Review

Unconventional viral gene expression mechanisms as therapeutic targets

https://doi.org/10.1038/s41586-021-03511-5

Received: 8 June 2020

Accepted: 22 March 2021

Published online: 19 May 2021

Check for updates

Jessica Sook Yuin Ho[1,3], Zeyu Zhu[1,3] & Ivan Marazzi[1,2]✉

Unlike the human genome that comprises mostly noncoding and regulatory sequences, viruses have evolved under the constraints of maintaining a small genome size while expanding the efficiency of their coding and regulatory sequences. As a result, viruses use strategies of transcription and translation in which one or more of the steps in the conventional gene–protein production line are altered. These alternative strategies of viral gene expression (also known as gene recoding) can be uniquely brought about by dedicated viral enzymes or by co-opting host factors (known as host dependencies). Targeting these unique enzymatic activities and host factors exposes vulnerabilities of a virus and provides a paradigm for the design of novel antiviral therapies. In this Review, we describe the types and mechanisms of unconventional gene and protein expression in viruses, and provide a perspective on how future basic mechanistic work could inform translational efforts that are aimed at viral eradication.

Expression of a gene in the human genome is a multistep and heavily regulated process that resembles a production line. Protein-coding genes are transcribed almost exclusively by RNA polymerase II (RNAPII). During transcription, quality-control checkpoints are implemented to ensure that a gene is properly recognized and transcribed. A number of factors (epigenetic enzymes, chromatin remodellers, transcription factors and activators–coactivators) ensure gene recognition and RNAPII progression on the genic template. The progression of RNAPII—which includes RNAPII initiation, pause–release, elongation and the termination of transcription—occurs in sync with co-transcriptional events (that is, 5′ capping, splicing and polyadenylation). The end result of gene transcription and RNA processing is the generation of a mature RNA, in which coding exons are fused in a linear order that depends on the isoform of the gene. Mature mRNA is subsequently exported from nucleus into the cytoplasm, where it is directed to ribosomes for translation. The canonical model of translation initiation starts with recognition of the 7-methylguanylate cap on the 5′ end of most eukaryotic mRNA by the initiation factor eIF4, which recruits a pre-initiation complex that comprises the 40S ribosomal subunit and several eukaryotic initiation factors (eIF3, eIF1, eIF1A and the ternary complex eIF2–GTP–Met–$tRNA_i^{Met}$). This complex then scans continuously from the 5′ to the 3′ end for the first initiation codon in an optimal context (the RCCAUGG Kozak sequence, in which R stands for purine)[1]. Once the start codon of a gene is read by the initiator $tRNA^{Met}$, translation progresses and ends when a stop codon in the mRNA (UAA, UAG or UGA) is recognized by release factors. Depending on the subcellular localization of a given protein, co- and post-translational events might take place to sort proteins to their destinations. In brief, this is the conventional eukaryotic production line through which a gene makes a protein ready to be used in the cell.

To overcome their small genomes and increase their coding capacity, viruses have evolved to co-opt the transcriptional, epigenetic and translational mechanisms of the infected host cell. To generate protein diversity, viruses can adopt the existing mechanisms of the host (for example, alternative splicing) or use unique strategies. Here we describe the diverse ways by which viral genomes give rise to genes and proteins that deviate from the canonical framework of human genes, restricting our analyses to eukaryotes and their viruses.

Small-genome solutions to big problems

A main strategy to increase the number of coded proteins from a small genome is the use of overlapping or overprinted genes. Nucleic acid sequences can simultaneously encode two or more proteins in alternative reading frames (ARFs). To synthesize these proteins, unconventional transcriptional ('copying') or translational ('reading') events need to take place (Fig. 1). Although a comprehensive characterization of gene overprinting in large mammalian genomes is lacking, estimates on the basis of simulating codon use[2] or ribosome footprinting[3] suggest that only 1% of human genes are overprinted. By contrast, gene overlapping is very common among viruses. Despite differences in the size and structure of viral genomes, 53% of sequenced viral genomes containing at least one pair of genes that overlap for more than 50 nucleotides[4]. Proteins that originate by overprinting often encode accessory proteins that feature short sequences, and can provide a selective advantage for viruses[5–7]. Many overlapping genes are fixed in viral genomes because of their functions as host antagonists, such as those that affect the interferon response of the host[8,9], suppress RNA interference[10], and induce apoptosis of host cells[11]. In addition, as a mutation in an overlapping genomic region affects both the canonical and the overprinted genes, overlapping genes may also serve as a safety mechanism that protects the virus from deleterious mutations. However, because proteins that are encoded by gene overprinting are often enriched in disordered regions and show a tendency to have no known homologues[12,13], many overprinting viral proteins are poorly characterized.

Another challenge that is inherent to a small genome is a lack of regulatory space for maintaining the correct stoichiometry and temporality

[1]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]These authors contributed equally: Jessica Sook Yuin Ho, Zeyu Zhu. ✉e-mail: ivan.marazzi@mssm.edu

