)90257-Q

Ganser, L. R., and Al-Hashimi, H. M. (2016). HIV-1 leader RNA dimeric interface revealed by NMR. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13263–13265. doi: 10.1073/pnas.1615789113

Gardner, P. P., and Giegerich, R. (2004). A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics* 5:140. doi: 10.1186/1471-2105-5-140

Gardner, P. P., Wilm, A., and Washietl, S. (2005). A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Res.* 33, 2433–2439. doi: 10.1093/nar/gki541

Giedroc, D. P., and Cornish, P. V. (2009). Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.* 139, 193–208. doi: 10.1016/j.virusres.2008.06.008

Gorodkin, J., Hofacker, I. L., and Ruzzo, W. L. (2014). Concepts and introduction to RNA bioinformatics. *Methods Mol. Biol.* 1097, 1–31. doi: 10.1007/978-1-62703-709-9_1

Grange, T., Bouloy, M., and Girard, M. (1985). Stable secondary structures at the 3′-end of the genome of yellow fever virus (17 D vaccine strain). *FEBS Lett.* 188, 159–163. doi: 10.1016/0014-5793(85)80895-4

Graveley, B. R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol. Cell* 1, 765–771. doi: 10.1016/S1097-2765(00)80076-3

Griffiths-Jones, S. (2005). RALEE–RNA aLignment editor in emacs. *Bioinformatics* 21, 257–259. doi: 10.1093/bioinformatics/bth489

Groom, H. C., Anderson, E. C., and Lever, A. M. (2009). Rev: beyond nuclear export. *J. Gen. Virol.* 90, 1303–1318. doi: 10.1099/vir.0.011460-0

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Res.* 36, W70–W74. doi: 10.1093/nar/gkn188

Gruber, A. R., Neuböck, R., Hofacker, I. L., and Washietl, S. (2007). The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. *Nucleic Acids Res.* 35, W335–W338. doi: 10.1093/nar/gkm222

Gultyaev, A. P., and Olsthoorn, R. C. L. (2010). A family of non-classical pseudoknots in influenza A and B viruses. *RNA Biol.* 7, 125–129. doi: 10.4161/rna.7.2.11287

Gultyaev, A. P., Spronken, M. I., Richard, M., Schrauwen, E. J. A., Olsthoorn, R. C. L., and Fouchier, R. A. M. (2016). Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes. *Sci. Rep.* 6:38892. doi: 10.1038/srep38892

Gultyaev, A. P., Tsyganov-Bodounov, A., Spronken, M. I. J., Van Der Kooij, S., Fouchier, R. M., and Olsthoorn, R. C. L. (2014). RNA structural constraints in the evolution of the influenza A virus genome NP segment. *RNA Biol.* 11, 942–952. doi: 10.4161/rna.29730

Hahn, C. S., Hahn, Y. S., Rice, C. M., Lee, E., Dalgarno, L., Strauss, E. G., et al. (1987). Conserved elements in the 3′ untranslated region of flavivirus RNAs and potential cyclization sequences. *J. Mol. Biol.* 198, 33–41. doi: 10.1016/0022-2836(87)90455-4

Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C. W., Mathews, D. H., and Weeks, K. M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5498–5503. doi: 10.1073/pnas.1219988110

Hartmann, G. (2017). Nucleic acid immunity. *Adv. Immunol.* 133, 121–169. doi: 10.1016/bs.ai.2016.11.001

Hatcher, E. L., Zhdanov, S. A., Bao, Y., Blinkova, O., Nawrocki, E. P., Ostapchuck, Y., et al. (2017). Virus variation resource - improved response to emergent viral outbreaks. *Nucleic Acids Res.* 45, D482–D490. doi: 10.1093/nar/gkw1065

Hellens, R. P., Brown, C. M., Chisnall, M. A., Waterhouse, P. M., and Macknight, R. C. (2016). The emerging world of small ORFs. *Trends Plant Sci.* 21, 317–328. doi: 10.1016/j.tplants.2015.11.005

Hermann, T. (2016). Small molecules targeting viral RNA. *Wiley Interdisc. Rev. RNA* 7, 726–743. doi: 10.1002/wrna.1373

Hermann, T. (2018). "Viral RNA targets and their small molecule ligands," in *RNA Therapeutics*, ed A. L. Garner (Cham: Springer International Publishing), 111–134.

