Siblings or doppelgängers? Deciphering the evolution of structured cis-regulatory RNAs beyond homology

Elizabeth C. Gray*, Daniel M. Beringer* and ⓘ Michelle M. Meyer

Department of Biology, Boston College, Chestnut Hill, MA 02467, U.S.A.

Correspondence: Michelle M. Meyer (m.meyer@bc.edu)

Structured cis-regulatory RNAs have evolved across all domains of life, highlighting the utility and plasticity of RNA as a regulatory molecule. Homologous RNA sequences and structures often have similar functions, but homology may also be deceiving. The challenges that derive from trying to assign function to structure and vice versa are not trivial. Bacterial riboswitches, viral and eukaryotic IRESes, CITEs, and 3′ UTR elements employ an array of mechanisms to exert their effects. Bioinformatic searches coupled with biochemical and functional validation have elucidated some shared and many unique ways cis-regulators are employed in mRNA transcripts. As cis-regulatory RNAs are resolved in greater detail, it is increasingly apparent that shared homology can mask the full spectrum of mRNA cis-regulator functional diversity. Furthermore, similar functions may be obscured by lack of obvious sequence similarity. Thus looking beyond homology is crucial for furthering our understanding of RNA-based regulation.

Introduction

Cis-regulatory RNA elements are structured regions of an mRNA that regulate the transcription, translational efficiency, or stability of the mRNA and are found throughout all domains of life. While most frequently found in the 5′ untranslated region (UTR), structured RNA cis-regulatory elements may be found throughout the transcript with diverse mechanisms in prokaryotic (Figure 1A,B) and eukaryotic systems (Figure 1C–E). RNA cis-regulators frequently control functions that require swift changes such as cellular stress response or, in the case of pathogens, host switching. In contrast with DNA or protein-coding elements, the primary sequence conservation of cis-regulatory RNA elements can be extremely low because the secondary structure or folded structure is often more highly conserved than the primary sequence. However, just as with DNA or protein-coding elements, identifying apparent homology is still an integral component of the process for connecting growing sequence databases with biological functions. In recent years, improvements to computational methods have made identifying new and homologous cis-regulatory RNA elements easier. Yet, due to the unique properties of structured RNA, the use of homology still has some serious limitations. In this review, we highlight a series of vignettes that illustrate limitations of using homology to find or predict cis-regulatory RNA element function and discuss recent insights into the evolution and function of riboswitches, internal ribosome entry sites (IRESes) and 3′ UTR elements such as cap-independent translation elements (CITEs).
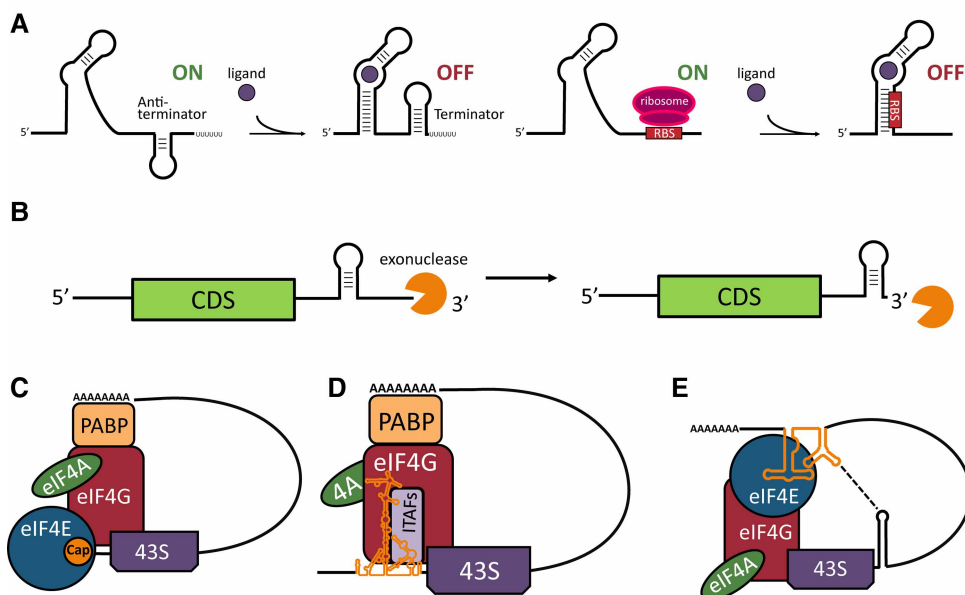
Riboswitches

A riboswitch is a cis-regulatory noncoding RNA element typically found in the 5′ UTR that bacteria employ to control stress response as well as flux through essential biosynthetic pathways. Consisting of a ligand-binding aptamer domain that controls a downstream expression platform, riboswitches comprise a diverse collection of sequences and structures that interact with a host of different cellular

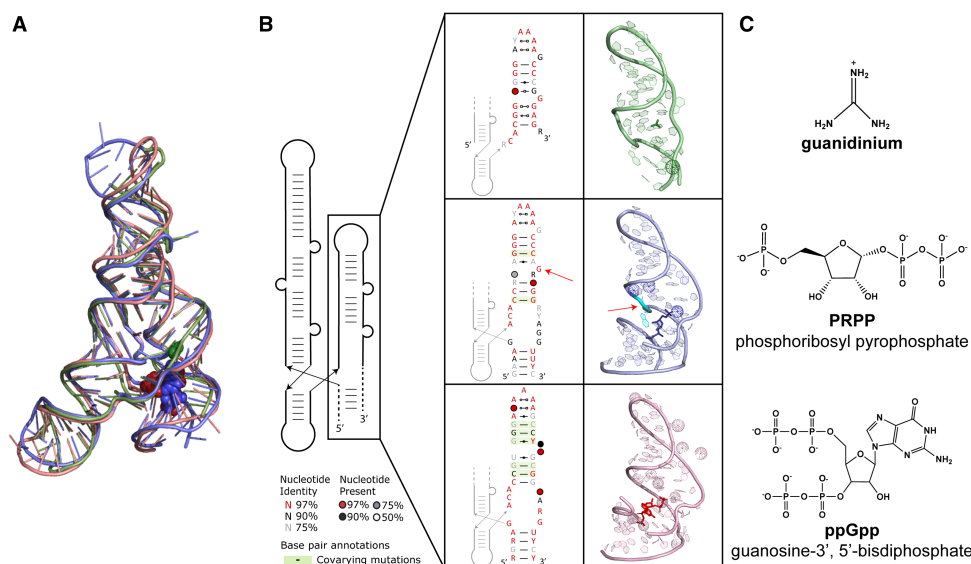*These authors contributed equally to this work.