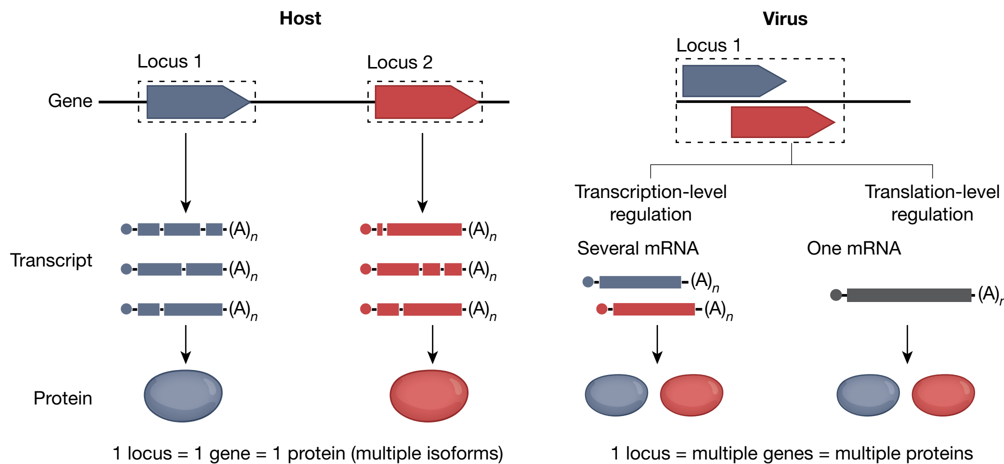

Fig. 1 | The host and virus adopt different strategies for gene expression as a result of differences in genome size. Left, in organisms with a large genome, expression of a cellular gene typically follows a linear pathway that leads to the synthesis of the respective canonical protein product. Right, viruses, which are confined by their much-smaller genome sizes, use unconventional pathways that mostly involve transcription-level (decoding multiple messages (mRNA)) or translation-level regulation to generate several protein products from a single locus.

of the expression of overprinted proteins. To overcome these limitations, viruses use several methods that include (1) intrinsic *cis* and *trans* regulation of polymerase and other enzymatic activities and (2) a codependency on host functions. We summarize the most relevant strategies used by viruses for expanding the coding and regulatory potentials of their overlapping genes, focusing mostly on viruses that are human pathogens and that represent current and future threats.

Expression of overlapping genes

Copying multiple messages

One set of strategies used by viruses to increase the efficiency of their small genomes involves transcriptional mechanisms that generate several mRNAs from overprinted coding sequences.

Transcriptional slippage. Transcriptional slippage is a process in which several overlapping transcripts are generated from the same gene via viral RNA polymerase stuttering, which results in the incorporation (and, occasionally, the deletion) of one or more nucleotides in the transcript (Fig. 2a). Sequences that are prone to transcriptional slippage include homopolymeric A/T tracts, the U6A motif in human immunodeficiency virus (HIV)[14], and the UC-rich slippery sequence in the paramyxoviruses[15]. The efficiency of transcriptional slippage is regulated by the stability and length of the nascent RNA relative to the template RNA, as well as by the structure of RNA-dependent RNA polymerase (RdRp)[15]. Owing to frameshift upon the insertion of nucleotides, the translation of overlapping transcripts typically results in proteins with a common N-terminus, but different C termini. Aside from using transcriptional slippage to generate mRNAs in different reading frames, some virus also use it to polyadenylate their mRNAs[16].

Transcriptional slippage was first identified in the synthesis of V proteins from the phosphoprotein (*P*) gene in *Parainfluenza virus 5* (previously known as *Simian virus 5*)[17], and has subsequently been observed in other pathogenic RNA viruses: mostly of members of *Mononegavirales*, including viruses in the *Paramxyoviridae* (such as Sendai virus) and *Filoviridae* (such as ebolavirus). Positive-strand viruses in the *Potyviridae*[18] and *Flaviviridae*[19] families have also been described as using this mechanism. In paramyxoviruses, transcriptional slippage can occur when RdRp encounters a 'slippery' sequence of 3′-UUUUUUCC̲C-5′ in the *P* gene and stutters at the underlined cytidine[15]. The polymerase then backtracks and realigns the newly synthesized mRNA with the template by non-destabilizing G:U base-pairing, which results in G insertions. The possible number of G insertions is limited to six by a sequence that contains adenosine that is located immediately upstream of the slippery site (as A:A base-pairing is not tolerated)[20]. In Sendai virus, at least three distinct mRNAs of the *P* gene are produced by transcriptional slippage. The unedited mRNA encodes P protein, which is a component of RdRp that regulates transcriptional fidelity and limits antiviral responses[21,22]. mRNA with +1 G or +2 G insertions code for two accessory proteins (V and W, respectively), both of which regulate viral replication kinetics and the activation of host responses[23,24]. Additionally, the unique hexameric genome-packaging rule of paramyxovirus might regulate the efficiency of mRNA editing mediated by transcriptional slippage in this virus[20,25], as it has been shown that mRNA editing is at its most extensive when the cytidine at which the RdDp stutters is in position 2 or 5 in a hexamer, which suggests that N proteins might remain in close proximity to RdRp during transcription[26]. Further examples of transcriptional slippage occur in ebolaviruses and Marburg viruses[27], both of which belong to the *Filoviridae* family. In ebolavirus, transcriptional slippage occurs at a 30% frequency on a stretch of seven uridines in the glycoprotein (*GP*) gene and results in the insertion of one or two additional adenines in the mRNA[28–31]. The unedited transcript translates into a nonstructural and secreted glycoprotein[28], and the +1 A and +2 A shifts result in an extended glycoprotein that bears a transmembrane domain and a small soluble glycoprotein, repsectively[28]. More recently, deep mRNA sequencing has revealed other possible polyuridine transcriptional slippage sites in the *GP*, *NP*, *VP30* and *L* mRNAs of ebolavirus[27], which suggests that there may be more uncharacterized polypeptide species expressed than has previously been believed.

RNA splicing. RNA splicing is a commonly used and tightly regulated eukaryotic mechanism of generating distinct mature transcripts from a single gene, and has also been exploited by several families of viruses that replicate in the host nucleus, such as members of the *Adenoviridae* and *Parvoviridae* (DNA viruses), retroviruses, and members of the *Bornaviridae* and *Orthomyxoviridae* (RNA viruses). However, because of the more compact nature of viral genomes, splicing in viruses—unlike in humans—often serves to express overprinted genes.

