

Functional long-range RNA–RNA interactions in positive-strand RNA viruses

Beth L. Nicholson and K. Andrew White

Abstract | Positive-strand RNA viruses are important human, animal and plant pathogens that are defined by their single-stranded positive-sense RNA genomes. In recent years, it has become increasingly evident that interactions that occur between distantly positioned RNA sequences within these genomes can mediate important viral activities. These long-range intragenomic RNA–RNA interactions involve direct nucleotide base pairing and can span distances of thousands of nucleotides. In this Review, we discuss recent insights into the structure and function of these intriguing genomic features and highlight their diverse roles in the gene expression and genome replication of positive-strand RNA viruses.

RNA-dependent RNA polymerase

(RdRp). A polymerase that uses an RNA template to synthesize a complementary RNA strand.

Subgenomic mRNAs

(sgmRNAs). Virally derived mRNAs that are transcribed by certain RNA viruses during infections and that enable the expression of a subset of viral proteins.

Cis-acting RNA elements

RNA sequences or structures that regulate the activity of the RNA in which they reside.

Trans-acting factors

Molecules, such as proteins or RNAs, that can elicit an effect in an interacting partner molecule.

Positive-strand RNA viruses constitute more than one-third of all virus genera and cause various diseases in plants and animals. Notable members of this group include hepaciviruses, coronaviruses, flaviviruses and tombusviruses. These viruses have single-stranded RNA genomes that have the same coding sense as host mRNAs and that thus function directly as templates for the translation of viral proteins (FIG. 1). One of the proteins that is expressed during infection is RNA-dependent RNA polymerase (RdRp), which is responsible for the replication of viral genomes and, in certain viruses, the transcription of smaller viral subgenomic mRNAs (sgmRNAs)¹ (FIG. 1). The regulation of fundamental viral processes generally involves local RNA elements within genomes, which recruit viral and host factors^{2,3}. Such regulatory units are known as *cis*-acting RNA elements, and the factors that they interact with, such as viral or cellular proteins, are generically known as *trans*-acting factors.

Although most of the regulatory *cis*-acting RNA elements in positive-strand RNA viruses are local sequences or confined secondary or tertiary structures, there is mounting evidence that long-range intragenomic RNA base pairing can also have essential functions, such as executing and modulating different viral processes^{4–7}. The definition of a long-range RNA–RNA interaction can vary and, depending on the scientific perspective, can range from tens to thousands of nucleotides. There are many examples of interactions in the lower range that regulate viral processes, and many of these have been described in other reviews^{3,4}. Conversely, interactions in the upper range tend to be

less prevalent, but can also have important functional consequences for viral gene expression and/or genome replication in plant and animal positive-strand RNA viruses (TABLE 1). Such long-range interactions were first reported for positive-strand RNA bacteriophages in the 1990s^{8–11} and, since then, have been discovered in various positive-strand RNA viruses that infect eukaryotes^{4–7}. All of the long-range interactions that have been characterized so far involve Watson–Crick RNA base pairing, and they often include sequences that are positioned in loop regions or in internal bulges within RNA structures. In this Review, we focus on the burgeoning field of interactions that span large distances, that is, $\geq 1,000$ nucleotides, and we discuss how these interactions modulate viral processes. In particular, we focus on: the regulation of translation initiation by 3' cap-independent translational enhancers (3' CITEs) and internal ribosome entry sites (IRESs); translational recoding events, including ribosomal frameshifting and stop codon readthrough; genome replication; and sgmRNA transcription. We then consider the possible roles and regulatory mechanisms of such interactions and discuss how they function in the complex context of viral RNA genomes. Last, we discuss outstanding questions that will determine the future directions of research into long-range intragenomic RNA–RNA interactions.

Translation of viral proteins

Positive-strand RNA viral genomes are directly translated into viral proteins by host ribosomes, and a

Department of Biology,
York University, Toronto,
Ontario M3J 1P3, Canada.
Correspondence to K.A.W.
e-mail: kawhite@uorku.ca
doi:10.1038/nrmicro3288

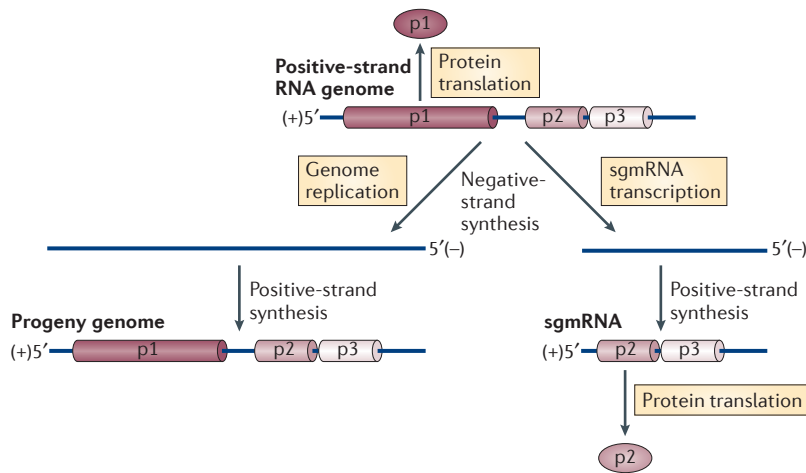


Figure 1 | Fundamental steps in positive-strand RNA virus replication. A linear representation of a generic positive-strand RNA viral genome is shown with the encoded proteins represented by cylinders. The viral protein that is encoded at the 5' end, p1 (for example, the viral RNA-dependent RNA polymerase (RdRp)), is translated directly from the genome. RdRp synthesizes a full-length complementary negative-strand RNA using the genome as a template, which is subsequently used for the synthesis of progeny genomes. Certain viral genomes also function as templates for the transcription of subgenomic mRNAs (sgmRNAs), which is a process that is also mediated by the viral RdRp and that involves a negative-strand intermediate. The ORFs of additional viral genes that are located downstream of the first ORF are generally not efficiently translated, owing to poor ribosome access (not shown). The transcription of smaller viral sgmRNAs enables the expression of these additional proteins (for example, p2), as downstream ORFs in sgmRNAs are relocated to 5'-proximal positions, which enables efficient ribosome access.

3' cap-independent translational enhancers (3' CITEs). Functional folded regions of RNA in or near to the 3' UTRs of certain positive-strand RNA virus genomes; they bind to translation initiation factors or ribosomal subunits to facilitate translation.

Internal ribosome entry sites (IRESs). Functional folded regions of RNA in or near to the 5' UTRs of certain positive-strand RNA virus genomes (and some cellular mRNAs) that recruit ribosomes directly to an internal location close to the initiation codon independently of a 5' cap structure.

Ribosomal frameshifting A translational recoding event in which a proportion of translating ribosomes change their reading frame at a defined point in the mRNA, resulting in the addition of an alternative carboxy-terminal extension.

minority are generated by translational recoding mechanisms that involve either ribosomal frameshifting or stop codon readthrough¹². Regardless of the translational strategy that is used, all viral genomes compete against abundant cellular mRNAs for ribosomes. Accordingly, these RNA genomes contain structural features that assist in the recruitment of the host translational machinery. Some viruses use the conventional terminal structures of mRNAs, that is, the 5' cap and the 3' poly(A) tail¹³. In other cases, less typical elements are used, such as 3' CITEs^{5,6} or IRESs¹⁴, which are located in genomic 3' and 5' UTRs, respectively. Some of the viruses that use 3' CITEs or IRESs, or that rely on translational recoding mechanisms, also require associated long-range RNA–RNA interactions.

3' CITE-mediated translation. 3' CITEs have been identified in several genera of the virus family *Tombusviridae*, as well as in the related genera *Luteovirus* (family *Luteoviridae*) and *Umbravirus* (family unclassified)^{5,6}. These positive-strand RNA genomes lack both 5' caps and 3' poly(A) tails and instead rely on 3' CITEs for protein translation. Several different structural classes of 3' CITE have been uncovered, including a 68 nucleotide RNA stem–loop¹⁵ and more complex multihelix pseudoknot^{16,17} or tRNA-shaped structures^{18,19}. Most 3' CITEs function by binding to the ribosome-recruiting eukaryotic translation initiation factor 4F (eIF4F) complex via one or both of its eIF4G and eIF4E subunits^{15,16,20–22}, whereas others bind directly to ribosomal subunits^{23,24}. Thus, 3' CITEs can recruit ribosomes by either direct or

indirect means. Despite their structural and functional diversity, 3' CITEs share a common relative position near to, or within, the 3' UTRs of viral genomes. Paradoxically, this 3'-proximal position places them, and the recruited translational machinery, at the opposite end to the site of translation initiation. Consequently, many of these viruses use long-range RNA-based interactions to relocate 3' CITEs that are bound to the translational machinery to positions that are close to their cognate 5' UTRs.

In the genome of barley yellow dwarf virus (BYDV; genus *Luteovirus*)²⁵, complementary sequences that are located in the 5' UTR and 3' CITE form a long-distance RNA–RNA bridge via a base-pair-mediated kissing-loop interaction²⁵. The importance of this interaction for viral translation was confirmed by compensatory mutational analysis, whereby disruption, and then restoration, of base-pairing potential by substitutions correlated with low and high levels of translational activity, respectively²⁵. Similar functional base-pairing interactions between 5' UTRs and 3' CITEs have also been shown for different members of the genus *Tombusvirus*, including maize necrotic streak virus (MNeSV), carnation Italian ringspot virus (CIRV) and tomato bushy stunt virus (TBSV)^{15,22,26–28}, as well as for saguaro cactus virus of the genus *Carmovirus*²⁹. These findings have led to a general model for 3' CITE activity, in which simultaneous binding of the 3' CITE to both the eIF4F complex and the 5' UTR enables the eIF4F-mediated recruitment of the 40S ribosomal subunit to the 5' end^{15,16,21,25,27,30} (FIG. 2a). *In vitro* studies with the MNeSV 3' CITE have confirmed the formation of the proposed tripartite 5' UTR–3' CITE–eIF4F complex and its requirement for efficient ribosome recruitment to the 5'-proximal start codon¹⁵.