Figure 1. Cis-regulatory RNAs in Bacteria and Eukaryotes.

(**A**) Riboswitches are RNA elements located in the 5′ UTR of many bacterial transcripts, regulating the expression of downstream genes through conformational changes induced by ligand binding. Riboswitches can act transcriptionally or translationally and depending on the changes in RNA structure induced by ligand binding may act as ON or OFF switches. (**B**) Bacterial 3′-UTRs often contain stem-loop structures which are a result of Rho-independent transcription termination, but are also present in Rho-dependent transcripts [57], suggesting a conserved mechanism where these structures protect both Rho-dependent and independent transcripts from 3′–5′ exonucleases. Cap-dependent (**C**), Type-II IRES-mediated (**D**) and CITE-mediated (**E**) Eukaryotic Translation Initiation Complex. In the absence of a 5′ cap, RNA structures in the 5′- or 3′-UTR (gold) can recruit initiation factors required for translation. Exact mechanisms and initiation factors required depend on the specific IRES or CITE. Specifically shown here are the encephalomyocarditis virus IRES [58] and the panicum mosaic virus-like CITE with the kissing stem-loop [51].

ligands and employ diverse mechanisms of action such as transcription termination and translation inhibition (Figure 1A) (for review see [1,2]). Most riboswitches were discovered through comparative genomic analyses to identify structured noncoding RNAs. The majority of motifs defined by such searches have properties suggesting regulatory activity, but as of yet have no identified triggering ligand [3]. Determining the correct ligand for a riboswitch candidate is typically aided by the functional characterization of the genes under its regulatory control. Riboswitches whose ligands were straightforward to identify have largely been associated with well-characterized metabolic pathways, such as coenzyme or amino acid biosynthesis. Riboswitch candidates whose ligands resist identification, collectively known as orphan riboswitches, are often associated with genes coding for proteins of unknown function, or genes for various proteins with no established link to one another. Even following ligand identification, homologous aptamer examples may utilize different mechanisms of regulation, or even have distinct ligand-binding preferences. However, experimental validation and structural determination for such examples has yielded important insights into the structural and mechanistic diversity employed by RNA to regulate gene expression.

Sequence and secondary structure similarity often suggest common ligands for homologous riboswitch aptamers, but the detailed biochemical characterization and subsequent three-dimensional structures can reveal minor sequence changes that lead to differences in ligand specificity. This phenomenon is exemplified by the *ykkC* riboswitches (Figure 2). Originally classified as a single type of riboswitch, the original *ykkC* aptamer evaded characterization for over a decade [4]. Furthermore, the discovery of non-homologous elements regulating similar sets of genes (mini-*ykkC* and *ykkC*-III) [5,6] only served to increase interest in these elements. Eventually, biochemical characterization subdivided the original *ykkC* aptamer into multiple sub-classes that recognize more than 5 distinct ligands: guanidine (*ykkC* subtype 1) [7], guanosine-3′, 5′-bisdiphosphate

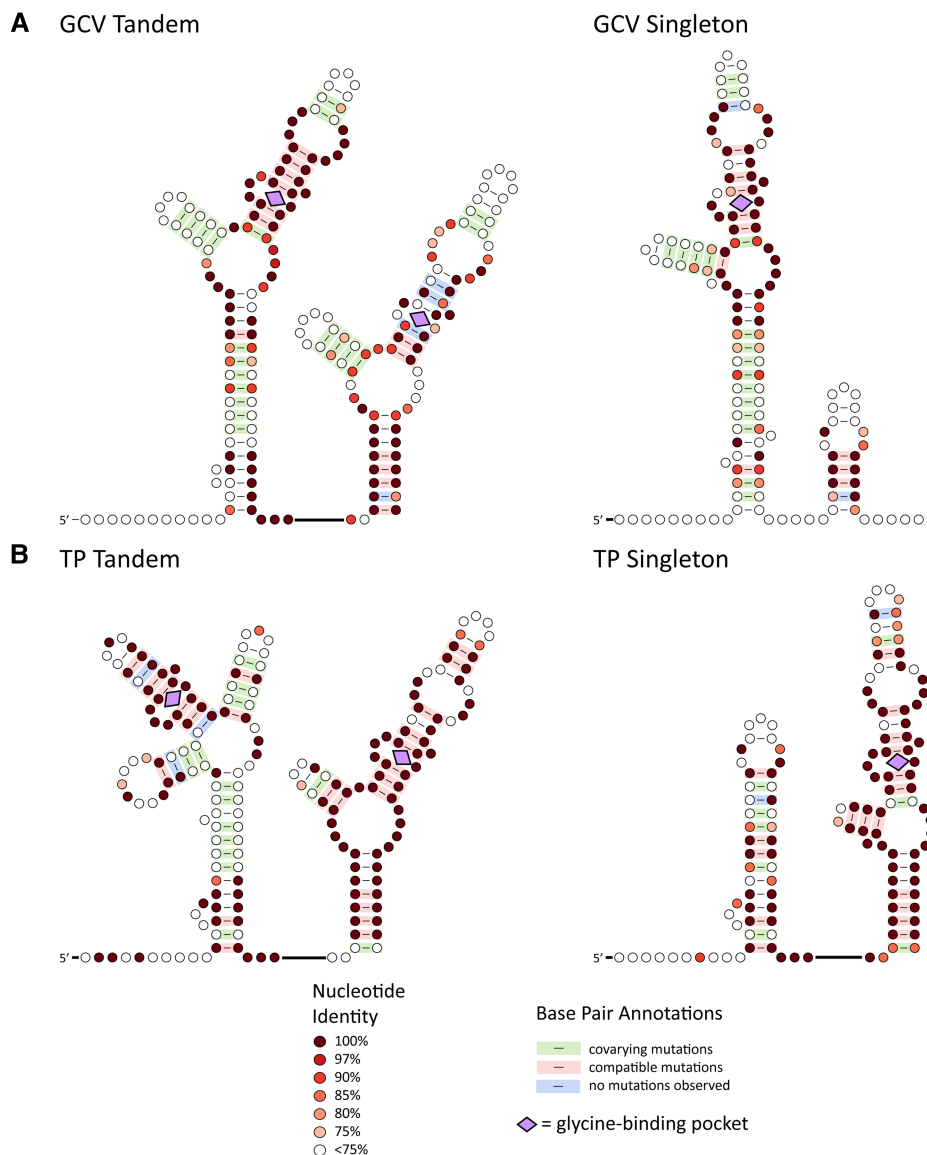**Figure 2.** *ykkC* **subtype 1, 2a, and 2b riboswitch comparison.**
(**A**) An overlay of the three riboswitches in the ligand-bound state shows nearly overlapping structural scaffolds. (**B**) The cartoon schematic of the *ykkC* riboswitch motif and subtypes highlights the shared structural core of the guanidine-I (top), PRPP (middle), and ppGpp (bottom) riboswitches. Ligand selectivity for PRPP and ppGpp is conferred by an additional helical element (dashed lines). The corresponding consensus secondary structures are adapted from [18] and [11]. Crystal structures are shown with co-ordinated Mg^{2+} ions as spheres and ligands as ball and stick models (pdb codes: 5T83, 6DLT, and 6DMC, from [8] and [18]). Assessing the individual active site structures shows that PRPP and ppGpp aptamers bind their ligands in overlapping positions, while the guanidine binds higher within the stem. The G96A point mutation (aqua, arrow) switches ligand specificity from PRPP to ppGpp. (**C**) Chemical structures of the ligands bound by the *ykkC* subtype 1, 2a, and 2b riboswitches.