Hilimire, T. A., Chamberlain, J. M., Anokhina, V., Bennett, R. P., Swart, O., Myers, J. R., et al. (2017). HIV-1 frameshift RNA-targeted triazoles inhibit propagation of replication-competent and multi-drug-resistant HIV in human cells. *ACS Chem. Biol.* 12, 1674–1682. doi: 10.1021/acschembio.7b00052

Hill, A. C., Bartley, L. E., and Schroeder, S. J. (2016). Prohead RNA: a noncoding viral RNA of novel structure and function. *Wiley Interdisc. Rev. RNA* 7, 428–437. doi: 10.1002/wrna.1330

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte Chem. Chem. Monthly* 125, 167–188. doi: 10.1007/BF00818163

Hofacker, I. L., Stadler, P. F., and Stocsits, R. R. (2004). Conserved RNA secondary structures in viral genomes: a survey. *Bioinformatics* 20, 1495–1499. doi: 10.1093/bioinformatics/bth108

Holmes, E. C. (2010). The RNA virus quasispecies: fact or fiction? *J. Mol. Biol.* 400, 271–273. doi: 10.1016/j.jmb.2010.05.032

Hong, J.-J., Wu, T.-Y., Chang, T.-Y., and Chen, C.-Y. (2013). Viral IRES prediction system - a web server for prediction of the IRES secondary structure *in silico*. *PLoS ONE* 8:e79288. doi: 10.1371/journal.pone.0079288

Huang, C.-H., Lu, C. L., and Chiu, H.-T. (2005). A heuristic approach for detecting RNA H-type pseudoknots. *Bioinformatics* 21, 3501–3508. doi: 10.1093/bioinformatics/bti568

Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., et al. (2011). ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 39, D576–D582. doi: 10.1093/nar/gkq901

Jacks, T., and Varmus, H. E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 230, 1237–1242. doi: 10.1126/science.2416054

Janssen, S., and Giegerich, R. (2015). The RNA shapes studio. *Bioinformatics* 31, 423–425. doi: 10.1093/bioinformatics/btu649

Jiang, T., Nogales, A., Baker, S. F., Martinez-Sobrido, L., and Turner, D. H. (2016). Mutations designed by ensemble defect to misfold conserved RNA structures of influenza A segments 7 and 8 affect splicing and attenuate viral replication in cell culture. *PLoS ONE* 11:e0156906. doi: 10.1371/journal.pone.0156906

Johansson, H. E., Dertinger, D., Lecuyer, K. A., Behlen, L. S., Greef, C. H., and Uhlenbeck, O. C. (1998). A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc. Natl. Acad. Sci. U.S.A.* 95, 9244–9249. doi: 10.1073/pnas.95.16.9244

Johnson, S. F., and Telesnitsky, A. (2010). Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathog.* 6:e1001007. doi: 10.1371/journal.ppat.1001007

Kendra, J. A., de la Fuente, C., Brahms, A., Woodson, C., Bell, T. M., Chen, B., et al. (2017). Ablation of programmed−1 ribosomal frameshifting in venezuelan equine encephalitis virus results in attenuated neuropathogenicity. *J. Virol.* 91:e01766-16. doi: 10.1128/JVI.01766-16

King, A. M. Q., Lefkowitz, E., Adams, M. J., and Carstens, E. B. (2011). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier.

Kobayashi, Y., Dadonaite, B., van Doremalen, N., Suzuki, Y., Barclay, W. S., and Pybus, O. G. (2016). Computational and molecular analysis of conserved influenza A virus RNA secondary structures involved in infectious virion production. *RNA Biol.* 13, 883–894. doi: 10.1080/15476286.2016.1208331

Kolekar, P., Pataskar, A., Kulkarni-Kale, U., Pal, J., and Kulkarni, A. (2016). IRESPred: web server for prediction of cellular and viral internal ribosome entry site (IRES). *Sci. Rep.* 6:27436. doi: 10.1038/srep27436

Kuiken, C., Thurmond, J., Dimitrijevic, M., and Yoon, H. (2012). The LANL hemorrhagic fever virus database, a new platform for analyzing biothreat viruses. *Nucleic Acids Res.* 40, D587–D592. doi: 10.1093/nar/gkr898

Kuiken, C., Yusim, K., Boykin, L., and Richardson, R. (2005). The Los alamos hepatitis C sequence database. *Bioinformatics* 21, 379–384. doi: 10.1093/bioinformatics/bth485

Lai, D., and Meyer, I. M. (2016). A comprehensive comparison of general RNA-RNA interaction prediction methods. *Nucleic Acids Res.* 44, e61. doi: 10.1093/nar/gkv1477

Lai, D., Proctor, J. R., Zhu, J. Y. A., and Meyer, I. M. (2012). R-CHIE: a web server and R package for visualizing RNA secondary structures. *Nucleic Acids Res.* 40:e95. doi: 10.1093/nar/gks241

Lange, S. J., Maticzka, D., Möhl, M., Gagnon, J. N., Brown, C. M., and Backofen, R. (2012). Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res.* 40, 5215–5226. doi: 10.1093/nar/gks181

Lauring, A. S., and Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6:e1001005. doi: 10.1371/journal.ppat.1001005