3' CITE-dependent translation is more complicated in pea enation mosaic virus (PEMV; *Umbravirus*)³¹. PEMV contains two different types of 3' CITE: a panicum mosaic virus-like translational enhancer (PTE), which is located in its 3' UTR and binds to eIF4F^{16,17}, and a kissing-loop T-shaped structure (kl-TSS), which is positioned immediately upstream of the PTE and binds directly to 60S ribosome subunits²⁴ (FIG. 2b). Accordingly, both direct and indirect modes of ribosome recruitment probably occur in PEMV. However, unlike the PTEs that have been identified in other viruses²⁹, the PTE in PEMV does not interact with the 5' end of the viral genome¹⁶. Instead, the adjacent kl-TSS engages in a long-distance interaction with a 5'-proximal hairpin, thereby uniting the terminal regions³¹ (FIG. 2b). The specific contribution of each of the 3' CITEs to the PEMV translational process remains to be fully determined; however, it is clear that the kl-TSS-mediated long-range interaction could be beneficial to the activity of both of the 3' CITEs by repositioning them close to the site of translation initiation.

IRES-mediated translation. IRESs are structured RNA elements that recruit ribosomes — either directly or with the assistance of cellular proteins — to the vicinity of a start codon¹⁴. Certain viral IRESs are modulated by

Table 1 | **Functional intragenomic interactions in positive-strand RNA viruses**

Virus (genus)	RNA–RNA interaction	Viral process regulated	Refs
Many plant viruses in the <i>Tombusviridae</i> family and the Umbravirus and Luteovirus genera	3' CITE–5' UTR or 3' CITE–5' coding region	Translation initiation ^{5,6}	5,6
BYDV (Luteovirus)	3' CITE–5' UTR	Translation initiation	21,25,30
	Frameshift site–3' UTR	Ribosomal frameshifting	47
CIRV (Tombusvirus)	3' CITE–5' UTR	Translation initiation	22,28
	PRTE–DRTE	Stop codon readthrough	50
TBSV (Tombusvirus)	3' CITE–5' UTR	Translation initiation	26,27
	UL–DL	Genome replication	52
	AS1–RS1; AS2–RS2; DE–CE	sgmRNA transcription	81–84
FMDV (Aphthovirus)	IRES–3' UTR	Translation initiation	32,33
	S-region–3' UTR	Possibly genome replication	33
CSFV (Pestivirus)	IRES–3' terminus	Translation	34
HCV (Hepacivirus)	IRES–5BSL3.2	Translation initiation	35–37
	5BSL3.2–3' UTR	Genome replication	38–42
DENV and WNV (Flavivirus)	5' UAR–3' UAR; 5' DAR–3' DAR; 5' CS–3' CS	Genome replication	53–61,65,66
TGEV (Coronavirus)	DE–PE; cBM–BM	sgmRNA–N transcription	77–79

3' CITE, 3' cap-independent translational enhancer; 5' CS–3' CS, 5' cyclization sequence–3' cyclization sequence; AS–RS, activator sequence 1–receptor sequence 1; cBM–BM, complementary B motif–B motif; BYDV, barley yellow dwarf virus; CIRV, carnation Italian ringspot virus; CSFV, classical swine fever virus; DAR, downstream of AUG region; DE–CE, distal element–core element; DENV, dengue virus; DRTE, distal readthrough element; FMDV, foot-and-mouth disease virus; HCV, hepatitis C virus; IRES, internal ribosome entry site; DE–PE, DE–proximal element; PRTE, proximal readthrough element; sgmRNA, subgenomic RNA; sgmRNA–N, sgmRNA–nucleocapsid protein; TBSV, tomato bushy stunt virus; TGEV, transmissible gastroenteritis virus; UL–DL, upstream linker–downstream linker; UAR, upstream of AUG region; WNV, West Nile virus.

Stop codon readthrough

A translational recoding event in which a proportion of ribosomes read a stop codon as a sense codon, which results in a carboxy-terminally extended protein.

5' cap

A 7-methyl guanine nucleotide that is linked via a 5'-to-5' triphosphate to the 5'-terminal nucleotide of eukaryotic mRNAs and some viral RNA genomes. It is bound by the cap-binding protein, eukaryotic translation initiation factor 4E (eIF4E), which is part of the eIF4F complex that recruits the small ribosomal subunit.

3' poly(A) tail

A chain of consecutive adenine nucleotides that is present at the 3' end of eukaryotic mRNAs and of some viral RNAs. The poly(A) tail is bound by poly(A)-binding protein, which mediates increased mRNA stability and enhanced protein translation.

UTRs

Regions within cellular mRNAs or viral RNA genomes that do not encode proteins. These regions, which are most commonly located at 5' or 3' ends, often contain *cis*-acting RNA elements that regulate RNA activities.

Pseudoknot

A higher-order RNA structure that involves base pairing between the loop of a stem-loop structure and an upstream or downstream RNA sequence that is located outside of the stem-loop.

Eukaryotic translation initiation factor 4F (eIF4F)

A translation initiation factor complex that consists of three subunits: the cap-binding protein eIF4E; the RNA helicase eIF4A; and eIF4G, which recruits the small ribosomal subunit.

regions that are far downstream of translation initiation sites. For example, the 5'-uncapped but 3'-polyadenylated genome of the picornavirus foot-and-mouth disease virus (FMDV) contains a 5' IRES that is positively regulated by interactions with the genomic 3' UTR³². *In vitro*, this 3' UTR participates in two long-range interactions with the 5' UTR — one with the IRES and the other with a 5'-terminal S-region that is involved in genome replication³³ (FIG. 2c). Although the specific sequences involved were not identified, the interaction of the 3' UTR with the IRES was found to be independent of its interaction with the S-region, and the two interactions could not form simultaneously. Therefore, the detected contacts could potentially modulate both translation and genome replication.

Negative regulators of viral IRES activity have also been identified. The genome of the pestivirus classical swine fever virus (CSFV) lacks both a 5' cap and a 3' poly(A) tail, and the 3' UTR inhibits the translational activity of the IRES in the 5' UTR³⁴. The negative regulatory sequence that affected IRES activity was mapped to a 3'-terminal RNA hairpin that ends with CGGCC-OH. This terminal sequence was also found to be complementary to a sequence that is located in the ribosome-binding region of the IRES (FIG. 2d), which suggests that a CGGCC–IRES base-pairing interaction inhibits ribosome recruitment³⁴.

Regulation in hepatitis C virus (HCV) is more complex and involves a network of RNA–RNA interactions. The HCV genome does not contain a 5' cap or a 3' poly(A) tail but instead has an IRES that binds directly

to the 40S ribosomal subunit. IRES activity is downregulated by a long-range RNA–RNA interaction^{35–37} that occurs between the apical loop of helix IIIId in the IRES and a bulge in an essential 3'-proximal *cis*-acting replication element in the coding region of the non-structural protein 5B (NS5B), which is known as 5BSL3.2 (REF. 35) (FIG. 2e). Interestingly, the same 5BSL3.2 bulge sequence also mediates genome replication by interacting with a nearby upstream sequence located around nucleotide 9110 (REFS 38–41) (FIG. 2e). Since the two interactions are equally probable in a thermodynamic context⁴¹, shifting of the conformational equilibrium between the two interactions could regulate viral translation and genome replication^{39,41,42}. In addition, there is both genetic and structural evidence that the terminal loop of 5BSL3.2 base pairs with the 3' SL2 element, which is an RNA element that is located in the 3' UTR and is involved in genome replication^{39–41,43} (FIG. 2e). Accordingly, 5BSL3.2 functions as a central hub for a network of interactions that collectively modulate IRES-mediated translation and genome replication.

Intriguingly, some uncapped and non-polyadenylated virus genomes use both a 5' IRES and a 3' CITE; for example, the plant nepovirus blackcurrant reversion virus (order Picornavirales) uses a hybrid 5' IRES–3' CITE-mediated translation mechanism and also requires a long-distance interaction between the 5' and 3' genomic ends for optimal translation^{44–46}. Although the details of the individual or combined functions of the 5' IRES and 3' CITE remain to be determined, it was suggested