(ppGpp) (*ykkC* subtype 2a), and phosphoribosyl pyrophosphate (PRPP) (*ykkC* subtype 2b) [8–11]. Further validation of the *ykkC* subtype 2c expanded the ligands bound by this motif to include adenosine- and cytidine 5′ diphosphates (in either their deoxyribose or ribose forms), while subtype 2d remains an orphan riboswitch whose ligand is unknown [12].

Subsequent structural data have demonstrated that the guanidine type I, ppGpp, and PRPP-binding riboswitches that appear homologous based on sequence and secondary structure do share a highly conserved structural core, but are able to distinguish between ligands specifically. In these riboswitches, the guanidine-binding site is above the site of PRPP and ppGpp binding (Figure 2A). Furthermore, ppGpp and PRPP are distinguished by a helical element revealed by structure determination [8]. The shared structural domain makes switching ligand specificity facile through changes to the ligand-binding helix. A single G96A point mutation switches the ligand affinity for the PRPP-binding aptamer to ppGpp with a 40 000-fold change in selectivity (Figure 2B) [11]. A single mutation to the structural core along with swapping the ligand-binding helix converts the PRPP aptamer to a guanidine aptamer [13]. The selective pressures driving the evolution of the guanidine, ppGpp, and PRPP-sensing *ykkC* riboswitch classes thus show how subtle sequence changes and modularity can lead to the adaptability of RNA-based regulation from a shared core structure.

Conversely, a common ligand may be recognized by aptamers with no obvious homology. Following the characterization of *ykkC* subtype 1, the mini-*ykkC* and *ykkC*-III elements were demonstrated to also be guanidine binders [14,15]. Recent crystal structures of the three non-homologous guanidine-binding riboswitches (*ykkC*, mini-*ykkC*, and *ykkC*-III) revealed some common elements but overall distinct structures that specifically interact with the same ligand, suggesting independent evolution of these elements [7,16–20]. These aptamers, and other similar examples such as the PreQ$_1$ [21–23] and S-adenosyl methionine (SAM) binding aptamers [24–28] highlight the flexibility and plasticity of RNA as a regulatory molecule.

While the *ykkC* aptamers make clear that RNA regulators preceding diverse sets of genes can evolve different specificities through subtle changes to shared core structures, divergent gene sets can also promote the selection

Figure 3. Genomic context drives conservation of the glycine riboswitches.

Heat maps of the consensus structures for the tandem and singlet glycine riboswitches regulating glycine cleavage metabolism (GCV) or glycine transport proteins (TP), with the glycine-binding pocket indicated (diamond), alignments used from [31]. (**A**) Aptamer 1 is more highly conserved in the tandem GCV riboswitch and type-1 singleton, while aptamer 2 degrades into a conserved ghost aptamer important for tertiary structure. (**B**) Conversely, aptamer 2 is more conserved in the tandem TP riboswitch and type-2 singleton, while aptamer 1 degrades into a conserved ghost aptamer.

of more drastic changes in aptamer structure that may not ultimately affect biological function significantly. This is best exemplified by the glycine riboswitch, which exists in both singlet and tandem conformations and can function as an ON or OFF switch, primarily regulating either glycine cleavage metabolism (GCV) or glycine transport (TP), respectively (Figure 3). A library of point mutations to the *Bacillus subtilis* glycine riboswitch showed that the first aptamer of the tandem ON switch (regulating glycine cleavage) is more essential for ligand binding and regulation [29,30]. Yet, extensive biochemical analysis of the tandem OFF switch (regulating transport) from *Vibrio cholerae* showed that ligand binding by the second aptamer is more important [30]. A recent computational analysis of glycine aptamers suggested that most singlet glycine riboswitches may derive from these tandem riboswitches [31]. In this scenario, the 'ghost' aptamer associated with most singlet

riboswitches is a degraded form of the less critical aptamer within a tandem aptamer pair. Singlet glycine ribo-switches are thus divided into type-1 switches if the ghost aptamer follows the ligand-binding aptamer (and most frequently precede glycine cleavage operons) or type-2 switches if the ghost aptamer precedes it (and most frequently precede transporters) [31,32]. The regulated genes appear to have driven whether the first or second aptamer is conserved, as the other aptamer degrades to a stem–loop ghost aptamer [30]. These studies of glycine riboswitch homologs give insights into the context-dependent selective pressures driving the evolution of the glycine riboswitch that likely apply to riboswitches more generally.