Le Grice, S. F. (2015). Targeting the HIV RNA genome: high-hanging fruit only needs a longer ladder. *Curr. Top. Microbiol. Immunol.* 389, 147–169. doi: 10.1007/82_2015_434

Leamy, K. A., Assmann, S. M., Mathews, D. H., and Bevilacqua, P. C. (2016). Bridging the gap between *in vitro* and *in vivo* RNA folding. *Q. Rev. Biophys.* 49:e10. doi: 10.1017/S003358351600007X

Lee, K.-M., Chen, C.-J., and Shih, S.-R. (2017). Regulation mechanisms of viral IRES-driven translation. *Trends Microbiol.* 25, 546–561. doi: 10.1016/j.tim.2017.01.010

Lee, M.-K., Kim, H.-E., Park, E.-B., Lee, J., Kim, K.-H., Lim, K., et al. (2016). Structural features of influenza A virus panhandle RNA enabling the activation of RIG-I independently of 5′-triphosphate. *Nucleic Acids Res.* 44, 8407–8416. doi: 10.1093/nar/gkw525

Li, H., Havens, W. M., Nibert, M. L., and Ghabrial, S. A. (2015). An RNA cassette from Helminthosporium victoriae virus 190S necessary and sufficient for stop/restart translation. *Virology* 474, 131–143. doi: 10.1016/j.virol.2014.10.022

Lilley, D. M., Clegg, R. M., Diekmann, S., Seeman, N. C., Von Kitzing, E., and Hagerman, P. J. (1995). A nomenclature of junctions and branchpoints in nucleic acids. *Nucleic Acids Res.* 23, 3363–3364. doi: 10.1093/nar/23.17.3363

Lim, C. S., and Brown, C. M. (2016). Hepatitis B virus nuclear export elements: RNA stem-loop α and β, key parts of the HBV post-transcriptional regulatory element. *RNA Biol.* 13, 743–747. doi: 10.1080/15476286.2016.1166330

Lim, F., and Peabody, D. S. (1994). Mutations that increase the affinity of a translational repressor for RNA. *Nucleic Acids Res.* 22, 3748–3752. doi: 10.1093/nar/22.18.3748

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* 227, 1435–1441. doi: 10.1126/science.2983426

Liu, Y., Wimmer, E., and Paul, A. V. (2009). Cis-acting RNA elements in human and animal plus-strand RNA viruses. *Biochim. Biophys. Acta* 1789, 495–517. doi: 10.1016/j.bbagrm.2009.09.007

Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26. doi: 10.1186/1748-7188-6-26

Lorenz, R., Wolfinger, M. T., Tanzer, A., and Hofacker, I. L. (2016). Predicting RNA secondary structures from sequence and probing data. *Methods* 103, 86–98. doi: 10.1016/j.ymeth.2016.04.004

Löytynoja, A., and Goldman, N. (2010). webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* 11:579. doi: 10.1186/1471-2105-11-579

Lozano, G., Fernandez, N., and Martinez-Salas, E. (2016). Modeling three-dimensional structural motifs of Viral IRES. *J. Mol. Biol.* 428, 767–776. doi: 10.1016/j.jmb.2016.01.005

Madhugiri, R., Fricke, M., Marz, M., and Ziebuhr, J. (2016). Coronavirus cis-Acting RNA Elements. *Adv. Virus Res.* 96, 127–163. doi: 10.1016/bs.aivir.2016.08.007

Magnus, M., Boniecki, M. J., Dawson, W., and Bujnicki, J. M. (2016). SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Res.* 44, W315–W319. doi: 10.1093/nar/gkw279

Mailler, E., Bernacchi, S., Marquet, R., Paillart, J.-C., Vivet-Boudou, V., and Smyth, R. P. (2016). The life-cycle of the HIV-1 Gag-RNA complex. *Viruses* 8:E248. doi: 10.3390/v8090248

Mandl, C. W., Holzmann, H., Kunz, C., and Heinz, F. X. (1993). Complete genomic sequence of Powassan virus: evaluation of genetic elements in tick-borne versus mosquito-borne flaviviruses. *Virology* 194, 173–184. doi: 10.1006/viro.1993.1247

Manzano, M., Reichert, E. D., Polo, S., Falgout, B., Kasprzak, W., Shapiro, B. A., et al. (2011). Identification of cis-acting elements in the 3′-untranslated region of the dengue virus type 2 RNA that modulate translation and replication. *J. Biol. Chem.* 286, 22521–22534. doi: 10.1074/jbc.M111.234302

Marz, M., Beerenwinkel, N., Drosten, C., Fricke, M., Frishman, D., Hofacker, I. L., et al. (2014). Challenges in RNA virus bioinformatics. *Bioinformatics* 30, 1793–1799. doi: 10.1093/bioinformatics/btu105