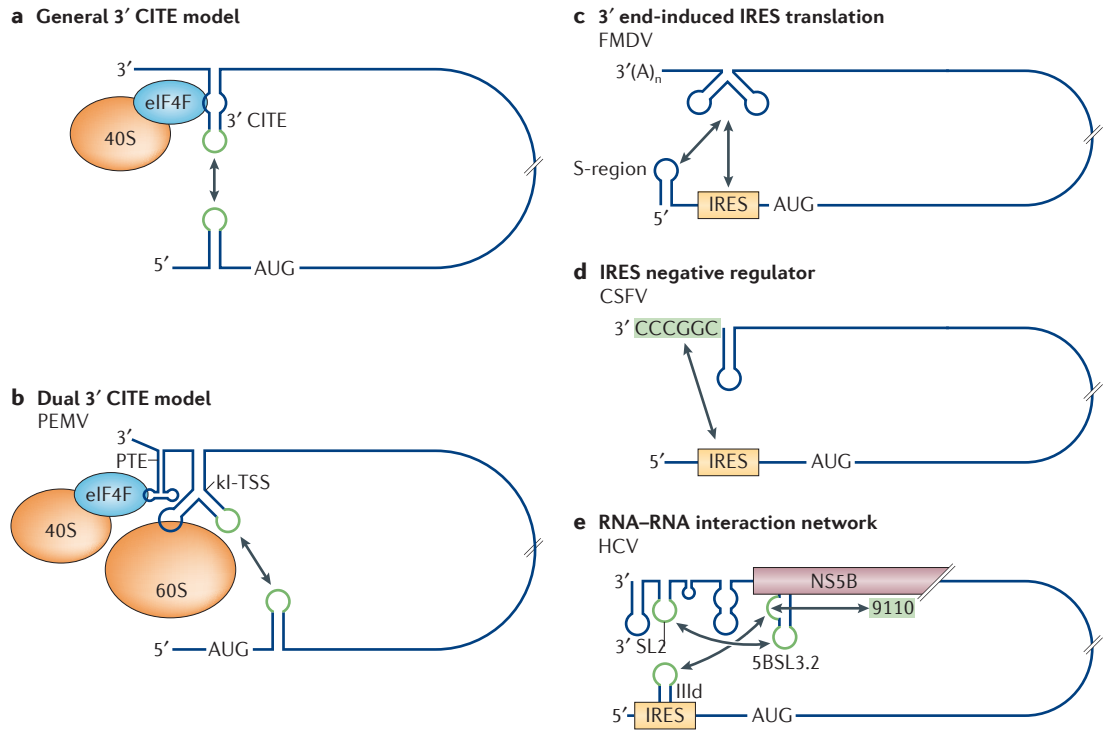


Figure 2 | Translation initiation regulated by long-distance interactions. **a,b** | 3' cap-independent translational enhancer (3' CITE)-mediated translation. **a** | General 3' CITE model showing the long-range interaction (indicated by the double-headed arrow) that positions the eukaryotic translation initiation factor 4F (eIF4F)-bound 3' CITE close to the 5' end of the genome, where it enables eIF4F-mediated recruitment of the 40S ribosomal subunit to the 5' end to initiate translation. Interacting sequences are shown in green. **b** | Dual 3' CITE model. Pea enation mosaic virus (PEMV) contains two different types of 3' CITE: a panicum mosaic virus-like translational enhancer (PTE), which is located in the 3'UTR and binds to eIF4F; and a kissing-loop T-shaped structure (kl-TSS), which is positioned immediately upstream and binds directly to the 60S ribosome subunit and mediates long-distance RNA–RNA base pairing with a 5'-proximal hairpin. **c–e** | Internal ribosome entry site (IRES)-mediated translation. **c** | The foot-and-mouth disease virus (FMDV) IRES is stimulated by the 3' UTR, which engages in long-range contacts with two regions in the 5' UTR: the IRES and a region that has been shown to be involved in genome replication, which is known as the S-region. The specific sequences that are involved in this interaction have not been identified and the interaction with the S-region may modulate genome replication. **d** | The 3'-terminal hexamer CCGCCC in classical swine fever virus (CSFV) is a negative modulator of IRES-mediated translation and may confer its inhibition by pairing with a ribosome-binding region in the IRES, thus blocking ribosome binding. **e** | Hepatitis C virus (HCV) IRES activity is negatively regulated by an interaction between helix IIIId of the IRES and a bulge in the structure 5BSL3.2, which is located in the coding region of non-structural protein 5B (NS5B). The same bulge in 5BSL3.2 also interacts with a genomic sequence around position 9110 of the genome, and the terminal loop of 5BSL3.2 can pair with the 3' SL2 element located in the 3' UTR, which may modulate genome replication. These interactions may coordinate viral translation and genome replication.

that this terminal interaction might help to facilitate the re-recruitment of terminated ribosomes⁴⁶. Indeed, this potential function in ribosome recycling is also applicable to some of the above examples in which 5'–3' interactions increase translational efficiency.

Translational recoding. Recoding via stop codon readthrough or ribosomal frameshifting leads to the production of carboxy-terminally extended proteins. In certain viruses, functional long-range base-pairing interactions were found to be required for both of these types of recoding events^{47–50}, and the involvement of such interactions is particularly prevalent in positive-strand RNA plant viruses¹².

The most common form of ribosomal frameshifting involves a small proportion of elongating ribosomes

moving backwards one base and then resuming translation in the new –1 reading frame¹². This process is facilitated by a 'slippery' heptanucleotide sequence at the frameshift site and a stimulatory RNA structure that is located a few nucleotides downstream¹² (FIG. 3a). In addition, in the plant viruses BYDV and red clover necrotic mosaic virus (RCNMV; genus Dianthovirus), the efficient –1 frameshifting that produces their viral RdRps requires base pairing between their proximal stimulatory RNA structures and complementary sequences that are located ~4,000 and ~2,500 nucleotides downstream, respectively^{47,48}. In BYDV, a bulge in the stimulatory RNA structure next to the frameshift site interacts with the terminal stem-loop of an RNA hairpin near to the 3' end of the genome (FIG. 3a), and a similar interaction occurs in RCNMV⁴⁸. In addition to mediating

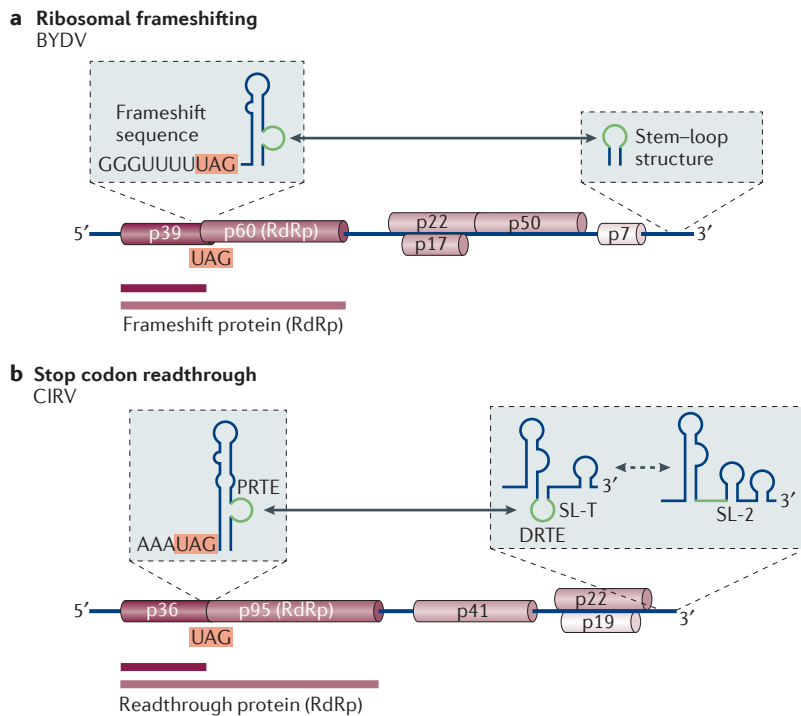


Figure 3 | Translational recoding facilitated by long-distance interactions.

a | Linear representation of the barley yellow dwarf virus (BYDV) RNA genome, which shows coding regions as cylinders. p39 and p60 correspond to proteins of 39 kDa and 60 kDa, respectively, which are encoded in two separate ORFs in different reading frames. p39 is produced when no ribosome frameshifting occurs. However, when a -1 frameshift occurs near the very 3' end of the p39 ORF, ribosomes are shifted into the p60 reading frame and translate the p60 ORF as a carboxy-terminal extension of p39. The resulting frameshift protein is the viral RNA-dependent RNA polymerase (RdRp), which is approximately 99 kDa. Frameshifting is stimulated by a long-range interaction (double-headed arrow) between a bulge in an RNA structure that is near to the frameshift site and the terminal loop of a 3'-proximal stem-loop. Interacting sequences are shown in green. **b** | Linear representation of the carnation Italian ringspot virus (CIRV) RNA genome, which shows readthrough translation of its RdRp. Ribosomes that initiate at the 5' end of the genome normally terminate at the stop codon at the end of the p36 ORF, producing a protein of 36 kDa. Translational readthrough of the p36 stop codon produces a C-terminally extended readthrough protein of 95 kDa (p95), which is the viral RdRp. Readthrough requires base pairing between the proximal readthrough element (PRTE) that is located near to the stop codon and the 3'-proximal distal readthrough element (DRTE). The DRTE 3' sequence is associated with one of two mutually exclusive RNA conformations. The SL-T conformation facilitates readthrough and prevents genome replication, whereas the alternative conformation, which contains SL-2, promotes replication and inhibits readthrough. These two conformations thus represent a type of RNA switch that probably coordinates translation and replication.

frameshifting, the interaction is also proposed to assist in the coordination of translation and negative-strand synthesis, which are directionally opposed processes for an RNA genome^{47,48}.

The interactions between the stimulatory RNA structures and the 3'-proximal RNA hairpins in BYDV and RCNMV are presumed to occur intramolecularly; however, there has been a recent report of a -1 ribosomal frameshifting event that is enhanced by an intermolecular genomic interaction⁴⁹. In the severe acute respiratory syndrome coronavirus (SARS-CoV), such an interaction involves a palindromic loop sequence in a local pseudoknot that is positioned just downstream of the

frameshift site. The palindromic loop sequences in two SARS-CoV genomes form a kissing-loop structure that increases frameshifting efficiency *in vitro*. Mutations that disrupted the base pairing abolished dimerization, reduced frameshifting and inhibited the accumulation of viral RNA in infected cells⁴⁹. The disruption also affected the ratio of genomic RNA to sgRNA levels and growth kinetics, which suggests that this intramolecular interaction has a genuine regulatory role in the viral life cycle⁴⁹.

Another common viral recoding strategy is stop codon readthrough¹², whereby, instead of ribosome termination, the stop codon is decoded as a sense codon. Translation then proceeds in the original reading frame, which results in an extended protein that is produced at a low frequency. As with frameshifting, the efficiency of codon readthrough is typically influenced by RNA sequences and structures that immediately surround the stop codon¹². In the plant tobravirus CIRV, stop codon readthrough generates the viral RdRp, and this process requires a long-distance interaction between an RNA structure that is immediately downstream of the readthrough site (which is known as the proximal readthrough element (PRTE)) and a sequence in the 3' UTR (which is known as the 3'-proximal distal readthrough element (DRTE))⁵⁰ (FIG. 3b). The DRTE is associated with one of two mutually exclusive stem-loop structures, SL-T and SL-2, of which SL-2 is essential for genome replication⁵⁰. Formation of SL-T positions the complementary 3' sequence in its terminal loop, which facilitates the establishment of the long-distance interaction that improves translational readthrough and simultaneously inhibits genome replication by precluding the formation of SL-2. Conversely, the SL-2-containing conformation promotes genome replication and impedes readthrough (FIG. 3b). On the basis of these observations, SL-T and SL-2 were proposed to function as an RNA switch that assists in the coordination of translation and replication⁵⁰.

Viral genome replication

The replication of positive-strand RNA virus genomes occurs via the synthesis of a complementary negative-strand RNA, which is subsequently used as a template for the production of progeny positive-strand RNA genomes (FIG. 1). This process is catalysed by a virally encoded RdRp and is assisted by viral and host proteins. The initiation of negative-strand synthesis involves the RdRp accessing the 3' terminus of a genome, and RNA sequences and structures that facilitate this are usually located near to the 3' end. However, there is compelling evidence that RNA elements that are considerably distal to 3' ends can also influence the efficiency of complementary strand production^{51,52}.