Although the vast majority of predicted cis-regulatory RNA structures remain uncharacterized, the *ykkC* and glycine riboswitches highlight the deeper understanding we can gain from detailed mechanistic and structural studies. Furthermore, we are now beginning to appreciate that a single riboswitch example does not necessarily represent the whole picture. Even aptamers binding the same ligand can operate co-transcriptionally or translationally and be driven by kinetic or thermodynamic mechanisms [33]. While homology is a good place to start, evolution yields multiple solutions to the same problem that may not share homology and effective regulatory solutions are frequently adapted for other purposes.
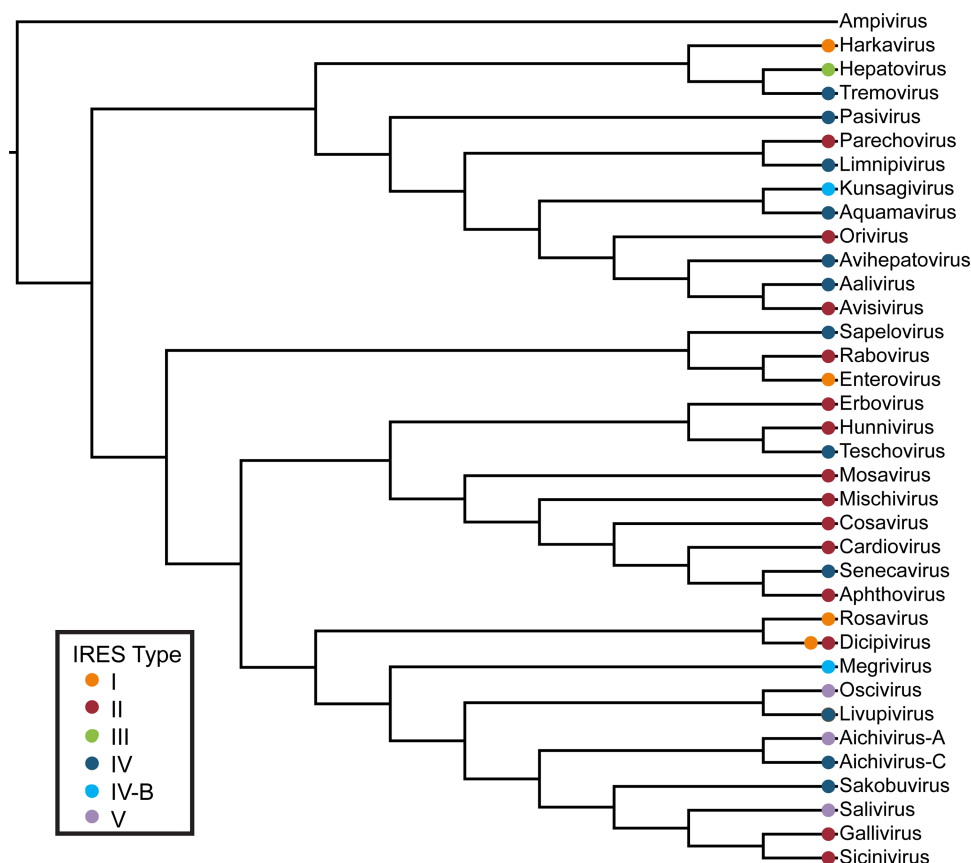
Internal ribosome entry sites (IRESes)

Unlike riboswitches, where close structural homologs may not have the same regulatory activity or bind the same ligand, internal ribosome entry sites (IRESes) are all functionally similar, but share little recognizable homology. IRESes are cis-regulatory RNA elements that recruit cellular translation machinery such as ribosomal-binding proteins to mRNAs for cap-independent translation (Figure 1C–E) and are found in eukaryotic and viral mRNA. In eukaryotes, although functional IRES sequences have been found and experimentally validated, there is very little primary or secondary conservation between these elements, and it is unclear how many transcripts have functional IRESes [34,35]. In contrast, viral IRESes do display some homology and are organized into types based on structural similarities; each type is often associated with specific viral clades [36]. However, some viruses contain a different IRES than other members of their clade, allowing IRES sequences to provide insight into viral evolution.

Lack of homology has made computational identification and comparison of eukaryotic IRESes difficult and controversial. A new method using comparative genomics and machine learning identified over 6 000 predicted IRESes in 20 fungal genomes [37]. Analysis of associated GO-terms showed IRESes predicted near genes involved in cell stress response. The conservation and distribution of the predicted IRESes supports the idea that cap-independent translation is an important part of the cell stress response. Another study used a fluorescent reporter method to identify thousands of viral and human sequences that could cause cap-independent translation, suggesting that this method of translation could be more widespread than previously thought [38]. Further investigation into methods to study eukaryotic IRESes that do not rely on sequence homology is a worthwhile investment and will likely lead to new insights into the evolutionary history of eukaryotic viruses.

In the past few years, crystal structures of picornavirus IRESes (which are generally categorized by the presence of a conserved tertiary structural core) have revealed surprising relationships between the different IRES types. Comparisons between the crystal structures of a type II and type III IRES found a conserved three-way junction critical for IRES-mediated translation [39]. This three-way junction may also be conserved in type I IRESes and bears similarity to a three-way junction in 3′ CITEs, another viral cis-regulator that allows for cap-independent translation. These examples reflect a potential role for the three-way junction in the recruitment of translation machinery. Similar studies of IRES crystal structures have also uncovered an alternative IRES strategy–tRNA mimicry [40]. Type IV picornavirus IRESes and the flavivirus hepatitis C IRES mimic the structure of the tRNA acceptor stem to directly bind the 40s subunit [40,41]. These recent studies of IRES structure highlight tertiary structure commonalities between IRESes lacking apparent primary sequence and secondary structural homology and provide a foundation for further understanding the molecular mechanisms and improved tool development of these curiously diverse cis-regulators.