Mathew, S. F., Crowe-Mcauliffe, C., Graves, R., Cardno, T. S., Mckinney, C., Poole, E. S., et al. (2015). The highly conserved codon following the slippery sequence supports−1 frameshift efficiency at the HIV-1 frameshift site. *PLoS ONE* 10:e0122176. doi: 10.1371/journal.pone.0122176

Mathews, D. H., and Turner, D. H. (2002). Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.* 317, 191–203. doi: 10.1006/jmbi.2001.5351

Mathews, D. H., Burkard, M. E., Freier, S. M., Wyatt, J. R., and Turner, D. H. (1999). Predicting oligonucleotide affinity to nucleic acid targets. *RNA* 5, 1458–1469. doi: 10.1017/S1355838299991148

Mccormack, J. C., Yuan, X., Yingling, Y. G., Kasprzak, W., Zamora, R. E., Shapiro, B. A., et al. (2008). Structural domains within the 3' untranslated region of Turnip crinkle virus. *J. Virol.* 82, 8706–8720. doi: 10.1128/JVI.00416-08

Mcwilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., et al. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* 41, W597–W600. doi: 10.1093/nar/gkt376

Melchers, W. J., Hoenderop, J. G., Bruins Slot, H. J., Pleij, C. W., Pilipenko, E. V., Agol, V. I., et al. (1997). Kissing of the two predominant hairpin loops in the coxsackie B virus 3′ untranslated region is the essential structural feature of the origin of replication required for negative-strand RNA synthesis. *J. Virol.* 71, 686–696.

Melian, E. B., Hinzman, E., Nagasaki, T., Firth, A. E., Wills, N. M., Nouwens, A. S., et al. (2009). NS1' of Flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. *J. Virol.* 84, 1641–1647. doi: 10.1128/JVI.01979-09

Meyer, I. M. (2017). In silico methods for co-transcriptional RNA secondary structure prediction and for investigating alternative RNA structure expression. *Methods* 120, 3–16. doi: 10.1016/j.ymeth.2017.04.009

Miao, Z., Adamiak, R. W., Antczak, M., Batey, R. T., Becka, A. J., Biesiada, M., et al. (2017). RNA-puzzles round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* 23, 655–672. doi: 10.1261/rna.060368.116

Miao, Z., and Westhof, E. (2017). RNA Structure: advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* 46, 483–503. doi: 10.1146/annurev-biophys-070816-034125

Miller, W. A., Jackson, J., and Feng, Y. (2015). Cis- and trans-regulation of luteovirus gene expression by the 3′ end of the viral genome. *Virus Res.* 206, 37–45. doi: 10.1016/j.virusres.2015.03.009

Miller, W. A., Waterhouse, P. M., and Gerlach, W. L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA. *Nucleic Acids Res.* 16, 6097–6111. doi: 10.1093/nar/16.13.6097

Miras, M., Miller, W. A., Truniger, V., and Aranda, M. A. (2017). Non-canonical Translation in Plant RNA Viruses. *Front. Plant Sci.* 8:494. doi: 10.3389/fpls.2017.00494

Miyazaki, Y., Irobalieva, R. N., Tolbert, B. S., Smalls-Mantey, A., Iyalla, K., Loeliger, K., et al. (2010). Structure of a conserved retroviral RNA packaging element by NMR spectroscopy and cryo-electron tomography. *J. Mol. Biol.* 404, 751–772. doi: 10.1016/j.jmb.2010.09.009

Mokrejs, M., Masek, T., Vopálensky, V., Hlubucek, P., Delbos, P., and Pospísek, M. (2010). IRESite–a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res.* 38, D131–D136. doi: 10.1093/nar/gkp981

Moomau, C., Musalgaonkar, S., Khan, Y. A., Jones, J. E., and Dinman, J. D. (2016). Structural and functional characterization of programmed ribosomal frameshift signals in West Nile virus strains reveals high structural plasticity among cis-acting RNA elements. *J. Biol. Chem.* 291, 15788–15795. doi: 10.1074/jbc.M116.735613

Moss, W. N., Dela-Moss, L. I., Kierzek, E., Kierzek, R., Priore, S. F., and Turner, D. H. (2012a). The 3′ splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. *PLoS ONE* 7:e38323. doi: 10.1371/journal.pone.0038323

Moss, W. N., Dela-Moss, L. I., Priore, S. F., and Turner, D. (2012b). The influenza A segment 7 mRNA 3′ splice site pseudoknot/hairpin family. *RNA Biol.* 9, 1305–1310. doi: 10.4161/rna.22343.