Flavivirus genome replication. Several members of the genus *Flavivirus*, including dengue virus (DENV), West Nile virus (WNV) and yellow fever virus (YFV), require genome circularization, which is mediated by base-pairing interactions between sequences in their genomic termini, for replication^{51,53–61}. For DENV, three different complementary sequences are involved in these interactions; these sequences are known as the cyclization

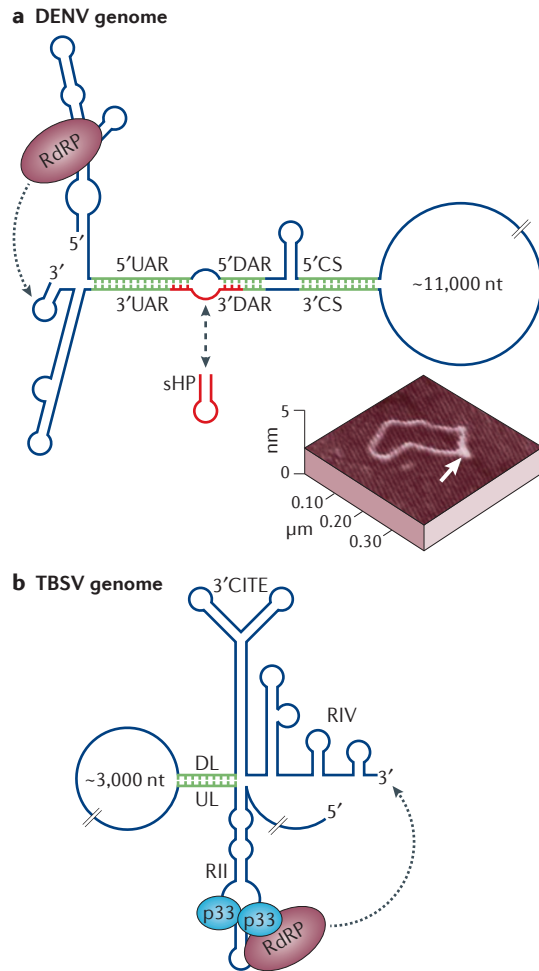


Figure 4 | Viral genome replication directed by long-distance interactions. **a** | A simplified representation of relevant RNA secondary structures in the dengue virus (DENV) RNA genome. Genome replication requires three sets of long-distance interactions: 5'-3' upstream of AUG region (UAR), 5'-3' downstream of AUG region (DAR) and 5' cyclization sequence-3' cyclization sequence (CS). Formation of these interactions results in genome circularization, which repositions the 5'-bound RNA-dependent RNA polymerase (RdRp) to the 3' end of the genome, where it initiates negative-strand synthesis. Interacting sequences are shown in green. The formation of the alternate regulatory structure small hairpin (sHP; shown in red) inhibits circularization. A circularized version of the genome, as observed by atomic force microscopy (AFM), is shown, with the 5'-3' base-paired region denoted by a white arrow⁶⁵. **b** | A simplified representation of the RNA secondary structures in the tomato bushy stunt virus (TBSV) RNA genome that are involved in genome replication. The 3' cap-independent translational enhancer (3' CITE) region is also included, as the downstream linker is located immediately upstream of it. The RdRp interacts with its auxiliary replication protein p33, which binds as a dimer to an internal RNA element, RII. A long-distance interaction that involves the upstream linker and downstream linker (UL-DL) sequences brings RdRp, which is associated with RII close to the 3' end of the genome, where it can initiate negative-strand synthesis. The UL-DL interaction also mediates the formation of a RII-RIV RNA platform that facilitates the assembly of the viral replicase complex. The AFM image in part **a** is reproduced, with permission, from Filomatori, C. V. *et al.* A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev.* 20, 2238-2249 (2006) © Cold Spring Harbor Laboratory Press.

sequence, the upstream of AUG region (UAR) and the downstream of AUG region (DAR) (FIG. 4a). The resulting RNA circularization has been observed directly by atomic force microscopy (AFM) in the absence of proteins, which shows that circularization can be entirely RNA-based⁶⁶ (FIG. 4a). In addition, structural analysis by chemical probing also supports protein-independent interactions between the 5' and 3' termini of the genome⁶². However, although circularization can occur autonomously, protein factors might assist in the process, as both the flavivirus core protein and NS3 helicase have been shown to mediate 5'- and 3'-end base pairing of genomic RNA *in vitro*^{63,64}. The observed circularization is required for flavivirus genome replication because RdRp binds to an RNA stem-loop in the 5'UTR, which positions RdRp ~11 kb upstream of the 3' terminus where negative-strand synthesis initiates^{65,66}. The long-distance RNA-RNA interaction between the genomic termini thus repositions the RdRp to the 3' end, where it can commence initiation⁶⁵ (FIG. 4a).

DENV genome circularization can be modulated by localized regulatory elements, such as a conserved RNA pseudoknot that is located adjacent to the cyclization sequence in the 5' UTR⁶⁷. Moreover, it was recently shown that the DENV genome requires a specific balance between circularized and linear (that is,

non-circularized) conformations⁶⁸. Parts of the 3' UAR and 3' DAR sequences can fold into a small local RNA hairpin in the 3' UTR, which is known as sHP, and the formation of sHP inhibits the interaction with 5'-proximal partner sequences⁶⁸ (FIG. 4a). Virus replication was found to be sensitive to mutations that altered the natural balance between local sHP formation and long-range pairing, which indicates that a defined ratio between circularized and linear conformations is necessary for viability⁶⁸. The regulatory function of sHP might be even more complex, as only base pairing of sHP is important for replication in mammalian cells, whereas, unexpectedly, both base pairing and sequence identity are important in mosquito cells⁶⁹.

Although the requirement for genome circularization in flavivirus replication is now generally accepted, it should be noted that an alternative model has recently been proposed, whereby the 5' and 3' complementary sequences would function in *trans* and generate dimers and/or oligomers of flavivirus genomes⁷⁰. The formation of such concatamers would presumably be concentration-dependent and could have regulatory effects that differ from those of circularized monomers⁷⁰. Although such alternative pairing scenarios are theoretically feasible, their existence and possible biological relevance remain to be investigated.

Atomic force microscopy (AFM). A high-resolution type of scanning probe microscopy that can be used to assess the surface topography of biological molecules.

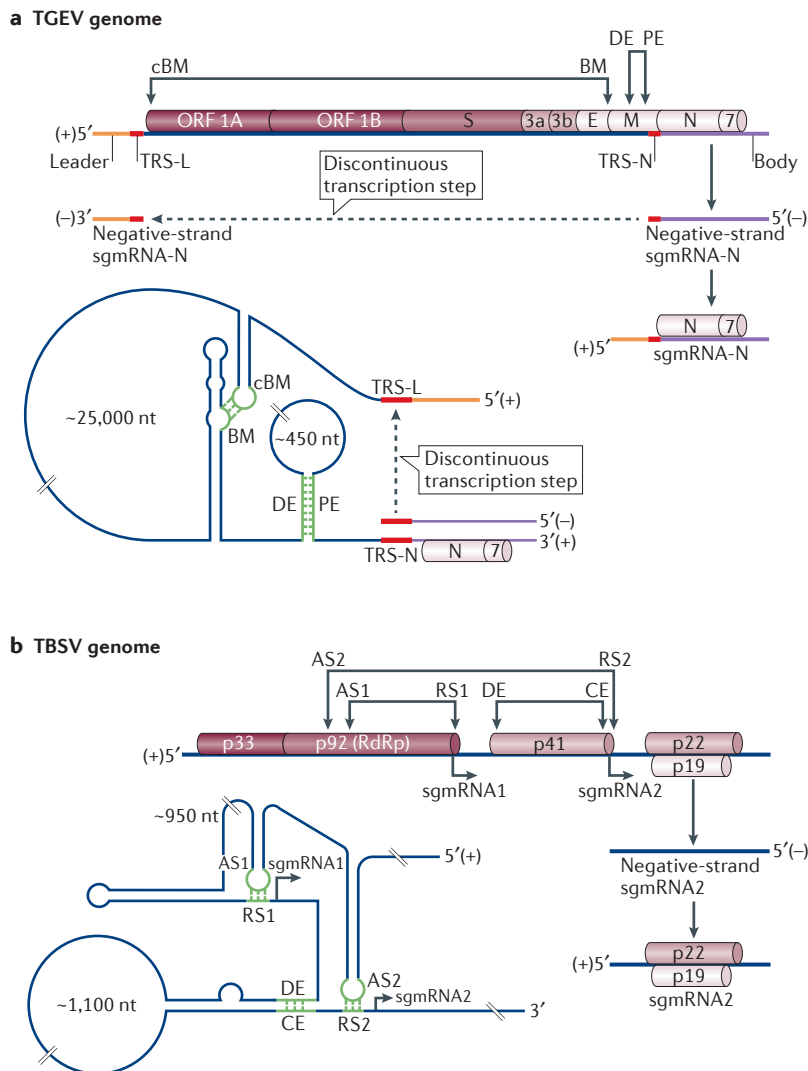


Figure 5 | sgmRNA transcription mediated by long-range interactions. a | The upper panel shows a linear representation of the transmissible gastroenteritis virus (TGEV) RNA genome with different encoded viral proteins, 1A, 1B, S, 3a, 3b, E, M, N and 7. The relative positions of long-range interactions that are involved in the transcription of the subgenomic mRNA (sgmRNA) encoding the viral nucleocapsid protein (sgmRNA-N) are indicated (double-headed arrows; complementary B-motif–B-motif (cBM–BM) and distal element–proximal element (DE–PE)). Transcription of sgmRNA-N involves discontinuous synthesis (dashed arrow) of a negative-strand RNA that contains sequences from the body (purple) region, which is located at the 3' end of the genome, and leader (orange) region, which is located at the 5' end of the genome. The RNA-dependent RNA polymerase (RdRp) repriming step is guided by the transcription-regulating sequences (TRS; red), and the negative strand that is generated is used as a template for sgmRNA-N transcription. In the lower panel, a simplified RNA secondary structure shows the two interactions, cBM–BM and DE–PE (green), bringing the TRS-N and TRS-leader (TRS-L) into close proximity to mediate RdRp repriming. **b |** The upper panel is a linear representation of the tomato bushy stunt virus (TBSV) genome, which shows the relative positions of long-range interactions that are involved in sgmRNA transcription. The transcription pathway that leads to sgmRNA2 production is shown on the right-hand side. The negative-strand template is generated by premature termination of RdRp, which is caused by an attenuation signal in the genome that is formed by a long-range interaction between an activator sequence and a receptor sequence (AS2–RS2). The lower left-hand panel shows a simplified RNA secondary structure that depicts the interactions that are involved in transcription of both sgmRNA1 and sgmRNA2. The long-range interactions AS1–RS1 and AS2–RS2 constitute the attenuation signals for sgmRNA1 and sgmRNA2, respectively. SgmRNA2 transcription requires an additional interaction between a distal element and a core element (DE–CE).