In the segmented RNA genome of influenza A viruses (IAV), splicing occurs in viral segments 8 (which encodes the *NS* gene), 7 (which encodes the *M* gene) and 2 (which encodes the *PB1* gene). Depending on the viral strain, up to three or four unique mRNAs can be generated from segments 8 and 7, respectively. The noncanonical proteins that are produced by splicing are involved in important functions, such as the nuclear export of viral RNA and host adaptation[32,33]. Importantly, the splicing of segments 7and 8 is regulated by an array of viral and host factors that includes *trans* regulators of splicing, such as NS1-BP, HNRNPK[34], SRSF1 (also known as SF2/ASF)[35], SRSF3[36] and protein kinase CLK1[36]. Finally, *cis*-regulatory RNA secondary structures at the 3′ splice site of segment 7 have been suggested to be potential regulators of splicing efficiency in IAV[37,38], and a determinant of host tropism[37].

Circular RNA is a relatively stable and exonuclease-resistant RNA that is produced by backsplicing, and has recently been identified[39]

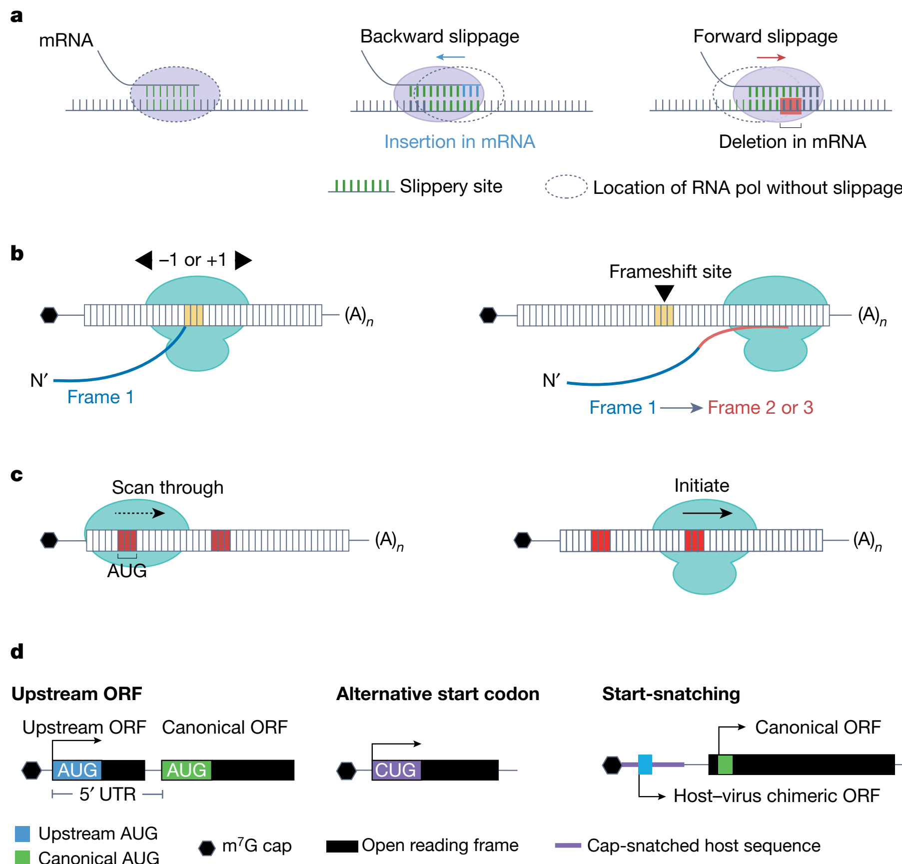

Fig. 2 | Small-genome solutions to expanding coding potential. a, Polymerase frameshifting, in which backward or forward slippage of RNA polymerase (pol) results in nucleotide insertions or deletions, and generates a heterogeneous population of viral mRNAs. **b**, PRFs lead to the synthesis of viral proteins from several reading frames. **c**, Leaky scanning, in which the ribosome scans through and skips an AUG start codon that is typically located in a less-optimal sequence context, and initiates at a downstream start codon. **d**, Generation of noncanonical sites of translation initiation through upstream ORFs or non-AUG start codons. In start-snatching, an upstream AUG start codon is obtained via cap-snatching of host RNA (which enables the translation of novel proteins on the basis of both host and viral genetic information).

across many viruses—including members of the gammaherpesvirus family (Epstein–Barr virus and Kaposi sarcoma virus) and the oncogenic human papillomaviruses. The functions of circular RNA in viruses are largely unknown, but a recent study has shown that knockdown of the E7 circular RNA produced by human papillomavirus 16 using short hairpin RNA inhibits oncogenic transformation of infected cells[40].

Reading multiple messages

Other mechanisms used by viruses to expand the set of proteins expressed from their small genomes include those that act at the level of mRNA translation, which allow for the expression of multiple overprinted proteins from one mRNA.

Programmed ribosome frameshifting. Programmed ribosomal frameshifts (PRFs) (Fig. 2b) occur when elongating ribosomes slip by one base upstream (5′, known as a −1 PRF) or downstream (3′, known as a +1 PRF), thus shifting the ribosomal reading frame. PRFs allow for the expression of overprinted proteins from the same mRNA and can also serve to regulate the stoichiometry of viral proteins. There are two prerequisites for a −1 PRF: (1) a slippery site with the sequence motif XXXYYYZ (in which X is any three identical nucleotides, Y represents U or A, and Z is A, C or U (although with some exceptions, such as GGU); as has previously been reviewed in detail[41,42]) and (2) a downstream pseudoknot structure that comprises two stems and a connecting loop as a stimulatory element for ribosomal pausing at the slippery site[43,44]. In +1 PRFs, ribosome pausing is also directed by the presence of rare or ‘hungry’ codons at the slippery site, which shifts the ribosomal A site onto a more abundant codon to resume elongation.