While the type of IRES carried by a virus generally varies by clade, some viruses do not share the same type of IRES as the rest of their clade (Figure 4) and recent work suggests that IRESes are more frequently exchanged in comparison with other viral elements. In picornaviruses, a recent analysis of type IV IRESes found greater structural diversity outside of the core elements involved in ribosome binding and subdomain orientation [42]. Additionally, IRESes are often more varied between viruses that were otherwise more similar, suggesting that the IRESes spread through horizontal gene transfer and are further altered through recombination outside of the core subdomains required for function [42]. For example, pasiviruses (infecting swine)

Figure 4. Diversity of IRES type within Picornavirus clades.

Maximum likelihood cladogram of picornavirus genera with a defined IRES type was generated with the alignment of 3D^{pol} protein using RaxML [59,60]. IRES type (I–V) is indicated by the colored circle next to the genus name. *Kombuvirus* is listed as specific species (Aichivirus-A and -C) since there is more than one IRES type in the clade. *Dicipiviruses* contain multiple IRES types in a single genome. Distribution of IRES type supports the evidence that IRESes are exchanged between more distantly related species.

show strong similarity over the entire genome to parechoviruses (infecting humans), supporting a common origin and both belong to the picornavirus family. However, despite their resemblance, these viruses display significant differences in their IRES sequences, as the pasivirus IRES is the same type employed by swine pesti-viruses (family: *Flaviviridae*) [43]. A second example is illuminated by the discovery of a novel picornavirus clade that infects birds rather than mammals, Falcovirus A1 (Harkavirus). This virus contains a type of IRES previously only found in enteroviruses, suggesting that shuffling of IRESes between viruses may be facile [44]. These findings suggest that the 5′ region of the virus containing the IRES could be prone to recombination, perhaps reflecting adaptation to different hosts. In a practical application of IRES heterogeneity, pestivirus D (also known as border disease virus), which causes significant disease in livestock around the world, is differen-tiated from similar viruses and classified into different types based on the IRES secondary structure. Characterization of additional strains of the virus has expanded the number of known types from 8 to 10, and phenotypic analysis showed that types were separated geographically [45]. Work to determine functional differ-ences between IRES types may provide further insight into the evolutionary pressures driving IRES diversity.

Aside from a potential connection to altered host-specificity, cis-regulatory elements including IRESes can have important clinical implications for viruses causing disease. A recent phylogenetic study of Enterovirus D68 (a picornavirus) found that the 5′ end of the virus containing the IRES mutated over time. Experimentally, these IRESes were found to have increased activity in neuronal cell types, perhaps explaining an uptick in neurological symptoms of patients during this time [46]. This is similar to findings that another cis-regulatory element in HIV, the Rev response element (RRE), was found to evolve over time to become

more active through small changes to the RNA secondary structure [47]. These studies demonstrate the impact the evolution of cis-regulatory elements has on viral behavior in a clinical setting.

IRESes provide an example of structurally diverse cis-regulators that have largely the same function with similar mechanisms. Despite the similar function of all lRESes, structural and sequence heterogeneity has hindered identification efforts. Eukaryotic IRESes remain somewhat controversial because they lack a degree of similarity that is shared among even the divergent viral IRESes. Investigations into viral IRESes bring to light the importance of the IRES for successful host infection and suggest that changes to the IRES through horizontal transfer or nucleotide substitution may be driven by adaptation to new cell types or species. Increased understanding into the evolutionary drivers of IRESes in viruses and eukaryotes will continue to lead to better methods to identify IRESes and improve our understanding of the biological mechanisms of individual IRESes and IRES classes.

3′ Untranslated regions (UTRs)

Cis-regulatory elements in the 3′ UTR may have a profound effect on the translation of an mRNA, but the types of elements and mechanisms of action are heterologous and diverse. In eukaryotes, the 3′ UTR is well known to be involved in mRNA post-translational regulation, stability, and localization through binding regulatory proteins or microRNAs. Viral 3′ UTRs also regulate translation and stability, but can include additional regulators for specific viral functions like recoding and host switching. Conversely, the 3′ UTRs of bacterial transcripts have not been as widely explored. However, recent work suggests the bacterial 3′ UTR could contain a wealth of information about evolutionary history as well as underappreciated regulatory mechanisms.

Viral 3′ UTR elements affect viral replication and are often critical for successful proliferation inside a host, including host-specificity and host-switching. RNA thermometers (RNATs) involved in host-switching are well known in bacteria [48], but an RNAT with a similar host-switching function was recently identified in a flavivirus. According to researchers, West Nile virus has a 3′ RNAT to aid in the switching between cold-blooded and warm-blooded hosts [49]. This viral RNAT was found to alter the circularization rate, and thus the replication rate of the virus, allowing for persistent, low-level infection in insects and acute, high-level infection in warm-blooded hosts. To replicate effectively, plant viruses frequently use CITEs to initiate translation through binding host eIF4E (Figure 1E). Like IRESes, CITEs are organized into different types defined by the RNA secondary structure, but there is little obvious homology between these types. Despite the lack of homology, CITEs appear to function similarly as each class requires the formation of a kissing loop between the 3′ CITE and a stem-loop in the 5′ UTR as well as binding of eIF4F for efficient translation [50]. To test the robustness of these elements to diverse cellular backgrounds, researchers tested the ability of CITEs from nine different viruses to initiate translation in both plant and mammalian cell lines and for eIF4E binding to the CITE, since eIF4E is conserved between these kingdoms [51]. Results showed that while many of the CITEs did not initiate translation in mammalian cells, one CITE from thin paspalum asymptomatic virus could initiate translation in both plant and mammalian cells. Later results found binding between host eIF4E and the CITE and subsequent translation was determined by the presence of a guanosine-rich domain in the CITE pseudoknot. This finding suggests the potential for a similar element currently undiscovered in mammalian viruses. Indeed, a 3′ RNA element with structural similarity to putative plant virus CITEs was recently discovered in Sindbis Virus, a virus infecting insects [52]. While this element did not cause cap-independent translation, the 3′ element did confer host-specificity to the virus, allowing for translation of viral proteins in insect cells, but not mammalian cells. When this 3′ element was added onto a mammalian virus not known to infect insect cells, the virus was able to translate proteins and replicate within insect cells. This example demonstrates how homology may not be completely predictive of function, and how many viral changes are required for host switching.