Tombusvirus genome replication. Similarly to flaviviruses, RdRp of the tombusvirus TBSV associates with the viral genome far upstream of the 3' terminus. In this case, RdRp forms a complex with its auxiliary replication protein, p33, which binds specifically to an internal RNA element, known as RII, that is located more than 3 kb upstream of the 3' end^{71,72} (FIG. 4b). The 3' terminus of the genome contains an RNA element known as RIV, which is essential for genome replication. RII and RIV are united by a long-range base-pairing interaction that occurs between an upstream linker sequence, which is 3'-proximal to RII, and a partner downstream linker sequence, which is near to the 3' terminus⁵² (FIG. 4b). In addition to facilitating 3' end access to the RdRp, the association between these linker sequences generates a bipartite RII–RIV RNA platform that is necessary for the assembly of the replicase complex, which is composed of viral and host proteins^{52,73}. As RII-like internal replication elements are also present in other viruses of the *Tombusviridae* family, it is probable that other members of this family might also require similar long-range intragenomic interactions for genome replication⁷⁴.

Viral sgmRNA transcription

Many positive-strand RNA viruses are polycistronic, which means that they encode multiple viral proteins within a single genome segment. ORFs that are located downstream are usually not efficiently translated, owing to poor ribosome access. Thus, to enable the robust expression of these proteins, these viruses transcribe smaller viral sgmRNAs in which the downstream ORFs are relocated to 5'-proximal positions, which thereby enables efficient ribosome access (FIG. 1). The mechanism that is involved in the generation of sgmRNAs depends on the virus, and some viruses that use discontinuous template synthesis or premature termination mechanisms require long-range RNA–RNA interactions.

Coronavirus sgmRNA transcription. Coronavirus genomes are extraordinarily large — they can be up to ~32 kb in length — and expression of their 3'-proximal genes depends on sgmRNA transcription⁷⁵. These sgmRNAs consist of a common 5'-terminal region (known as the leader) that is fused to a variable 3'-terminal segment (known as the body). They are transcribed from complementary negative-strand templates, which are generated by a discontinuous template synthesis mechanism⁷⁶ (FIG. 5a). During negative-strand synthesis, RdRp dissociates from the positive-strand genome at specified locations for each sgmRNA (which defines the body segment) and then repriming on the template within the 5' UTR (where it copies the common leader sequence). The positions of RdRp release and repriming are guided by transcription-regulating sequences (TRSs)⁷⁶. The discontinuous RNA template is then used to transcribe corresponding sgmRNAs, which contain a common 5' leader that is connected to different 3'-terminal body sections that encode the different viral ORFs. In transmissible gastroenteritis virus (TGEV), two long-distance interactions control the transcription of the sgmRNA that encodes the nucleocapsid protein

(sgmRNA-N)^{77–79} (FIG. 5a). One of these interactions spans almost 26 kb, which is the longest that has been reported so far; it forms between complementary sequences known as the B-motif (BM) and the complementary BM (cBM), which are located upstream of the TRS for sgmRNA-N (TRS-N) and downstream of the leader TRS (TRS-L), respectively⁷⁹. This interaction, in combination with a shorter-range interaction between a distal element and a proximal element (DE–PE), brings the two TRS elements into close proximity, which promotes efficient RdRp transfer^{77–79}.

Tombusvirus sgmRNA transcription. An alternative mechanism for the production of sgmRNAs involves the premature termination of the viral RdRp during negative-strand synthesis of the genome. This results in the generation of a 3'-truncated negative-strand that is then used as a template for the synthesis of positive-strand sgmRNAs^{1,80} (FIG. 5b). Premature termination is facilitated by an RNA attenuation signal in the positive-strand genome, which is formed by a base-paired RNA segment that is located just upstream of the sgmRNA start site. This signal functions as a physical 'roadblock' that induces polymerase termination. Notably, the attenuation signals that promote the transcription of sgmRNA1 and sgmRNA2 in TBSV are formed by long-range interactions that involve activator and receptor sequences, which are known as AS1–RS1 and AS2–RS2, respectively^{80,81} (FIG. 5b). AS1–RS1 spans ~1.1 kb, whereas AS2–RS2 spans ~2.2 kb and requires an auxiliary ~1.0 kb long-range interaction between a distal element and a core element (DE–CE)^{82,83}. Interestingly, the identity of the nucleotides that form the attenuation signals is not important⁸³; however, the stability of the base-paired segments was found to be important⁸⁴.

Nodavirus and dianthovirus sgmRNA transcription. sgmRNA transcription in the bisegmented insect nodavirus flock house virus (FHV) is also likely to occur via a premature termination mechanism. A sequence that is located in the central region of its larger genome segment interacts with two different downstream sequences, one of which is located more than 1.4 kb away and is positioned directly in front of the sgmRNA transcription start site⁸⁵. The double-stranded RNA structure that forms ahead of the initiation site is probably functionally comparable to the attenuation signals in TBSV.

It has also been proposed that another bisegmented virus — the dianthovirus RCNMV — uses a premature termination mechanism for transcription⁸⁶. However, for RCNMV, the attenuation signal forms in *trans* between two complementary sequences that are located separately in the RNA1 and RNA2 genome segments. During infections, the increase in the concentrations of the two genome segments promotes the formation of this bimolecular interaction, which, in turn, activates sgmRNA transcription. Accordingly, it was proposed that this interaction provides a concentration-dependent mechanism to temporally synchronize the transcription of the capsid protein-encoding sgmRNA with the accumulation of the two genomic segments that are co-packaged⁸⁶.

Roles and regulation of long-range interactions

As described in the above sections, long-range interactions have diverse functions. These include the relocalization of bound proteins to a distal genomic location (for example, repositioning 3' CITE-bound factors near to 5' termini or repositioning 5'-proximally bound RdRps near to 3' termini), the generation of a bipartite RNA platform for the assembly of protein complexes (for example, the RII–RIV RNA platform that is used for tombusvirus replicase complex assembly), the colocalization of two RNA elements that require proximity for function (for example, TRS elements involved in RdRp repriming in coronavirus transcription) and the formation of RNA structures that directly regulate a viral process (for example, the double-stranded attenuation signals that direct the premature termination of RdRps during sgmRNA transcription). In some cases, the need for a long-range interaction is clear, such as in relocating proteins; however, in other cases, the requirement is less obvious. For example, the attenuation signals that are involved in sgmRNA transcription in tombusviruses can be functionally replaced by local RNA hairpins, which indicates that the long-distance interactions are not essential⁸³. Interestingly, some viruses that are related to tombusviruses (for example, carmoviruses and necroviruses) use local, rather than long-range, transcription-attenuation signals^{87–89}. Thus, the differences that are observed could simply reflect the random nature of the emergence of local versus long-range interactions during virus evolution. Nevertheless, long-range interactions may provide yet-to-be discovered regulatory advantages that could be mediated via genome-level RNA rearrangement.

Long-distance base-pairing interactions can be regulated by several mechanisms. The most basic strategy is to modulate the stability of the base-paired region by altering the composition and/or number of nucleotides that are involved. The presence of competing RNA structures provides an additional mechanism, which is exemplified by the RNA switches that regulate the interactions that are required for translation and/or replication in HCV⁴⁰, replication in DENV⁶⁸ and readthrough in CIRV⁵⁰ (as described in the previous sections). Furthermore, long-range interactions might also be regulated by proteins that could facilitate or prevent the formation of an interaction and/or destabilize or stabilize an interaction. In line with such concepts, the flavivirus core protein and NS3 helicase have been shown to facilitate circularization *in vitro*^{63,64}, and it seems probable that proteins also modulate some of the interactions in other systems. Interactions can also be mediated by several different contacts, as has been reported for flavivirus circularization^{53–61}, and this suggests that cooperative effects and different regulatory mechanisms for individual contacts could also have a role⁶⁰.

Intramolecular interactions are not generally predicted to be influenced by the local genome concentration; however, as crowding agents have been shown to increase the folding of small RNAs *in vitro*⁹⁰, it is possible that high concentrations of viral RNAs *in vivo* could also

Replicase complex

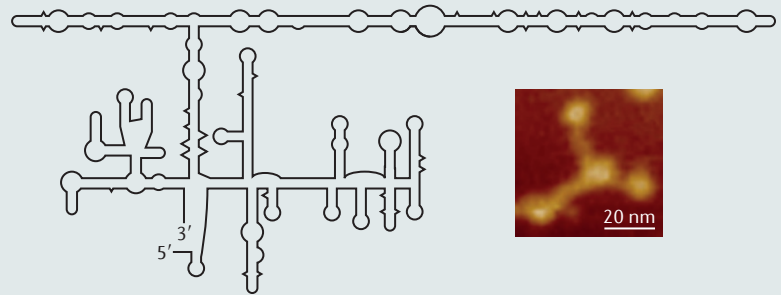
A membrane-associated complex that consists of the viral RNA-dependent RNA polymerase and viral or host proteins; it mediates the replication of viral RNA genomes and the transcription of viral subgenomic mRNAs.