Moss, W. N., Priore, S. F., and Turner, D. H. (2011). Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA* 17, 991–1011. doi: 10.1261/rna.2619511

Napthine, S., Lever, R. A., Powell, M. L., Jackson, R. J., Brown, T. D. K., and Brierley, I. (2009). Expression of the VP2 protein of murine norovirus by a translation termination-reinitiation strategy. *PLoS ONE* 4:e8390. doi: 10.1371/journal.pone.0008390

Napthine, S., Treffers, E. E., Bell, S., Goodfellow, I., Fang, Y., Firth, A. E., et al. (2016). A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Res.* 44, 5491–5503. doi: 10.1093/nar/gkw480

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 43, D130–D137. doi: 10.1093/nar/gku1063.

Newburn, L. R., and White, K. A. (2015). Cis-acting RNA elements in positive-strand RNA plant virus genomes. *Virology* 479–480, 434–443. doi: 10.1016/j.virol.2015.02.032

Nicholson, B. L., and White, K. A. (2014). Functional long-range RNA–RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.* 12, 493–504. doi: 10.1038/nrmicro3288

Nicholson, B. L., and White, K. A. (2015). Exploring the architecture of viral RNA genomes. *Curr. Opin. Virol.* 12, 66–74. doi: 10.1016/j.coviro.2015.03.018

Niu, S., Cao, S., and Wong, S.-M. (2014). An infectious RNA with a hepta-adenosine stretch responsible for programmed−1 ribosomal frameshift derived from a full-length cDNA clone of Hibiscus latent Singapore virus. *Virology* 449, 229–234. doi: 10.1016/j.virol.2013.11.021

Nussinov, R., Pieczenik, G., Griggs, J. R., and Kleitman, D. J. (1978). Algorithms for loop matchings. *SIAM J. Appl. Math.* 35, 68–82. doi: 10.1137/0135006

O'leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., Mcveigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Olsthoorn, R. C., and Bol, J. F. (2001). Sequence comparison and secondary structure analysis of the 3′ noncoding region of flavivirus genomes reveals multiple pseudoknots. *RNA* 7, 1370–1377.

Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004). Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* 2, 461–472. doi: 10.1038/nrmicro903

Panjaworayan, N., and Brown, C. M. (2011). Effects of HBV genetic variability on RNAi strategies. *Hepat. Res. Treat.* 2011:367908. doi: 10.1155/2011/367908

Patel, N., Wroblewski, E., Leonov, G., Phillips, S. E. V., Tuma, R., Twarock, R., et al. (2017). Rewriting nature's assembly manual for a ssRNA virus. *Proc. Natl. Acad. Sci. U.S.A.* 114, 12255–12260. doi: 10.1073/pnas.1706951114

Paul, C. P., Barry, J. K., Dinesh-Kumar, S. P., Brault, V., and Miller, W. A. (2001). A sequence required for −1 ribosomal frameshifting located four kilobases downstream of the frameshift site 1 1Edited by D. E. Draper. *J. Mol. Biol.* 310, 987–999. doi: 10.1006/jmbi.2001.4801

Peabody, D. S. (1990). Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein-RNA interaction. *J. Biol. Chem.* 265, 5684–5689.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* 85, 2444–2448. doi: 10.1073/pnas.85.8.2444

Petrov, A. I., Zirbel, C. L., and Leontis, N. B. (2013). Automated classification of RNA 3D motifs and the RNA 3D motif atlas. *RNA* 19, 1327–1340. doi: 10.1261/rna.039438.113

Phan, A., Mailey, K., Saeki, J., Gu, X., and Schroeder, S. J. (2017). Advancing viral RNA structure prediction: measuring the thermodynamics of pyrimidine-rich internal loops. *RNA* 23, 770–781. doi: 10.1261/rna.059865.116

Pijlman, G. P., Funk, A., Kondratieva, N., Leung, J., Torres, S., van der Aa, L., et al. (2008). A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* 4, 579–591. doi: 10.1016/j.chom.2008.10.007

Pilipenko, E. V., Maslova, S. V., Sinyakov, A. N., and Agol, V. I. (1992). Towards identification of cis-acting elements involved in the replication of enterovirus and rhinovirus RNAs: a proposal for the existence of tRNA-like terminal structures. *Nucleic Acids Res.* 20, 1739–1745. doi: 10.1093/nar/20.7.1739

Poidinger, M., Hall, R. A., and Mackenzie, J. S. (1996). Molecular characterization of the Japanese encephalitis serocomplex of the flavivirus genus. *Virology* 218, 417–421. doi: 10.1006/viro.1996.0213

Pooggin, M. M., Rajeswaran, R., Schepetilnikov, M. V., and Ryabova, L. A. (2012). Short ORF-dependent ribosome shunting operates in an RNA picorna-like virus and a DNA pararetrovirus that cause rice Tungro disease. *PLoS Pathog.* 8:e1002568. doi: 10.1371/journal.ppat.1002568