Like IRES sequences, the use of 3′-elements to regulate cap-independent translation is by no means restricted to viruses. One example of native 3′ UTR impact on translation efficiency in eukaryotes is the transcription factor c-myc mRNA, which includes many cis-regulatory elements. c-myc mRNA is known to be translated in a cap-independent manner using an IRES if cap-dependent translation is inhibited. A recent study found that the activity of both cap-dependent and independent translation is increased by a 3′ element [53]. Thus, cis-regulatory RNA elements in eukaryotic 3′ UTRs may tune translation in eukaryotes to adjust to diverse cellular conditions.

The post-transcriptional regulatory role of 3′ UTRs in bacterial mRNAs is far less clear due to the tight coupling of transcription and translation. However, the 3′ UTR is known to affect mRNA stability [54], and changes in the sequence potentially lead to changes in expression. Comparative analysis of the 3′ UTRs of

orthologous genes across *Staphylococcus* species showed that conservation is lost downstream of the coding sequence, and that chimeric mRNAs with 3′ UTRs from orthologous genes showed different expression levels. Despite lack of homology across these elements, there are clearly important determinants for RNA stability conferred by these sequence changes. This suggests that divergent 3′ UTRs significantly alter gene expression levels and could thus be important for generating bacterial diversity [55]. Furthermore, technologies such as Term-Seq [56] have yielded insights into evolutionary differences between and within bacterial species affecting mRNA stability and transcript processing. An *in vivo* mapping of Rho-dependent transcripts in *E. coli* showed that the 3′ termini of essentially all protein-coding transcripts include stable structured RNA elements, protecting them from 3′–5′ exonucleases [57]. These 3′ UTR stem–loops resemble those classically associated with Rho-independent termination, but lack a polyU tract. This finding suggests that switching between termination mechanisms may occur facilely, depending only on the presence or absence of polyU sequences.

Conclusions

Structured cis-regulatory RNAs are found across mRNA transcripts and across all domains of life. While homology-based tools allow for the identification of putative RNA elements based on sequence and structure relatively easily, assigning biological function is much more laborious. Mechanistic and validation studies have greatly expanded our understanding of the novel roles cis-regulators play. However, we are now beginning to appreciate that one example may not be representative of the whole. Cis-regulators that have evolved are increasingly found to be more unique than initially presumed, with diverse properties driven by a host of selective pressures. Only following characterization of homologous RNAs can we begin to distinguish them as siblings, doppelgängers, or distant relatives. The examples outlined in this review emphasize that while homology can be a promising starting point, it can also be a restrictive paradigm. Growing beyond homology in our understanding of and search for new regulators will likely reveal yet unexplored aspects of RNA cis-regulatory elements.

Perspectives

- Structured cis-regulatory RNAs are important regulators found throughout the domains of life, but identification efforts have been stymied by lack of primary sequence and sometimes secondary structure homology.

- Cis-regulatory RNAs have been classically identified through homology-based bioinformatic searches and validation, but new discoveries from structural and mechanistic studies highlight the sometimes unexpected evolutionary relationships between seemingly homologous and divergent regulators.

- Recent advances have led to novel experimental approaches that identify RNA regulators lacking obvious homology. Continuing technical developments will allow researchers to better explore the evolutionary origins of such regulators in the absence of significant sequence similarity.

Abbreviations

CITE, cap-independent translation element; IRES, internal ribosomal entry site; ITAF, IRES trans-acting factor; ppGpp, guanosine-3′, 5′-bisdiphosphate; PRPP, phosphoribosyl pyrophosphate; RNAT, RNA thermometer; UTR, untranslated region.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