Much of our early understanding of −1 PRFs came from studies of the Rous sarcoma virus[7] and HIV-1[45], in both of which the structural protein precursor (Gag) and the enzyme precursor (Pol) are translated from the same viral mRNA. Gag is produced through conventional translation. A −1 PRF midway through Gag synthesis occurs in 2–10% of translating ribosomes and results in a fusion protein that is known as Gag–Pol, which is later cleaved by viral proteases to generate full-length Pol[5,46,47]. PRFs also have an important role in members of the *Coronaviridae* (for example, severe acute respiratory syndrome coronavirus (SARS-CoV), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and Middle Eastern respiratory syndrome coronavirus) and *Flaviviridae* (for example, West Nile virus)[48,49]. In the *Coronaviridae*, the replicase gene is organized into two partially overlapping open reading frames (ORFs) known as ORF1a and ORF1ab that encode polyprotein 1a and the fused polyprotein 1a–1b, respectively, the latter of which is generated by a −1 PRF. This frameshift event occurs at a frequency of 14–27%[50], and has been suggested as a mechanism that maintains the ratio of ORF1a to ORF1ab[51]. Unlike members of the *Retroviridae*, SARS-CoV contains an atypical three-stem pseudoknot and an additional, structurally conserved attenuator sequence that is 5′ of the PRF signal[50–52], which has been shown to control the frequency of −1 PRFs in coronaviruses[51,52]. Notably, lowering the efficiency of frameshifts markedly reduces viral replication and infectivity[6,51,53–55], which underscores the importance of the −1 PRF for these viruses. Importantly, host factors have been identified that interfere with virus PRFs. For instance, the human protein C19Orf66—first identified for its inhibitory effect on the replication of dengue virus[56]—has been shown to inhibit −1 PRFs in Gag–Pol synthesis[57]. C19Orf66 has further been shown to exhibit broad-spectrum activity in blocking PRFs in HIV-2, Rous sarcoma virus, human T lymphotropic virus and mouse mammary tumour virus[57]. Whether C19Orf66 functions only by limiting PRFs requires investigation, but targeting PRF factors could provide a selective and powerful antiviral strategy.

Leaky scanning. In ribosomal leaky scanning, the ribosome skips a translation initiation site (especially if this site is located in the context of a weak Kozak sequence) and initiates at a downstream one (Fig. 2c). Many viruses—including retroviruses[58], paramyxoviruses[59], papillomaviruses[60] and bunyaviruses—adopt leaky scanning to express

several proteins from one transcript[61,62]. In pandemic strains of HIV, a bicistronic mRNA transcript encodes a conserved upstream, small 81-amino-acid protein known as Vpu, which confers a fitness advantage by degrading the CD4 viral receptor and enhancing virion release[58,63–66]. The bypassing of the Vpu start codon leads to initiation on a downstream start codon, which results in the synthesis of the viral envelope protein[58]. In the segmented RNA genome of IAV, leaky scanning can generate four proteins in addition to the canonical protein that is encoded by segment 2[67]. For example, a downstream AUG leads to the synthesis of PB1-F2, a protein that localizes to mitochondria and elicits a pro-inflammatory and pro-apoptotic effect on host cells[11,68–70].

Translation of upstream ORFs. Although viruses have a relatively short 5′ untranslated region, an increasing body of evidence suggests that upstream ORFs that are led by upstream start codons (AUGs) can be translated (Fig. 2d). Upstream translation has widely been observed in DNA viruses and positive- and negative-sense RNA viruses, as well as in mammalian genomes[71–81]. Upstream ORFs in viruses have been suggested to have two major functional consequences. First, and similar to mammalian upstream ORFs[78–82], many viral upstream ORFs suppress the translation of the downstream canonical ORF. For instance, in ebolavirus, an upstream ORF of the *L* gene (which is important for replication and RNA capping) suppresses the translation of the *L* ORF under normal conditions and enhances it under stress conditions[75]. This bimodal regulation fine tunes the synthesis of L protein and helps to maintain optimal polymerase activity[75]. Similarly, upstream ORFs can regulate the expression of viral proteins in coronaviruses (such as murine hepatitis virus and bovine coronaviruses) and in several DNA viruses (such as hepatitis B virus and human cytomegalovirus)[72,74–77]. Second, the products of upstream ORFs can be involved in regulating virulence and tropism. In the monopartite genome of enteroviruses, a highly conserved upstream ORF partially overprints the canonical polyprotein ORF[71] and encodes a putative transmembrane protein that facilitates viral release and invasion of echovirus 7 in human gut epithelial cells[71].

Initiation of translation from non-AUG codons. The translation of many virus genes has been shown to initiate on noncanonical start codons that are typically found upstream of the canonical AUG codon[81] (Fig. 2d). These noncanonical start codons fall mainly into two categories. First, a near-cognate start codon that normally varies by one nucleotide from AUG can be recognized by the initiator $tRNA_i^{Met}$, which occurs at the P-site of the ribosome. For instance, the polycistronic P/C mRNA of Sendai virus and parainfluenza virus type 1 encodes five proteins (P, C, C′, Y1 and Y2) from overlapping ORFs. The C′ protein is generated by the efficient initiation of translation from an upstream non-AUG codon (ACG for Sendai virus and GUG for parainfluenza virus type 1), which has a N′ extension compared to the C protein[83,84]. Similar uses of non-AUG start codons (most frequently CUG, and sometimes GUG) have been identified in viruses that infect a wide range of hosts, including murine leukaemia virus[85], human T cell lymphotropic virus type 1[86], influenza virus[87], soil-borne wheat mosaic virus[88] and equine infectious anaemia virus[89]. Second, the non-AUG start codon can be recognized by a non-methionine tRNA. In this case, the initiator $tRNA_i^{Met}$ is not required and translation initiates in the A site. This leads to proteins that start with non-methionine amino acids, which have mainly been identified in insect viruses[90,91].

Start-snatching to generate hybrid proteins. Translation in eukaryotic cells requires the recognition of the 5′ methyl-7-guanosine (m^7G) cap on mRNA. Segmented negative-sense RNA viruses in the order *Bunyavirales* and the families *Orthomyxoviridae* (for example, IAV) and *Arenaviridae* (for example, Lassa virus) do not encode capping enzymes, but instead rely on a process known as 'cap-snatching' to access cap-dependent translation. In this process, viral polymerase binds to the m^7G cap of host RNA and cleaves off a short stretch (7–20 nucleotides in the case of IAV and about 7 nucleotides for Lassa virus) of host capped-RNA[92,93]. These host-derived fragments are then used as a primer to initiate the transcription of viral mRNAs[94]. As a consequence, mRNAs of segmented negative-sense RNA viruses exist as genetic hybrids, in which 5′ sequence heterogeneity is provided by snatched host-derived sequences[92,95–97].