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., et al. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* 40:e112. doi: 10.1093/nar/gks339

Powell, M. L. (2010). Translational termination–reinitiation in RNA viruses. *Biochem. Soc. Trans.* 38, 1558–1564. doi: 10.1042/BST0381558

Priore, S. F., Kauffmann, A. D., Baman, J. R., and Turner, D. H. (2015). The influenza A PB1-F2 and N40 start codons are contained within an RNA pseudoknot. *Biochemistry* 54, 3413–3415. doi: 10.1021/bi501564d

Prostova, M. A., Deviatkin, A. A., Tcelykh, I. O., Lukashev, A. N., and Gmyl, A. P. (2017). Independent evolution of tetraloop in enterovirus oriL replicative element and its putative binding partners in virus protein 3C. *PeerJ* 5:e3896. doi: 10.7717/peerj.3896

Proutski, V., Gould, E. A., and Holmes, E. C. (1997). Secondary structure of the 3' untranslated region of flaviviruses: similarities and differences. *Nucleic Acids Res.* 25, 1194–1202. doi: 10.1093/nar/25.6.1194

Puton, T., Kozlowski, L. P., Rother, K. M., and Bujnicki, J. M. (2013). CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Res.* 41, 4307–4323. doi: 10.1093/nar/gkt101

Rackham, O., and Brown, C. M. (2004). Visualization of RNA–protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J.* 23, 3346–3355. doi: 10.1038/sj.emboj.7600341

Rauscher, S., Flamm, C., Mandl, C. W., Heinz, F. X., and Stadler, P. F. (1997). Secondary structure of the 3'-noncoding region of flavivirus genomes: comparative analysis of base pairing probabilities. *RNA* 3, 779–791.

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA* 10, 1507–1517. doi: 10.1261/rna.5248604

Rivas, E., and Eddy, S. R. (1999). A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* 285, 2053–2068. doi: 10.1006/jmbi.1998.2436

Rivas, E., Clements, J., and Eddy, S. R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* 14, 45–48. doi: 10.1038/nmeth.4066

Rognes, T. (2011). Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. *BMC Bioinformatics* 12:221. doi: 10.1186/1471-2105-12-221

Roll, J., Zirbel, C. L., Sweeney, B., Petrov, A. I., and Leontis, N. (2016). JAR3D Webserver: scoring and aligning RNA loop sequences to known 3D motifs. *Nucleic Acids Res.* 44, W320–W327. doi: 10.1093/nar/gkw453

Romero-López, C., and Berzal-Herranz, A. (2013). Unmasking the information encoded as structural motifs of viral RNA genomes: a potential antiviral target. *Rev. Med. Virol.* 23, 340–354. doi: 10.1002/rmv.1756

Sagan, S. M., Chahal, J., and Sarnow, P. (2015). cis-Acting RNA elements in the hepatitis C virus RNA genome. *Virus Res.* 206, 90–98. doi: 10.1016/j.virusres.2014.12.029

Schepetilnikov, M., Schott, G., Katsarou, K., Thiébeauld, O., Keller, M., and Ryabova, L. A. (2009). Molecular dissection of the prototype foamy virus (PFV) RNA 5'-UTR identifies essential elements of a ribosomal shunt. *Nucleic Acids Res.* 37, 5838–5847. doi: 10.1093/nar/gkp609

Sealfon, R. S., Lin, M. F., Jungreis, I., Wolf, M. Y., Kellis, M., and Sabeti, P. C. (2015). FRESCo: finding regions of excess synonymous constraint in diverse viruses. *Genome Biol.* 16:38. doi: 10.1186/s13059-015-0603-7

Shi, P. Y., Brinton, M. A., Veal, J. M., Zhong, Y. Y., and Wilson, W. D. (1996). Evidence for the existence of a pseudoknot structure at the 3' terminus of the flavivirus genomic RNA. *Biochemistry* 35, 4222–4230. doi: 10.1021/bi952398v

Shi, Y., Wu, Y., Zhang, W., Qi, J., and Gao, G. F. (2014). Enabling the "host jump": structural determinants of receptor-binding specificity in influenza A viruses. *Nat. Rev. Microbiol.* 12, 822–831. doi: 10.1038/nrmicro3362

Simmonds, P. (2012). SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res. Notes* 5:50. doi: 10.1186/1756-0500-5-50

Simon, A. E., and Miller, W. A. (2013). 3' cap-independent translation enhancers of plant viruses. *Annu. Rev. Microbiol.* 67, 21–42. doi: 10.1146/annurev-micro-092412-155609

Sloma, M. F., and Mathews, D. H. (2016). Exact calculation of loop formation probability identifies folding motifs in RNA secondary structures. *RNA* 22, 1808–1818. doi: 10.1261/rna.053694.115