PORTLAND
PRESS

References
1 Breaker, R.R. (2018) Riboswitches and translation control. *Cold Spring Harb. Perspect. Biol.* **10**, a032797 https://doi.org/10.1101/cshperspect.a032797
2 Bastet, L., Turcotte, P., Wade, J.T. and Lafontaine, D.A. (2018) Maestro of regulation: riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay. *RNA Biol.* **15**, 679–682
3 Sherlock, M. and Breaker, R.R. (2020) Former orphan riboswitches reveal unexplored areas of bacterial metabolism, signaling and gene control processes. *RNA* **26**, 675–693 https://doi.org/10.1261/rna.074997.120
4 Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J. et al. (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6421–6426 https://doi.org/10.1073/pnas.0308014101
5 Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J. et al. (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35**, 4809–4819 https://doi.org/10.1093/nar/gkm487
6 Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H. et al. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* **11**, R31 https://doi.org/10.1186/gb-2010-11-3-r31
7 Nelson, J.W., Atilho, R.M., Sherlock, M.E., Stockbridge, R.B. and Breaker, R.R. (2017) Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol. Cell* **65**, 220–230 https://doi.org/10.1016/j.molcel.2016.11.019
8 Peselis, A. and Serganov, A. (2018) Ykkc riboswitches employ an add-on helix to adjust specificity for polyanionic ligands. *Nat. Chem. Biol.* **14**, 887–894 https://doi.org/10.1038/s41589-018-0114-4
9 Sherlock, M.E., Sudarsan, N., Stav, S. and Breaker, R.R. (2018) Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *eLife* **7**, e33908 https://doi.org/10.7554/eLife.33908
10 Sherlock, M.E., Sudarsan, N. and Breaker, R.R. (2018) Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6052–6057 https://doi.org/10.1073/pnas.1720406115
11 Knappenberger, A.J., Reiss, C.W. and Strobel, S.A. (2018) Structures of two aptamers with differing ligand specificity reveal ruggedness in the functional landscape of RNA. *eLife* **7**, e36381 https://doi.org/10.7554/eLife.36381
12 Sherlock, M.E., Sadeeshkumar, H. and Breaker, R.R. (2019) Variant bacterial riboswitches associated with nucleotide hydrolase genes sense nucleoside diphosphates. *Biochemistry* **58**, 401–410 https://doi.org/10.1021/acs.biochem.8b00617
13 Knappenberger, A.J., Reiss, C.W., Focht, C.M. and Strobel, S.A. (2020) A modular RNA domain that confers differential ligand specificity. *Biochemistry* **59**, 1361–1366 https://doi.org/10.1021/acs.biochem.0c00117
14 Sherlock, M.E. and Breaker, R.R. (2017) Biochemical validation of a third guanidine riboswitch class in bacteria. *Biochemistry* **56**, 359–363 https://doi.org/10.1021/acs.biochem.6b01271
15 Sherlock, M.E., Malkowski, S.N. and Breaker, R.R. (2017) Biochemical validation of a second guanidine riboswitch class in bacteria. *Biochemistry* **56**, 352–358 https://doi.org/10.1021/acs.biochem.6b01270
16 Huang, L., Wang, J., Wilson, T.J. and Lilley, D.M.J. (2017) Structure of the guanidine III riboswitch. *Cell Chem. Biol.* **24**, 1407–1415.e2 https://doi.org/10.1016/j.chembiol.2017.08.021
17 Huang, L., Wang, J. and Lilley, D.M.J. (2017) The structure of the guanidine-II riboswitch. *Cell Chem. Biol.* **24**, 695–702.e2 https://doi.org/10.1016/j.chembiol.2017.05.014
18 Reiss, C.W., Xiong, Y. and Strobel, S.A. (2017) Structural basis for ligand binding to the guanidine-I riboswitch. *Structure* **25**, 195–202 https://doi.org/10.1016/j.str.2016.11.020
19 Reiss, C.W. and Strobel, S.A. (2017) Structural basis for ligand binding to the guanidine-II riboswitch. *RNA* **23**, 1338–1343 https://doi.org/10.1261/rna.061804.117
20 Battaglia, R.A., Price, I.R. and Ke, A. (2017) Structural basis for guanidine sensing by the ykkC family of riboswitches. *RNA* **23**, 578–585 https://doi.org/10.1261/rna.060186.116
21 McCown, P.J., Liang, J.J., Weinberg, Z. and Breaker, R.R. (2014) Structural, functional, and taxonomic diversity of three preQ1 riboswitch classes. *Chem. Biol.* **21**, 880–889 https://doi.org/10.1016/j.chembiol.2014.05.015
22 Eichhorn, C.D., Kang, M. and Feigon, J. (2014) Structure and function of preQ1 riboswitches. *Biochim. Biophys. Acta* **1839**, 939–950 https://doi.org/10.1016/j.bbagrm.2014.04.019
23 Liberman, J.A., Suddala, K.C., Aytenfisu, A., Chan, D., Belashov, I.A., Salim, M. et al. (2015) Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3485–E3494 https://doi.org/10.1073/pnas.1503955112
24 Price, I.R., Grigg, J.C. and Ke, A. (2014) Common themes and differences in SAM recognition among SAM riboswitches. *Biochim. Biophys. Acta* **1839**, 931–938 https://doi.org/10.1016/j.bbagrm.2014.05.013
25 Trausch, J.J., Xu, Z., Edwards, A.L., Reyes, F.E., Ross, P.E., Knight, R. et al. (2014) Structural basis for diversity in the SAM clan of riboswitches. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6624–6629 https://doi.org/10.1073/pnas.1312918111

26 Huang, L. and Lilley, D.M.J. (2018) Structure and ligand binding of the SAM-V riboswitch. *Nucleic Acids Res.* **46**, 6869–6879 https://doi.org/10.1093/nar/gky520

27 Weickhmann, A.K., Keller, H., Wurm, J.P., Strebitzer, E., Juen, M.A., Kremser, J. et al. (2019) The structure of the SAM/SAH-binding riboswitch. *Nucleic Acids Res.* **47**, 2654–2665 https://doi.org/10.1093/nar/gky1283

28 Sun, A., Gasser, C., Li, F., Chen, H., Mair, S., Krasheninina, O. et al. (2019) SAM-VI riboswitch structure and signature for ligand discrimination. *Nat. Commun.* **10**, 1–13 https://doi.org/10.1038/s41467-018-07882-8

29 Babina, A.M., Lea, N.E. and Meyer, M.M. (2017) In vivo behavior of the tandem glycine riboswitch in *Bacillus subtilis*. *MBio* **8**, e01602-17 https://doi.org/10.1128/mBio.01602-17

30 Torgerson, C.D., Hiller, D.A. and Strobel, S.A. (2020) The asymmetry and cooperativity of tandem glycine riboswitch aptamers. *RNA* **26**, 564–580 https://doi.org/10.1261/rna.073577.119

31 Crum, M., Ram-Mohan, N. and Meyer, M.M. (2019) Regulatory context drives conservation of glycine riboswitch aptamers. *PLOS Comput. Biol.* **15**, e1007564 https://doi.org/10.1371/journal.pcbi.1007564

32 Ruff, K.M., Muhammad, A., McCown, P.J., Breaker, R.R. and Strobel, S.A. (2016) Singlet glycine riboswitches bind ligand as well as tandem riboswitches. *RNA* **22**, 1728–1738 https://doi.org/10.1261/rna.057935.116

33 Lemay, J.-F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P. et al. (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet* **7**, e1001278 https://doi.org/10.1371/journal.pgen.1001278

34 Baird, S.D., Lewis, S.M., Turcotte, M. and Holcik, M. (2007) A search for structurally similar cellular internal ribosome entry sites. *Nucleic Acids Res.* **35**, 4664–4677 https://doi.org/10.1093/nar/gkm483

35 Lozano, G., Francisco-Velilla, R. and Martinez-Salas, E. (2018) Deconstructing internal ribosome entry site elements: an update of structural motifs and functional divergences. *Open Biol.* **8**, 180155 https://doi.org/10.1098/rsob.180155

36 Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J. and Embarek, A.M. (2018) Insights into structural and mechanistic features of viral IRES elements. *Front. Microbiol* **8**, 2629 https://doi.org/10.3389/fmicb.2017.02629

37 Peguero-Sanchez, E., Pardo-Lopez, L. and Merino, E. (2015) IRES-dependent translated genes in fungi: computational prediction, phylogenetic conservation and functional association. *BMC Genom.* **16**, 1059 https://doi.org/10.1186/s12864-015-2266-x

38 Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A.A., Stern-Ginossar, N., Yakhini, Z. et al. (2016) Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, aad4939 https://doi.org/10.1126/science.aad4939