Box 1 | Global organization of the STMV and TBSV RNA genomes

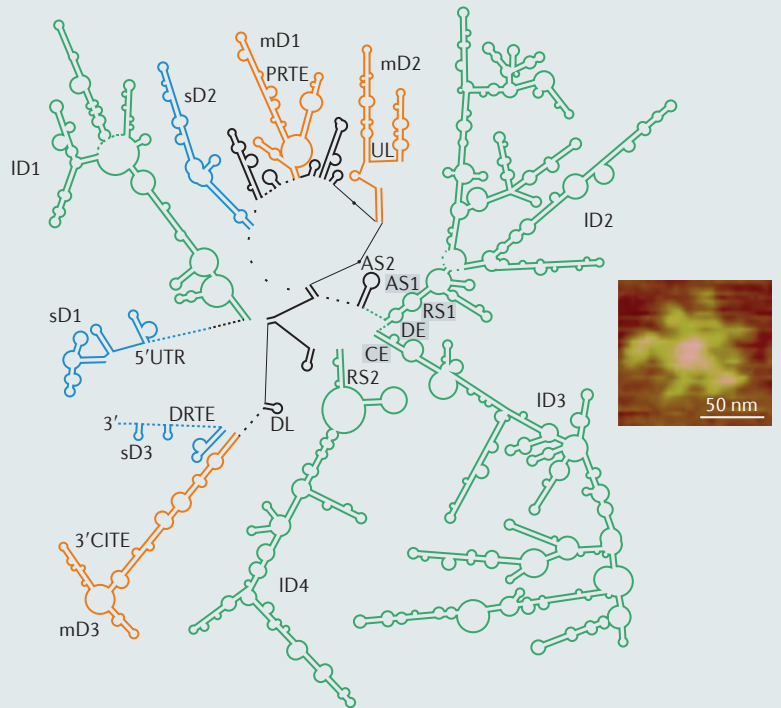
The RNA secondary structure models for satellite tobacco mosaic virus (STMV; see the figure, part a) and tomato bushy stunt virus (TBSV; see the figure, part b) were generated by incorporating SHAPE (selective 2'-hydroxyl acylation analysed by primer extension)-reactivity data into the 'RNAstructure' program. The global organization of the ~1.1 kb STMV RNA genome includes three domains that are formed by long-range base-pairing interactions⁹³. Consistent with this structural model, a fraction of the STMV genomes that were examined by atomic force microscopy (AFM) were found to adopt three-branch structures⁹³ (see the figure, part a). Although a second study that examined the STMV structure by SHAPE predicted different terminal configurations⁹⁴, the long-range interactions that define the large central domain were consistent in both models^{93,94}. The role of these long-range interactions in the STMV reproductive cycle remains to be determined; however, it was suggested that the helices that form these and other interactions could be important for genome packaging by interacting with capsid proteins within assembled particles^{93,94}.

SHAPE-based structural analysis was also used to study the larger ~4.8 kb TBSV RNA genome, which contains six known functional long-range interactions: 3' cap-independent translational enhancer (3' CITE)–5' UTR, proximal readthrough element–distal readthrough element (PRTE–DRTE), upstream linker–downstream linker (UL–DL), activator sequence 1–receptor sequence 1 (AS1–RS1), AS2–RS2 and distal element–core element (DE–CE)⁹⁵. Interestingly, only two (AS1–RS1 and DE–CE) of these six known interactions (TABLE 1) were predicted to form in the proposed genome structure. Although the complementary segments that were involved in the other four interactions were not paired, the global organization of the genome brings these partner sequences into proximity, which suggests that the existing framework could allow for the formation of these additional interactions without the need for large-scale rearrangements⁹⁵. The TBSV genome structure also revealed the presence of differently sized domains (small domain (sD), blue; medium domain (mD), orange; and large domain (ID), green), and the large domains (which are 500–2,000 nucleotides long) correspond to different coding regions. These domains emanate from a central hub that is formed by long-range base-pairing interactions, and AFM images of the TBSV genome showed that it had compact floret-like shapes that were consistent with a multidomain structure⁹⁵ (see the figure, part b). This structural model for the TBSV genome is one of several different possible conformations and provides a reference structure with which those that have been

a STMV genome



b TBSV genome



deduced under alternative conditions can be compared. Part a of the figure is adapted, with permission, from Archer, E. J. *et al.* Long-range architecture in a viral RNA genome. *Biochemistry* 52, 3182–3190 © 2013 American Chemical Society. Part b of the figure is adapted, with permission, from REF. 95 © 2013 Wu *et al.*

influence the formation or stabilization of intramolecular long-range interactions. Alternatively, if some of the proposed *cis*-interactions do actually occur in *trans*, as has been suggested⁷⁰, then genome abundance would clearly be an additional mode of control (as has been proposed for sgRNA transcription in RCNMV⁸⁶ and for frameshifting in SARS-CoV⁴⁹). However, a dependence on *cis*-only interactions could be advantageous, as it could provide a form of quality control for genetic completeness to viruses that use intragenomic interactions by selecting for viral genomes that maintain the interacting sequences and presumably intact intervening sequences^{4,65}.

Whole-genome context of long-range interactions

Although we are beginning to understand the function and regulation of long-range interactions, it remains unclear how they are able to function in the complex context of viral RNA genomes that have multiple functions. Some interactions involve overlapping sequences and are therefore mutually exclusive, whereas others promote processes that are opposed either physically (such as translation and replication) or temporally (such as replication and encapsidation). Accordingly, proper coordination of these interactions must be essential for their function. This regulation would be particularly relevant for viruses that have multiple long-range interactions,

SHAPE structural mapping (Selective 2'-hydroxyl acylation analysed by primer extension structural mapping). An RNA structure-probing technique that modifies the backbone in a base-nonspecific manner at positions of base flexibility. Highly modified positions correspond to unpaired nucleotides; this information is used to more accurately predict RNA secondary structure.

such as tombusviruses, which have at least six different functional long-distance interactions, all of which span distances of 1 kb or more (TABLE 1). In such cases, the global structure of the viral genome must have features that enable it to form each of the different interactions at the correct time. Thus, genomes must be dynamic and able to adopt alternative conformations, which would be influenced by the intrinsic features of the RNA and its environmental context. Indeed, the structure of the viral genome is likely to be distinct during different steps of the viral life cycle — for example, when the genome is encapsidated, has just been released from its capsid, is being translated, is being replicated or transcribed or is undergoing packaging. Other events, such as co-replicative 5'-to-3' folding of the genome during its synthesis, would also influence initial and ultimate structures and the ability to form different long-range interactions¹⁰. The factors that govern structural transitions at the genomic level are therefore of great interest.

The study of diverse genome states under different *in vitro* and *in vivo* conditions will assist in gaining an understanding of the complex coordination of alternative conformations. Ideally, long-range interactions should be studied in their natural genomic contexts, and a logical first step would be to obtain information about the secondary structure of viral RNA genomes. These studies are now possible, owing to technical advances, such as high-throughput SHAPE structural mapping (selective 2'-hydroxyl acylation analysed by primer extension structural mapping)^{91,92}. This chemical probing method provides information about nucleotide flexibility at each position in an RNA, which positively correlates with the likelihood that a residue is single-stranded in the structure. The SHAPE reactivity data can be incorporated into a thermodynamic-based RNA secondary structure-predicting program as a pseudo-free energy parameter to improve model prediction^{91,92}. Using this approach, the global organization of the 1,058 nucleotide-long satellite tobacco mosaic virus (STMV)^{93,94} RNA genome and the 4,778 nucleotide-long TBSV RNA genome⁹⁵ have been predicted, and the results have provided insights into the genomic contexts of long-range interactions (BOX 1).

Conclusions and perspectives

In this Review, we have discussed examples that highlight the important and varied roles of long-range RNA–RNA interactions in fundamental viral processes, including the translation of viral proteins, the replication of the viral genome and the transcription of sgmRNAs. It is clear that substantially different types of positive-strand RNA viruses use this distinctive regulatory strategy.

These viruses have evolved to integrate long-range interactions within their genomes in a manner that provides them with mechanisms to regulate a diverse array of viral functions. Indeed, it is quite remarkable that these relatively simple interactive structural features are able to carry out such a broad range of structure-based functions. Importantly, these interactions provide the viruses with a unique opportunity for regulation that is linked to genome structure; this, in turn, may provide benefits in addition to those of local RNA elements. Many important advances have been made in understanding the structure and function of long-range interactions. Unfortunately, not all interactions are amenable to the available investigative methodologies. For example, interactions that are transient or that require special conditions in order to form may be difficult to detect biochemically and interactions that have unknown, functionally relevant, alternative base-pair partners may be intractable to genetic techniques, such as compensatory mutational analyses. Regardless of the challenges, the interactions that are supported by both biophysical and genetic evidence should be viewed as having increased credibility. For the future, additional research and technological advances are needed to expand on existing findings and to clarify open questions; for example, it will be crucial to determine the detailed structures of different interactions and to establish how these structures influence activity. Other important challenges will be to identify the dynamics and energetic barriers of transitions and folding pathways. In addition, future research is needed to determine the regulatory advantage of long-range interactions, how these interactions are integrated and coordinated within viral genomes and whether any of these interactions occur in *trans*. Moreover, it will be crucial to investigate the possible involvement of viral and/or host proteins in regulating interactions and how the cellular environment affects interactions. An increased understanding of long-range interactions will also help to determine whether such interactions are plausible targets for antiviral therapies. Finally, another area for future consideration extends beyond viral contexts to cellular messages: how common are functional long-range RNA–RNA interactions in cellular mRNA biogenesis and function? A limited number of recent reports suggest that these structures can indeed participate in the regulation of pre-mRNA splicing⁹⁶ as well as in the control of eukaryotic⁹⁷ and bacterial⁹⁸ mRNA translation. Consequently, the phenomenon that is observed in viruses may foreshadow a similar prevalence and diversity of functional long-range RNA–RNA interactions in cellular mRNAs.