Smith, C., Heyne, S., Richter, A. S., Will, S., and Backofen, R. (2010). Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res.* 38, W373–W377. doi: 10.1093/nar/gkq316

Soszynska-Jozwiak, M., Michalak, P., Moss, W. N., Kierzek, R., and Kierzek, E. (2015). A conserved secondary structural element in the coding region of the influenza A Virus Nucleoprotein (NP) mRNA is important for the regulation of viral proliferation. *PLoS ONE* 10:e0141132. doi: 10.1371/journal.pone.0141132

Sperschneider, J., and Datta, A. (2010). DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. *Nucleic Acids Res.* 38, e103–e103. doi: 10.1093/nar/gkq011

Starck, S. R., Tsai, J. C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., et al. (2016). Translation from the 5' untranslated region shapes the integrated stress response. *Science* 351:aad3867. doi: 10.1126/science.aad3867

Stockley, P. G., Stonehouse, N. J., Murray, J. B., Goodman, S. T., Talbot, S. J., Adams, C. J., et al. (1995). Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein. *Nucleic Acids Res.* 23, 2512–2518. doi: 10.1093/nar/23.13.2512

Stombaugh, J., Zirbel, C. L., Westhof, E., and Leontis, N. B. (2009). Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* 37, 2294–2312. doi: 10.1093/nar/gkp011

Su, L., Chen, L., Egli, M., Berger, J. M., and Rich, A. (1999). Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Biol.* 6, 285–292. doi: 10.1038/6722

Sun, J., Yu, Y., and Deubel, V. (2012). Japanese encephalitis virus NS1' protein depends on pseudoknot secondary structure and is cleaved by caspase during virus infection and cell apoptosis. *Microbes Infect.* 14, 930–940. doi: 10.1016/j.micinf.2012.03.007

Sun, L.-Z., Zhang, D., and Chen, S.-J. (2017). Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annu. Rev. Biophys.* 46, 227–246. doi: 10.1146/annurev-biophys-070816-033920

Tajima, Y., Iwakawa, H.-O., Kaido, M., Mise, K., and Okuno, T. (2011). A long-distance RNA-RNA interaction plays an important role in programmed−1 ribosomal frameshifting in the translation of p88 replicase protein of Red clover necrotic mosaic virus. *Virology* 417, 169–178. doi: 10.1016/j.virol.2011.05.012

Taly, J.-F., Magis, C., Bussotti, G., Chang, J.-M., Di Tommaso, P., Erb, I., et al. (2011). Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat. Protoc.* 6, 1669–1682. doi: 10.1038/nprot.2011.393

Taylor, W. R., and Hamilton, R. S. (2017). Exploring RNA conformational space under sparse distance restraints. *Sci. Rep.* 7:44074. doi: 10.1038/srep44074

Theis, C., Janssen, S., and Giegerich, R. (2010). "Prediction of RNA secondary structure including kissing hairpin motifs," in *Lecture Notes in Computer Science*, eds V. Moulton and M. Singh (Berlin; Heidelberg: Springer-Verlag), 52–64.

Theis, C., Reeder, J., and Giegerich, R. (2008). KnotInFrame: prediction of −1 ribosomal frameshift events. *Nucleic Acids Res.* 36, 6013–6020. doi: 10.1093/nar/gkn578

Theis, C., Zirbel, C. L., Zu Siederdissen, C. H., Anthon, C., Hofacker, I. L., Nielsen, H., et al. (2015). RNA 3D modules in genome-wide predictions of RNA 2D structure. *PLoS ONE* 10:e0139900. doi: 10.1371/journal.pone.0139900

Torarinsson, E., and Lindgreen, S. (2008). WAR: webserver for aligning structural RNAs. *Nucleic Acids Res.* 36, W79–W84. doi: 10.1093/nar/gkn275

Tuplin, A., Wood, J., Evans, D. J., Patel, A. H., and Simmonds, P. (2002). Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* 8, 824–841. doi: 10.1017/S1355838202554066

Turner, D. H., and Mathews, D. H. (2016). *RNA Structure Determination: Methods and Protocols.* New York, NY: Springer.