Instead of merely providing a m^7G cap, cap-snatched host sequences that bear AUGs also allow segmented negative-sense RNA viruses to express cryptic ORFs within their 5′ untranslated regions (known as upstream viral ORFs). This process has been termed 'start-snatching' (Fig. 2d). During IAV infection, about 12% of host-derived cap-snatched sequences bear AUG start codons that confer translation. Depending on the reading frame of the host-derived AUG with respect to the viral RNA, these codons initiate the synthesis of either host–virus chimeric N-terminally extended viral proteins or novel polypeptides (up to 80 amino acids in length) that are overprinted with the major viral ORF[98]. Start-snatching and the genesis of upstream viral ORFs may be a way for segmented negative-sense RNA viruses to sample evolutionary space before gene functionalization. A recent study has shown that some strains of IAV have evolved to encode an AUG start codon in the untranslated region of the nucleoprotein segment. Expression of this N-terminally extended nucleoprotein increases viral virulence[99].

Additional mechanisms

Genome compaction in viruses has driven additional mechanisms that do not rely on genic overprinting to express several proteins from a single locus, which have previously been reviewed[81] and are summarized in Box 1.

Lessons for the development of therapeutic agents

A fundamental principle that underlies the development of antiviral drugs is to evaluate the benefit (for example, infection suppression) versus the cost (for example, off-target effects or toxicity on the host) provided by a drug (Fig. 3a). Two general strategies are currently used to combat microbial infections: training the host by vaccination and using small-molecule inhibitors to target the virus or the host. Here we provide perspectives on how common features of noncanonical viral gene expression could serve as a starting point for the development of antiviral therapies.

ARFs as vaccination targets

A goal of vaccination is to generate broadly protective antibodies and/or cross-reactive T cells that are directed against viral targets. However, the design of effective and universal vaccines is often hampered by rapid changes of viral antigens through mutation, recombination or re-assortment. For instance, antigenic drift and shift in the surface glycoproteins of IAV have hampered the development of a universal vaccine against influenza virus[100]. Thus, a major challenge remains to find ideal vaccination targets that are both highly immunogenic and genetically constrained from mutation owing to potential fitness loss.

ARFs have long been neglected as potential candidates for vaccine or drug development, and might provide a solution to this conundrum. ARFs (such as overprinted ORFs) feature an overall low synonymous divergence[101–103], and are therefore expected to be relatively constrained from accumulating mutations (as mutations in these regions are likely to disrupt more than one viral protein). Importantly, proteins encoded by ARFs have been shown to be abundantly synthesized during infections[104–107] and can be efficiently processed through class-I MHC processing pathways and induce cytotoxic T lymphocyte responses[108–110].

The use of ARF as epitopes has been proposed for HIV[108,111–113], influenza virus[110] and in some cancers[109] and has several major advantages. First, ARFs in simian immunodeficiency virus and HIV contribute greatly to $CD8^+$ T cell responses in infected individuals and trigger a stronger

Box 1

Nonoverlapping gene expression in viruses

Several proteins can be generated without overprinting from one coding sequence through an array of strategies that are used pervasively by viruses.

IRESs

IRESs (typically involving stem-loop and pseudoknot structures located upstream of a coding sequence) are widely adopted by viruses to circumvent cap dependency during translation. First identified in poliovirus and encephalomyocarditis viruses[156,157], IRESs have frequently been discovered in viruses with uncapped positive-strand RNA genomes (mostly picornaviruses), a few DNA viruses[158,159] and in mammalian genomes[160]. Although IRESs are functionally similar, there is no consensus sequence or structure for them. Mechanistically, IRESs can be divided into two types: a common type that recruits ribosome via binding to eIFs and other RNA-binding proteins[161,162], and a simpler type that directly recruits ribosome without eIFs (which has mainly been found in cricket paralysis virus)[163,164].

Ribosomal shunting

In ribosomal shunting (an alternative mechanism for the cap-dependent initiation of translation initiation), the 40S ribosome subunit bypasses the scanning of some segments of RNA by translocating to a downstream shunt acceptor site. This is typically enabled by stable hairpin structures that are formed by 5′-RNA leader sequences that block scanning. Shunting was first discovered in cauliflower mosaic virus[165], and has been found in many plant pararetroviruses[166] and in animal viruses that include adenoviruses[167] and Sendai virus[168].

Translation reinitiation

In translation reinitiation, post-termination ribosomes remain on a polycistronic viral mRNA and reinitiate translation from a nearby start codon either upstream or downstream of the termination codon of the preceding ORF. Translation reinitiation was first discovered in the polycistronic subgenomic RNA of rabbit haemorrhagic disease virus, in which two ORFs overlap by 17 nucleotides and encode the major and minor capsid proteins and several nonstructural proteins[169]. Translation reinitiation also occurs in caliciviruses and negative-sense RNA viruses such as influenza B virus[170], and human respiratory syncytial virus[171,172].

Read-through translation

In read-through translation, the ribosome reads a stop codon as a sense codon influenced by the stop codon context, which results in continued translation and protein products with extended C termini. Read-through translation has been observed on all three stop codons (most commonly on UGA) and can be mediated by either normal tRNAs or suppressor tRNAs. Read through is used extensively in many viruses, ranging from alphaviruses to mimivirus[173]. Although read through is relatively rare in mammalian genes (as has previously been reviewed[174]), it can result in insertions of selenocysteine[175,176].