39 Koirala, D., Shao, Y., Koldobskaya, Y., Fuller, J.R., Watkins, A.M., Shelke, S.A. et al. (2019) A conserved RNA structural motif for organizing topology within picornaviral internal ribosome entry sites. *Nat. Commun.* **10**, 3629 https://doi.org/10.1038/s41467-019-11585-z

40 Pisareva, V.P., Pisarev, A.V. and Fernández, I.S.) Dual tRNA mimicry in the cricket paralysis virus IRES uncovers an unexpected similarity with the Hepatitis C Virus IRES. *eLife* **7**, e34062 https://doi.org/10.7554/eLife.34062

41 Yamamoto, H., Collier, M., Loerke, J., Ismer, J., Schmidt, A., Hilal, T. et al. (2015) Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. *EMBO J.* **34**, 3042–3058 https://doi.org/10.15252/embj.201592469

42 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 https://doi.org/10.1016/j.virol.2015.02.016

43 Boros, Á., Fenyvesi, H., Pankovics, P., Biró, H., Phan, T.G., Delwart, E. et al. (2015) Secondary structure analysis of swine pasivirus (family Picornaviridae) RNA reveals a type-IV IRES and a parechovirus-like 3′ UTR organization. *Arch. Virol.* **160**, 1363–1366 https://doi.org/10.1007/s00705-015-2371-9

44 Boros, Á., Pankovics, P., Simmonds, P., Pollák, E., Mátics, R., Phan, T.G. et al. (2015) Genome analysis of a novel, highly divergent picornavirus from common kestrel (*Falco tinnunculus*): the first non-enteroviral picornavirus with type-I-like IRES. *Infect. Genet. Evol.* **32**, 425–431 https://doi.org/10.1016/j.meegid.2015.04.005

45 Giangaspero, M., Steinbach, F., Strong, R., Decaro, N., Buonavoglia, C., Domenis, L. et al. (2020) Characterization of internal ribosome entry sites according to secondary structure analysis to classify border disease virus strains. *J. Virol. Methods* **275**, 113704 https://doi.org/10.1016/j.jviromet.2019.113704

46 Furuse, Y., Chaimongkol, N., Okamoto, M. and Oshitani, H. (2019) Evolutionary and functional diversity of the 5′ untranslated region of enterovirus D68: increased activity of the internal ribosome entry site of viral strains during the 2010s. *Viruses* **11**, 626 https://doi.org/10.3390/v11070626

47 Sherpa, C., Jackson, P.E.H., Gray, L.R., Anastos, K., Le Grice, S.F.J., Hammarskjold, M.-L. et al. (2019) Evolution of the HIV-1 rev response element during natural infection reveals nucleotide changes that correlate with altered structure and increased activity over time. *J. Virol.* **93**, e02102-18 https://doi.org/10.1128/JVI.02102-18

48 Loh, E., Righetti, F., Eichner, H., Twittenhoff, C. and Narberhaus, F. (2018) RNA thermometers in bacterial pathogens. *Microbiol. Spectr.* **6**, RWR-0012-2017 https://doi.org/10.1128/microbiolspec.RWR-0012-2017

49 Meyer, A., Freier, M., Schmidt, T., Rostowski, K., Zwoch, J., Lilie, H. et al. (2020) An RNA thermometer activity of the west Nile virus genomic 3′-terminal stem-loop element modulates viral replication efficiency during host switching. *Viruses* **12**, 104 https://doi.org/10.3390/v12010104

50 Miller, W.A., Wang, Z. and Treder, K. (2007) The amazing diversity of cap-independent translation elements in the 3′-untranslated regions of plant viral RNAs. *Biochem. Soc. Trans.* **35**, 1629–1633 https://doi.org/10.1042/BST0351629

51 Kraft, J.J., Peterson, M.S., Cho, S.K., Wang, Z., Hui, A., Rakotondrafara, A.M. et al. (2019) The 3′ untranslated region of a plant viral RNA directs efficient cap-independent translation in plant and mammalian systems. *Pathogens* **8**, 28 https://doi.org/10.3390/pathogens8010028

52 Garcia-Moreno, M., Sanz, M.A. and Carrasco, L. (2016) A viral mRNA motif at the 3′-untranslated region that confers translatability in a cell-specific manner. Implications for virus evolution. *Sci. Rep.* **6**, 1–17 https://doi.org/10.1038/srep19217

53 Meristoudis, C., Trangas, T., Lambrianidou, A., Papadopoulos, V., Dimitriadis, E., Courtis, N. et al. (2015) Systematic analysis of the contribution of c-myc mRNA constituents upon cap and IRES mediated translation. *Biol. Chem.* **396**, 1301–1313 https://doi.org/10.1515/hsz-2015-0190

54 Ren, G.-X., Guo, X.-P. and Sun, Y.-C. (2017) Regulatory 3′ untranslated regions of bacterial mRNAs. *Front. Microbiol.* **8**, 1276 https://doi.org/10.3389/fmicb.2017.01276

55 Menendez-Gil, P., Caballero, C.J., Catalan-Moreno, A., Irurzun, N., Barrio-Hernandez, I., Caldelari, I. et al. (2020) Differential evolution in 3′UTRs leads to specific gene expression in Staphylococcus. *Nucleic Acids Res.* **48**, 2544–2563 https://doi.org/10.1093/nar/gkaa047

PORTLAND
PRESS

1951

56   Dar, D., Shamir, M., Mellin, J.R., Koutero, M., Stern-Ginossar, N., Cossart, P. et al. (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**, aad9822 https://doi.org/10.1126/science.aad9822

57   Dar, D. and Sorek, R. (2018) High-resolution RNA 3′-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res.* **46**, 6797–6805 https://doi.org/10.1093/nar/gky274

58   Mokrejš, M., Mašek, T., Vopálenský, V., Hlubuček, P., Delbos, P. and Pospíšek, M. (2010) IRESite—a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res.* **38**, D131–D136 https://doi.org/10.1093/nar/gkp981

59   Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenie. *Bioinformatics* **30**, 1312–1313 https://doi.org/10.1093/bioinformatics/btu033

60   Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G. and Smith, D.B. (2018) Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717 https://doi.org/10.1093/nar/gkx932