1. Sztuba-Solińska, J., Stollar, V. & Bujarski, J. J. Subgenomic messenger RNAs: mastering regulation of (+) strand RNA virus life cycle. *Virology* **412**, 245–255 (2011).
2. Li, Z. & Nagy, P. D. Diverse roles of host RNA binding proteins in RNA virus replication. *RNA Biol.* **8**, 305–315 (2011).
3. Liu, Y., Wimmer, E. & Paul, A. V. *Cis*-acting RNA elements in human and animal plus-strand RNA viruses. *Biochim. Biophys. Acta* **1789**, 495–517 (2009).
4. Miller, W. A. & White, K. A. Long-distance RNA–RNA interactions in plant virus gene expression and replication. *Annu. Rev. Phytopathol.* **44**, 447–467 (2006).
5. Nicholson, B. L. & White, K. A. 3' cap-independent translation enhancers of positive-strand RNA plant viruses. *Curr. Opin. Virol.* **1**, 373–380 (2011).
6. Simon, A. E. & Miller, W. A. 3' cap-independent translation enhancers of plant viruses. *Annu. Rev. Microbiol.* **67**, 21–42 (2013).
7. Romero-López, C. & Berzal-Herranz, A. Unmasking the information encoded as structural motifs of viral RNA genomes: a potential antiviral target. *Rev. Med. Virol.* **23**, 340–354 (2013).
8. van Himbergen, J., van Geffen, B. & van Duin, J. Translational control by a long range RNA–RNA interaction; a basepair substitution analysis. *Nucleic Acids Res.* **21**, 1713–1717 (1993). **This paper is the first report of a functional long-distance RNA–RNA interaction in a positive-strand RNA virus.**
9. Licitis, N., van Duin, J., Balklava, Z. & Berzins, V. Long-range translational coupling in single-stranded RNA bacteriophages: an evolutionary analysis. *Nucleic Acids Res.* **26**, 3242–3246 (1998).

10. Beekwilder, M. J., Nieuwenhuizen, R. & van Duin, J. Secondary structure model for the last two domains of single-stranded RNA phage Q β . *J. Mol. Biol.* **247**, 903–917 (1996).
11. Klovin, J., Berzins, V. & van Duin, J. A long-range interaction in Q β RNA that bridges the thousand nucleotides between the M-site and the 3' end is required for replication. *RNA* **4**, 948–957 (1998).
12. Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93**, 1385–1409 (2012).
13. Dreher, T. W. & Miller, W. A. Translational control in positive strand RNA plant viruses. *Virology* **344**, 185–197 (2006).
14. Balvay, L., Soto Rifo, R., Ricci, E. P., Decimo, D. & Ohlmann, T. Structural and functional diversity of viral IRESes. *Biochim. Biophys. Acta* **1789**, 542–557 (2009).
15. Nicholson, B. L., Wu, B., Chevchenko, I. & White, K. A. Tombusvirus recruitment of host translational machinery via the 3' UTR. *RNA* **16**, 1402–1419 (2010).
16. Wang, Z., Treder, K. & Miller, W. A. Structure of a viral cap-independent translation element that functions via high affinity binding to the eIF4E subunit of eIF4F. *J. Biol. Chem.* **284**, 14189–14202 (2009).
17. Wang, Z., Parisien, M., Scheets, K. & Miller, W. A. The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. *Structure* **19**, 868–880 (2011).
18. McCormack, J. C. et al. Structural domains within the 3' untranslated region of Turnip crinkle virus. *J. Virol.* **82**, 8706–8720 (2008).
19. Zuo, X. et al. Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3' UTR of turnip crinkle virus. *Proc. Natl Acad. Sci. USA* **107**, 1385–1390 (2010).
20. Gazo, B. M., Murphy, P., Gatchel, J. R. & Browning, K. S. A novel interaction of cap-binding protein complexes eukaryotic initiation factor (eIF) 4F and eIF(iso)4F with a region in the 3'-untranslated region of satellite tobacco necrosis virus. *J. Biol. Chem.* **279**, 13584–13592 (2004).
21. Treder, K. et al. The 3' cap-independent translation element of Barley yellow dwarf virus binds eIF4F via the eIF4G subunit to initiate translation. *RNA* **14**, 134–147 (2008).
22. Nicholson, B. L., Zaslav, O., Mayberry, L. K., Browning, K. S. & White, K. A. Tombusvirus Y-shaped translational enhancer forms a complex with eIF4F and can be functionally replaced by heterologous translational enhancers. *J. Virol.* **87**, 1872–1883 (2013).
23. Stupina, V. A. et al. The 3' proximal translational enhancer of Turnip crinkle virus binds to 60S ribosomal subunits. *RNA* **14**, 2379–2393 (2008).
24. Gao, F., Kasprzak, W., Stupina, V. A., Shapiro, B. A. & Simon, A. E. A ribosome-binding, 3' translational enhancer has a T-shaped structure and engages in a long-distance RNA–RNA interaction. *J. Virol.* **86**, 9828–9842 (2012).
25. Guo, L., Allen, E. M. & Miller, W. A. Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. *Mol. Cell* **7**, 1103–1109 (2001).
- This is the first report of a long-range RNA–RNA interaction that facilitates translation initiation in a viral genome.**
26. Fabian, M. R. & White, K. A. 5'–3' RNA–RNA interaction facilitates cap- and poly(A) tail-independent translation of tomato bushy stunt virus mRNA: a potential common mechanism for Tombusviridae. *J. Biol. Chem.* **279**, 28862–28872 (2004).
27. Fabian, M. R. & White, K. A. Analysis of a 3'-translation enhancer in a tombusvirus: a dynamic model for RNA–RNA interactions of mRNA termini. *RNA* **12**, 1304–1314 (2006).
28. Nicholson, B. L. & White, K. A. Context-influenced cap-independent translation of Tombusvirus mRNAs in vitro. *Virology* **380**, 203–212 (2008).
29. Chattopadhyay, M., Shi, K., Yuan, X. & Simon, A. E. Long-distance kissing loop interactions between a 3' proximal Y-shaped structure and apical loops of 5' hairpins enhance translation of Saguaro cactus virus. *Virology* **417**, 113–125 (2011).
30. Rakotondrafara, A. M., Polacek, C., Harris, E. & Miller, W. A. Oscillating kissing stem–loop interactions mediate 5' scanning-dependent translation by a viral 3'-cap-independent translation element. *RNA* **12**, 1893–1906 (2006).
31. Gao, F. et al. The kl-TSS translational enhancer of PEMV can bind simultaneously to ribosomes and a 5' proximal hairpin. *J. Virol.* **87**, 11987–12002 (2013).
32. López de Quinto, S., Sáiz, M., de la Morena, D., Sobrino, F. & Martínez-Salas, E. IRES-driven translation is stimulated separately by the FMDV 3'-NCR and poly(A) sequences. *Nucleic Acids Res.* **30**, 4398–4405 (2002).
33. Serrano, P., Pulido, M. R., Sáiz, M. & Martínez-Salas, E. The 3' end of the foot-and-mouth disease virus genome establishes two distinct long-range RNA–RNA interactions with the 5' end region. *J. Gen. Virol.* **87**, 3013–3022 (2006).
34. Huang, S. W., Chan, M. Y., Hsu, W. L., Huang, C. C. & Tsai, C. H. The 3'-terminal hexamer sequence of classical swine fever virus RNA plays a role in negatively regulating the IRES-mediated translation. *PLoS ONE* **7**, e33764 (2012).
35. Romero-López, C. & Berzal-Herranz, A. A long-range RNA–RNA interaction between the 5' and 3' ends of the HCV genome. *RNA* **15**, 1740–1752 (2009).
36. Romero-López, C. & Berzal-Herranz, A. The functional RNA domain 5BSL3.2 within the NS5B coding sequence influences hepatitis C virus IRES-mediated translation. *Cell. Mol. Life Sci.* **69**, 103–113 (2012).
- References 35 and 36 report the identification of a 5'–3' RNA–RNA interaction in HCV and its role in regulating IRES-mediated translation.**
37. Romero-López, C., Barroso-Deljesus, A., García-Sacristán, A., Briones, C. & Berzal-Herranz, A. The folding of the hepatitis C virus internal ribosome entry site depends on the 3'-end of the viral genome. *Nucleic Acids Res.* **40**, 11697–11713 (2012).
38. Diviney, S. et al. A hepatitis C virus cis-acting replication element forms a long-range RNA–RNA interaction with upstream RNA sequences in NS5B. *J. Virol.* **82**, 9008–9022 (2008).
39. Tuplin, A., Struthers, M., Simmonds, P. & Evans, D. J. A twist in the tail: SHAPE mapping of long-range interactions and structural rearrangements of RNA elements involved in HCV replication. *Nucleic Acids Res.* **40**, 6908–6921 (2012).
40. Palau, W., Masante, C., Ventura, M. & Di Primo, C. Direct evidence for RNA–RNA interactions at the 3' end of the hepatitis C virus genome using surface plasmon resonance. *RNA* **19**, 982–991 (2013).
41. Shetty, S., Stefanovic, S. & Mihailescu, M. R. Hepatitis C virus RNA: molecular switches mediated by long-range RNA–RNA interactions? *Nucleic Acids Res.* **41**, 2526–2540 (2013).
- This is a biophysical study of HCV RNA elements that are involved in long-range interactions and their potential roles as RNA switches in viral life cycle regulation.**
42. Romero-López, C., Barroso-Deljesus, A., García-Sacristán, A., Briones, C. & Berzal-Herranz, A. End-to-end crossstalk within the hepatitis C virus genome mediates the conformational switch of the 3'X-tail region. *Nucleic Acids Res.* **42**, 567–582 (2014).
43. Friebe, P., Boudet, J., Simorre, J. P. & Bartenschlager, R. Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication. *J. Virol.* **79**, 380–392 (2005).
44. Karetnikov, A., Keränen, M. & Lehto, K. Role of the RNA2 3' non-translated region of Blackcurrant reversion nepovirus in translational regulation. *Virology* **354**, 178–191 (2006).
45. Karetnikov, A. & Lehto, K. J. The RNA2 5' leader of Blackcurrant reversion virus mediates efficient *in vivo* translation through an internal ribosomal entry site mechanism. *J. Gen. Virol.* **88**, 286–297 (2007).
46. Karetnikov, A. & Lehto, K. Translation mechanisms involving long-distance base pairing interactions between the 5' and 3' non-translated regions and internal ribosomal entry are conserved for both genomic RNAs of Blackcurrant reversion nepovirus. *Virology* **371**, 292–308 (2008).
47. Barry, J. K. & Miller, W. A. 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. USA* **99**, 11133–11138 (2002).
- This paper provides the first experimental evidence for a long-range interaction that facilitates frameshifting.**
48. Tajima, Y., Iwakawa, H. O., Kaido, M., Mise, K. & Okuno, T. 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 (2011).
49. Ishimaru, D. et al. RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucleic Acids Res.* **41**, 2594–2608 (2013).
50. Cimino, P. A., Nicholson, B. L., Wu, B., Xu, W. & White, K. A. Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. *PLoS Pathog.* **7**, e1002423 (2011).
- This paper provides the first experimental evidence for a long-range interaction that facilitates readthrough.**
51. Villordo, S. M. & Gamarnik, A. V. Genome cyclization as strategy for flavivirus RNA replication. *Virus Res.* **139**, 230–239 (2009).
52. Wu, B. et al. A discontinuous RNA platform mediates RNA virus replication: building an integrated model for RNA-based regulation of viral processes. *PLoS Pathog.* **5**, e1000323 (2009).
53. You, S., Falgout, B., Markoff, L. & Padmanabhan, R. *In vitro* RNA synthesis from exogenous dengue viral RNA templates requires long range interactions between 5' and 3'-terminal regions that influence RNA structure. *J. Biol. Chem.* **276**, 15581–15591 (2001).
54. Khromykh, A. A., Meka, H., Guyatt, K. J. & Westaway, E. G. Essential role of cyclization sequences in flavivirus RNA replication. *J. Virol.* **75**, 6719–6728 (2001).
55. Corver, J. et al. Fine mapping of a cis-acting sequence element in yellow fever virus RNA that is required for RNA replication and cyclization. *J. Virol.* **77**, 2265–2270 (2003).
56. Alvarez, D. E., Lodeiro, M. F., Luduena, S. J., Pietrasanta, L. I. & Gamarnik, A. V. Long-range RNA–RNA interactions circularize the dengue virus genome. *J. Virol.* **79**, 6631–6643 (2005).
- This report uses AFM to visualize single viral RNA molecule circularization and identifies a second long-range interaction that is required for Dengue virus replication.**
57. Alvarez, D. E., Filomatori, C. V. & Gamarnik, A. V. Functional analysis of dengue virus cyclization sequences located at the 5' and 3'UTRs. *Virology* **375**, 223–235 (2008).
58. Zhang, B., Dong, H., Stein, D. A., Iversen, P. L. & Shi, P. Y. West Nile virus genome cyclization and RNA replication require two pairs of long-distance RNA interactions. *Virology* **373**, 1–13 (2008).
59. Zhang, B., Dong, H., Ye, H., Tilgner, M. & Shi, P. Y. Genetic analysis of West Nile virus containing a complete 3'CSI RNA deletion. *Virology* **408**, 138–145 (2010).
60. Friebe, P. & Harris, E. Interplay of RNA elements in the dengue virus 5' and 3' ends required for viral RNA replication. *J. Virol.* **84**, 6103–6118 (2010).
61. Friebe, P., Shi, P. Y. & Harris, E. The 5' and 3' downstream AUG region elements are required for mosquito-borne flavivirus RNA replication. *J. Virol.* **85**, 1900–1905 (2011).
62. Sztuba-Solinska, J. et al. Structural complexity of Dengue virus untranslated regions: cis-acting RNA motifs and pseudoknot interactions modulating functionality of the viral genome. *Nucleic Acids Res.* **41**, 5075–5089 (2013).
63. Ivanyi-Nagy, R. & Darlix, J. L. Core protein-mediated 5'–3' annealing of the West Nile virus genomic RNA *in vitro*. *Virus Res.* **167**, 226–235 (2012).
64. Gebhard, L. G., Kaufman, S. B. & Gamarnik, A. V. Novel ATP-independent RNA annealing activity of the dengue virus NS3 helicase. *PLoS ONE* **7**, e36244 (2012).
65. Filomatori, C. V. et al. A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev.* **20**, 2238–2249 (2006).
- This study reports the identification of a 5'-proximal RNA element that is recognized by the viral RdRp and proposes a model that involves the relocation of the RdRp to the 3'-terminus via long-range RNA–RNA base pairing.**
66. Dong, H., Zhang, B. & Shi, P. Y. Terminal structures of West Nile virus genomic RNA and their interactions with viral NS5 protein. *Virology* **381**, 123–135 (2008).
67. Liu, Z. Y. et al. Novel cis-acting element within the capsid-coding region enhances flavivirus viral-RNA replication by regulating genome cyclization. *J. Virol.* **87**, 6804–6818 (2013).
68. Villordo, S. M., Alvarez, D. E. & Gamarnik, A. V. A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA* **16**, 2325–2335 (2010).
69. Villordo, S. M. & Gamarnik, A. V. Differential RNA sequence requirement for dengue virus replication in mosquito and mammalian cells. *J. Virol.* **87**, 9365–9372 (2013).