Stop–carry on

In stop–carry on, the ribosome skips the formation of a peptidyl bond while reading two consecutive sense codons, which generates two peptides from one ORF in a stop-codon-independent manner. Stop–carry on was first identified in the peptidase 2A and 2B regions of the polyprotein of foot and mouth disease virus[177,178] (as has previously been reviewed[179]). The 2A peptide structurally hinders the binding of the last tRNA (tRNA-Pro) to the ribosomal A site but not to the release factors[180], thus causing the skipping of peptidyl bond formation. Stop–carry on occurs mainly on 2A and 2A-like peptides, which are widely present and conserved in picornaviruses and other mammalian and insect viruses[179].

cytotoxic T lymphocyte response compared to epitopes that target the canonical proteins[108,114]. The potential of ARFs as epitopes is further substantiated by the observation that codon-optimized recombinant HIV vaccines (in which ARFs are disrupted or skewed) trigger a reduced cytotoxic T lymphocyte response compared to non-codon optimized vaccines[112]. Second, cytotoxic T lymphocyte responses to at least some ARF epitopes do not drive viral escape[113] and presentation of ARF epitopes has been associated with favourable clinical outcomes[111]. Finally, overprinting ORFs tend to be highly conserved among strains of the same virus, as in IAV[98]. Taken together, these findings suggest that ARFs and overprinting ORFs present potential antigen candidates for the development of new vaccines and for therapies based on chimeric antigen receptor T cells[115].

Targeting viral nucleic acid structures

Many viruses rely on the presence of *cis*-acting structural elements in their genomes for protein expression. These elements tend to be highly conserved, and have both structural and sequence-specific properties; they therefore present excellent targets for drug development (Fig. 3b). These strategies require precise knowledge of the sequence and structure of the nucleic acid target region, as well as its viral and host binding partners.

Structure-targeting drugs can be designed following two strategies. First, a drug can disrupt or alter the structure of a *cis* element. For example, a compound (known as ligand 43) discovered from an in silico small-molecule screen has been shown to specifically inhibit −1 PRFs in SARS-CoV by altering the plasticity of a viral RNA pseudoknot[116–118]. Second, a drug can inhibit cofactor binding to a structural element. For example, benzimidazole (a potential inhibitor of hepatitis C virus (HCV)[119,120]) functions by widening the interhelical angle in the viral internal ribosomal entry site (IRES), which results in reduced interaction with ribosome subunits and thus the inhibition of translation[121,122].

In theory, the high conservation at structure and sequence levels makes viral *cis* elements ideal targets for antisense oligonucleotides, which work by disrupting structure formation or induce degradation of the RNA by recruitment of RNase H. Indeed, the first drug approved by the US Food and Drug Administration (fomivirsen) for treating cytomegalovirus retinitis in individuals infected with HIV is an antisense drug. Several other antisense-based antiviral drugs against HIV, HCV, ebolavirus and Marburg virus have entered clinical trials. However, antisense oligonucleotide technology has some caveats. Besides considerations of delivery method (which have previously been reviewed[123]), virus escape can occur. For example, an antisense oligonucleotide inhibitor (ISI-14803) of HCV that targets the IRES has been shown to exert selective pressure on the IRES sequence[124,125]. This resulted in mutations accumulating in the virus in patients during a phase-I clinical trial, although no mutations were detected at the antisense oligonucleotide binding site[124]. Taken together, these data suggest that the design of drugs based on antisense oligonucleotides requires a careful analysis of the surrounding structures. Alternatively, it may be necessary to use multiplex delivery of antisense oligonucleotides (that is, to target several regions of the structure at the same time), such that compensatory escape mutations will be unable to take hold.

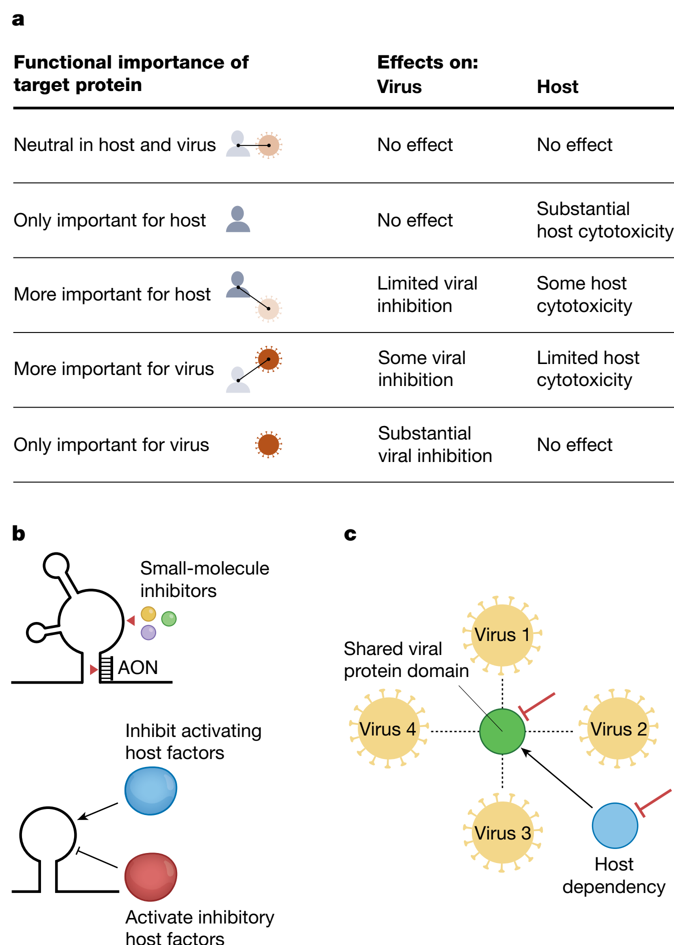

Fig. 3 | Strategy for therapeutic and prophylactic development of novel antiviral agents. a, A balance between viral inhibition and host toxicity underlies therapeutic development. Targeting viral-specific functions or host functions that are more important (in a given time frame) to the virus than the host paves the way for the generation of therapeutic agents. **b**, *Cis*-acting nucleic-acid structural elements that are involved in unconventional viral expression mechanisms (such as pseudoknots in PRFs, and stem loops in polymerase slippage sites and IRES) can be directly targeted by small molecules, host factors and antisense oligonucleotides (AON) or indirectly targeted by modulating the related host factors. **c**, Targeting of virus-specific processes in gene expression (such as cap-snatching and RdRp) that are shared among viruses and not found in hosts offers a high specificity for antiviral agents. The targeting of host dependencies that are used by several virus provides an alternative route to pan-viral therapeutic agents.