Umu, S. U., and Gardner, P. P. (2016). A comprehensive benchmark of RNA–RNA interaction prediction tools for all domains of life. *Bioinformatics* 33, 988–996. doi: 10.1093/bioinformatics/btw728

Villordo, S. M., Carballeda, J. M., Filomatori, C. V., and Gamarnik, A. V. (2016). RNA structure duplications and flavivirus host adaptation. *Trends Microbiol.* 24, 270–283. doi: 10.1016/j.tim.2016.01.002

Villordo, S. M., Filomatori, C. V., Sánchez-Vargas, I., Blair, C. D., and Gamarnik, A. V. (2015). Dengue virus RNA structure specialization facilitates host adaptation. *PLoS Pathog.* 11:e1004604. doi: 10.1371/journal.ppat.1004604

Wallner, G., Mandl, C. W., Kunz, C., and Heinz, F. X. (1995). The flavivirus 3'-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. *Virology* 213, 169–178. doi: 10.1006/viro.1995.1557

Walsh, D., and Mohr, I. (2011). Viral subversion of the host protein synthesis machinery. *Nat. Rev. Microbiol.* 9, 860–875. doi: 10.1038/nrmicro2655

Wang, J., Mao, K., Zhao, Y., Zeng, C., Xiang, J., Zhang, Y., et al. (2017). Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. *Nucleic Acids Res.* 45, 6299–6309. doi: 10.1093/nar/gkx386

Wang, Z., Parisien, M., Scheets, K., and Miller, W. A. (2011). The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. *Structure* 19, 868–880. doi: 10.1016/j.str.2011.03.013

Ward, A. M., Bidet, K., Yinglin, A., Ler, S. G., Hogue, K., Blackstock, W., et al. (2011). Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3′ UTR structures. *RNA Biol.* 8, 1173–1186. doi: 10.4161/rna.8.6. 17836

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/ btp033

Wattenberg, M. (2002). "Arc diagrams: visualizing structure in strings," in *IEEE Symposium on Information Visualization, 2002* (Boston, MA: INFOVIS).

Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W. Jr., Swanstrom, R., et al. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460, 711–716. doi: 10.1038/nature 08237

Weeks, K. M. (2015). Review toward all RNA structures, concisely. *Biopolymers* 103, 438–448. doi: 10.1002/bip.22601

Whitehead, S. S., Blaney, J. E., Durbin, A. P., and Murphy, B. R. (2007). Prospects for a dengue virus vaccine. *Nat. Rev. Microbiol.* 5, 518–528. doi: 10.1038/nrmicro1690

Xu, Z. Z., and Mathews, D. H. (2016). Prediction of secondary structures conserved in multiple RNA sequences. *Methods Mol. Biol.* 1490, 35–50. doi: 10.1007/978-1-4939-6433-8_3

Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W., and Chen, X. (2014). Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. *Bioengineered* 5, 300–304. doi: 10.4161/bioe.32110

Yao, Z., Weinberg, Z., and Ruzzo, W. L. (2006). CMfinder–a covariance model based RNA motif finding algorithm. *Bioinformatics* 22, 445–452. doi: 10.1093/bioinformatics/btk008

You, S., and Rice, C. M. (2008). 3′ RNA elements in hepatitis C virus replication: kissing partners and long poly(U). *J. Virol.* 82, 184–195. doi: 10.1128/JVI.01796-07

Zhang, B., Dong, H., Stein, D. A., and Shi, P.-Y. (2008). Co-selection of West Nile virus nucleotides that confer resistance to an antisense oligomer while maintaining long-distance RNA/RNA base pairings. *Virology* 382, 98–106. doi: 10.1016/j.virol.2008.08.044

Zhang, M., Perelson, A. S., and Tung, C.-S. (2011). *RNA Structural Motifs.* Chichester, UK: eLS.

Zhang, W., Thieme, C. J., Kollwig, G., Apelt, F., Yang, L., Winter, N., et al. (2016). tRNA-related sequences trigger systemic mRNA transport in plants. *Plant Cell* 28, 1237–1249. doi: 10.1105/tpc.15.01056

Zhao, Y., Huang, Y., Gong, Z., Wang, Y., Man, J., and Xiao, Y. (2012). Automated and fast building of three-dimensional RNA structures. *Sci. Rep.* 2:734. doi: 10.1038/srep00734

Zhu, J. Y., and Meyer, I. M. (2015). Four RNA families with functional transient structures. *RNA Biol.* 12, 5–20. doi: 10.1080/15476286.2015.1008373

Zinoviev, A., Hellen, C. U. T., and Pestova, T. V. (2015). Multiple mechanisms of reinitiation on bicistronic calicivirus mRNAs. *Mol. Cell* 57, 1059–1073. doi: 10.1016/j.molcel.2015.01.039

Zoll, J., Heus, H. A., van Kuppeveld, F. J. M., and Melchers, W. J. G. (2009). The structure-function relationship of the enterovirus 3′-UTR. *Virus Res.* 139, 209–216. doi: 10.1016/j.virusres.2008.07.014

Zong, L., Qin, Y., Jia, H., Ye, L., Wang, Y., Zhang, J., et al. (2017). Differential regulation of hepatitis B virus core protein expression and genome replication by a small upstream open reading frame and naturally occurring mutations in the precore region. *Virology* 505, 155–161. doi: 10.1016/j.virol.2017.02.020

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* 244, 48–52. doi: 10.1126/science.2468181

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595