70. Lott, W. B. & Doran, M. R. Do RNA viruses require genome cyclisation for replication? *Trends Biochem. Sci.* **38**, 350–355 (2013).
71. Rajendran, K. S. & Nagy, P. D. Kinetics and functional studies on interaction between the replicase proteins of Tomato bushy stunt virus: requirement of p33:p92 interaction for replicase assembly. *Virology* **345**, 270–279 (2006).
72. Pogany, J., White, K. A. & Nagy, P. D. Specific binding of tombusvirus replication protein p33 to an internal replication element in the viral RNA is essential for replication. *J. Virol.* **79**, 4859–4869 (2005).
73. Pathak, K. B., Pogany, J., Xu, K., White, K. A. & Nagy, P. D. Defining the roles of cis-acting RNA elements in tombusvirus replicase assembly *in vitro*. *J. Virol.* **86**, 156–171 (2012).
74. Nicholson, B. L., Lee, P. K. & White, K. A. Internal RNA replication elements are prevalent in Tombusviridae. *Front. Microbiol.* **3**, 279 (2012).
75. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37 (2006).
76. Sawicki, S. G., Sawicki, D. L. & Siddell, S. G. A contemporary view of coronavirus transcription. *J. Virol.* **81**, 20–29 (2007).
77. Moreno, J. L., Zúñiga, S., Enjuanes, L. & Sola, I. Identification of a coronavirus transcription enhancer. *J. Virol.* **82**, 3882–3893 (2008).
78. Mateos-Gómez, P. A., Zúñiga, S., Palacio, L., Enjuanes, L. & Sola, I. Gene N proximal and distal RNA motifs regulate coronavirus nucleocapsid mRNA transcription. *J. Virol.* **85**, 8968–8980 (2011).
79. Mateos-Gomez, P. A., Morales, L., Zúñiga, S., Enjuanes, L. & Sola, I. Long distance RNA–RNA interactions in the coronavirus genome form high-order structures promoting discontinuous RNA synthesis during transcription. *J. Virol.* **87**, 177–186 (2013). **This paper reports the characterization of an extremely long-range RNA–RNA interaction — which is the longest interaction that has been reported so far — spanning ~26,000 nucleotides,**
- and is required for the production of sgmRNAs in a coronavirus.**
80. Jiwan, S. D. & White, K. A. Subgenomic mRNA transcription in Tombusviridae. *RNA Biol.* **8**, 287–294 (2011).
81. Choi, I. R. & White, K. A. An RNA activator of subgenomic mRNA1 transcription in tomato bushy stunt virus. *J. Biol. Chem.* **277**, 3760–3766 (2002).
82. Choi, I. R., Ostrovsky, M., Zhang, G. & White, K. A. Regulatory activity of distal and core RNA elements in tombusvirus subgenomic mRNA2 transcription. *J. Biol. Chem.* **276**, 41761–41768 (2001).
83. Lin, H. & White, K. A. A complex network of RNA–RNA interactions controls subgenomic mRNA transcription in a tombusvirus. *EMBO J.* **23**, 3365–3374 (2004).
84. Wang, S., Mortazavi, L. & White, K. A. Higher-order RNA structural requirements and small-molecule induction of tombusvirus subgenomic mRNA transcription. *J. Virol.* **82**, 3864–3871 (2008).
85. Lindenbach, B. D., Sgro, J. Y. & Ahlquist, P. Long-distance base pairing in flock house virus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. *J. Virol.* **76**, 3905–3919 (2002).
86. Sit, T. L., Vaewhongs, A. A. & Lommel, S. A. RNA-mediated trans-activation of transcription from a viral RNA. *Science* **281**, 829–832 (1998).
87. Xu, W. & White, K. A. RNA-based regulation of transcription and translation of arenavirus subgenomic mRNA1. *J. Virol.* **83**, 10096–10105 (2009).
88. Jiwan, S. D., Wu, B. & White, K. A. Subgenomic mRNA transcription in tobacco necrosis virus. *Virology* **418**, 1–11 (2011).
89. Wu, B., Oliveri, S., Mandic, J. & White, K. A. Evidence for a premature termination mechanism of subgenomic mRNA transcription in a carmovirus. *J. Virol.* **84**, 7904–7907 (2010).
90. Strulson, C. A., Boyer, J. A., Whitman, E. E. & Bevilacqua, P. C. Molecular crowders and cosolutes promote folding cooperativity of RNA under physiological ionic conditions. *RNA* **20**, 331–347 (2014).
91. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
92. Low, J. T. & Weeks, K. M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150–158 (2010).
93. Archer, E. J. *et al.* Long-range architecture in a viral RNA genome. *Biochemistry* **52**, 3182–3190 (2013).
94. Athavale, S. S. *et al.* *In vitro* secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PLoS ONE* **8**, e54384 (2013).
95. Wu, B., Griggall, J., Ore, M. O., Morin, S. & White, K. A. Global organization of a positive-strand RNA virus genome. *PLoS Pathog.* **9**, e1003363 (2013). **This paper reports secondary-structure modelling of the ~4.8 kb RNA genome of TBSV and analyses newly identified structures and interactions.**
96. Lovci, M. T. *et al.* Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nature Struct. Mol. Biol.* **20**, 1434–1442 (2013).
97. Chen, J. & Kastan, M. B. 5'-3'-UTR interactions regulate p53 mRNA translation and provide a target for modulating p53 induction after DNA damage. *Genes Dev.* **24**, 2146–2156 (2010).
98. Ruiz de Los Mozos, I. *et al.* Base pairing interaction between 5'- and 3'-UTRs controls icaR mRNA translation in *Staphylococcus aureus*. *PLoS Genet.* **9**, e1004001 (2013).

Acknowledgements

The authors apologize to colleagues whose research could not be cited owing to space limitations. The work in the authors' laboratory is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Competing interests statement

The authors declare no competing interests.