Targeting virus-specific mechanisms of gene expression

Many viruses rely on their own proxies of host enzymes (for example, the capping machinery of the *Coronaviridae*) or pathways (for example, the cap-snatching of the *Orthomyxoviridae*) to express viral proteins (Box 2). Inhibitory drugs against these virus-specific proteins and pathways should achieve high specificity for the virus with minimal effect on the host (Fig. 3a).

Cap-snatching, which is used only by influenza viruses and other segmented negative-sense viruses, presents one such targetable pathway. To date, at least three small-molecule antiviral agents (favipiravir, pimodivir and baloxavir) that target the PB1, PB2 and PA subunits, respectively, of the influenza viral polymerase trimer have entered clinical development (as has previously been reviewed[126]). Baloxavir has been approved for treating influenza virus infections in the USA and Japan, and was generated through rational design against the cap-dependent endonuclease active site of the IAV PA protein[127]. Baloxavir has been shown to effectively inhibit cap-snatching activities in both IAV and influenza B virus[127], and has broader antiviral effects than current standard-of-care anti-influenza drugs[128,129]. Success with these drugs may pave the way for the development of antiviral agents against other highly pathogenic cap-snatching viruses.

Conserved protein domains across viral families might provide targets for broader-acting antiviral agents (Fig. 3c). For example, RdRp is essential to RNA viruses and shares a similar 3D structural conformation[130] and mechanism of action across species, which suggests that drugs that target RdRp could have activities in different viral families. Favipiravir—which was initially discovered on the basis of its antiviral activity against IAV—has been shown to exhibit antiviral activity against other RNA viruses, including viruses that cause fatal haemorrhagic fevers (arenaviruses, peribunyaviruses and filoviruses)[131].

Although viral-targeting drugs offer high specificity, a potential issue is the acquisition of drug-resistant mutations in the viral targets. In the case of baloxavir, IAV recovered from 1.1 to 19.5% of patients treated with the drug developed up to 138 compensatory mutations[132]. A possible solution is combination therapy: because the targets of combination therapy are often located in different pathways or proteins, it is more difficult for the viral to acquire resistance compared to monotherapies. Indeed, combination therapies have been shown to slow down the acquisition of resistance and yield effective viral clearance[133], as exemplified by the combinatorial 'highly active antiretroviral therapy' (HAART) used in controlling HIV infections[134], as well as similar strategies using in the treatments of cancers[135] and multidrug-resistant bacterial infections (as has previously been reviewed[136]).

Unfortunately most drugs—whether developed by academic or commercial institutions—are developed as single agents, and face a range of legal and regulatory issues that might hamper their use in the testing of combination therapies. Thus, a shift in drug-development paradigms towards a more collaborative environment among research bodies and clinicians is imperative for the future development of combinatorial strategies.

Host dependencies as targets of pan-viral therapies

Although the high mutation rates of viruses suggest an unlimited evolutionary potential, a virus that is fully co-adapted to its host will have very few neutral sites in its genome[137]—which locks the virus into evolutionary stasis and limits marked divergence over the long term. In support of this, an analysis of HBV genomes recovered from prehistoric periods has shown that these viruses were only 1.3–3% divergent from modern circulating strains[138,139]. This suggests that a viable strategy for antiviral development can be achieved by targeting host dependencies, which can result from indirect or direct interactions between a virus and its host (Fig. 3c).

When considering the inhibition of a host dependency a trade-off exists between viral inhibition and the potential disruption of host cellular functions. A parallel can be observed with cancer therapeutic agents: cancer cells that are heavily reliant on essential host functions can be killed by short-term or partial inhibition against these functions (for example, topoisomerase or proteasome inhibitors), while maintaining minimal long-term damage to the patient. The ideal therapeutic targets for viral infections would be host factors upon which viruses heavily depend, and the short-term or partial inhibition of which over the course of an infection is well-tolerated by the host. Furthermore, if commonalities in host dependencies exist among different viruses, targeting these dependencies might allow the development of broad-spectrum or pan-viral therapeutic agents. This could contribute to combating newly emerging infections that lack efficient antiviral therapies (for example, as in the current COVID-19 pandemic).

Direct dependencies. Viral proteins or RNA may directly interact with host factors to give rise to direct dependencies. The identification of direct host dependencies requires knowledge of host–viral

Box 2

Viral proxies of host molecules

Many host processes are encoded by large multimodal complexes that are confined to specific subcellular compartments. Viruses use proxies of host machinery and processes that can accordingly be considered as component-level and process-level proxies.

Component-level proxies

Viruses may encode their own simplified versions of host proteins. For instance, cellular RNA capping occurs co-transcriptionally in the nucleus by a series of capping enzymes through removal of a monophosphate from 5′ mRNA and the subsequent transferring and methylation of a GMP. The m^7G cap is a critical modification for viral mRNA, as it protects viral mRNA from degradation and allows translation. Although some viruses can use host capping enzymes, other have evolved to express viral substitutes to these host enzymes. In most cases, this entails simplified and/or non-canonical proteins (as compared to their host counterparts). The *Mononegavirales* rely on an unconventional multifunctional enzyme known as L protein that serves as both RdRp and capping enzyme. In this process, a covalent mRNA–enzyme intermediate is first formed between the 5′ monophosphorylated mRNA and the L protein[181]. This mRNA–enzyme intermediate is then transferred to a GDP receptor. Subsequently, two methylation modifications sequentially occur on the ribose-2′-*O* position of the first nucleotide and on the guanine *N*-7 position of the cap[181]. Another unconventional capping mechanism is adopted by *Alphaviridae* (as has previously been reviewed[182]), in which pre-methylated GTP is donated to 5′ diphosphorylated mRNA, forming the cap[183]. Other viruses have innovated structural mimics of the 5′ cap. For example, picornaviruses encode the protein Vpg, which covalently links to the 5′ of the genome and mimics m^7G caps. The full spectrum of this strategy has previously been reviewed[184].

Process-level proxies

Viruses may evolve completely unique strategies that are analogous to host processes. For example, RNA splicing typically occurs co-transcriptionally in the nucleus of the host. Viruses (such as coronaviruses and other viruses in the order *Nidovirales*) that replicate in the cytoplasm therefore do not have ready access to the splicing machinery of the host. Instead, these positive-strand viruses use a mechanism of discontinuous transcription to generate a nested set of minus-strand subgenomic mRNAs. These subgenomic mRNAs all share a leader sequence derived from the 5′ end of the viral genome and serve as templates for mRNA production. Discontinuous transcription is regulated by transcription regulating sequences. Transcription regulating sequences mediate long-range RNA–RNA interactions that promote viral polymerase template-switching during transcription (as previously reviewed[185]). Notably, a recent analysis of SARC-CoV-2 indicates that many unidentified ORFs can be generated by discontinuous transcription[186,187]. Discontinuous transcription can thus be considered as different from alternative splicing mechanistically, but analogous to splicing in terms of the end result (that is, the generation of several transcripts from one gene or genomic region). Similarly, polyprotein processing is commonly adopted by RNA viruses and retroviruses in which a polyprotein is expressed from a single RNA species and subsequently cleaved by viral or host proteases into functional proteins. These unique viral enzymes or pathways, and the host factors that might regulate these processes, represent potential targets in strategies for viral eradication.

protein–protein and protein–nucleic acid interactions that are shared and important among different viral families. The inhibition of these proteins or processes is therefore likely to have broad-spectrum antiviral effects.

Several viral species require a common set of host factors (collectively known as the IRES *trans*-acting factors) for viral IRES translation. The inhibition of these factors therefore blocks replication of viruses from several unrelated families. For example, the inhibition of the host ribosome-binding protein receptor for activated C kinase 1 (which is co-opted by many viruses in IRES-mediated translation[140]) effectively inhibited HCV and herpes simplex virus infection with no significant effect on the viability or proliferation of the human host cells[140,141].

Another host dependency is protein localization to the endoplasmic reticulum, which is shared by several evolutionarily distant viruses such IAV, HIV and dengue virus[142]. As predicted, treatment with small-molecule inhibitors of SEC61 (a protein complex that mediates co-translational translocation in endoplasmic reticulum and endoplasmic reticulum–Golgi intermediate compartments) showed suppression of replication of all three of these viruses in vitro[142]. Different iterations of SEC61 inhibitors have been shown to effectively suppress Zika virus and coronavirus replication in vitro[143,144]. Further work is needed to evaluate their activity in vivo, but the underlying general concept is that viruses have a strong requirement—in a small temporal window of active infection—for oxidative folding and modification associated with apical trafficking[142–144]. Along similar lines, host glycosylation enzymes (which are extensively used for viral surface protein modification) have inspired the development of vaccines and therapeutic agents—for example, the use of glycans as vaccine adjuvants for HIV[145,146] and antiviral drugs (zanamivir and oseltamivir) for IAV.

Indirect dependencies. An indirect host dependency arises from indirect functional interactions between the virus and a host protein or process. One example of such a dependency is the importance of the host splicing machinery for viruses that replicate in cytosol. For instance, infections with SARS-CoV-2 have been shown to cause a marked increase in spliceosome components in host cells[147]. Viruses can disrupt host splicing function by triggering nucleo-cytoplasmic translocation and the sequestering of spliceosome components (in the case of rotavirus[148,149], which has previously been reviewed[150]) or by inducing changes in splicing patterns of host cellular genes (in the case of influenza virus[149], Zika virus[151], human cytomegalovirus[152], and in hepatitis B virus- and HCV-related hepatocellular carcinoma[153]).

The therapeutic targeting of alternative splicing by small molecules or protein inhibitors and antisense oligonucleotides has been proposed in the treatment of cancer, on the basis of the observation of pro-oncogenic isoforms generated by defective alternative splicing (as previously reviewed[154,155]). Altering the splice pattern of a receptor for viral entry using antisense oligonucleotides could generate a decoy receptor and prevent infection. Overall, the pervasive involvement of host splicing machinery in viral gene expression suggests that modulation of splicing might serve as a promising antiviral therapeutic strategy.

Conclusion

Viruses use a diverse array of noncanonical transcriptional and translational strategies to greatly expand the coding potential of, and add novel functionality to, their small genomes. However, to do so they have relied on unique enzymatic activities or become dependent on host

functions. Viral enzymes that have no homology with human enzymes represent ideal targets for the development of virus-specific inhibitors. Host dependencies are also valuable targets as—in many cases—these dependencies exist broadly across different viruses. We surmise that future developments in our biochemical and detailed mechanistic understanding of how viruses make proteins will inform the development of therapeutic agents and vaccines.

1. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
2. Chung, W. Y., Wadhawan, S., Szklarczyk, R., Pond, S. K. & Nekrutenko, A. A first look at ARFome: dual-coding genes in mammalian genomes. *PLoS Comput. Biol*. **3**, e91 (2007).
3. Michel, A. M. et al. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res*. **22**, 2219–2229 (2012).
4. Schlub, T. E. & Holmes, E. C. Properties and abundance of overlapping genes in viruses. *Virus Evol*. **6**, veaa009 (2020).
5. Biswas, P., Jiang, X., Pacchia, A. L., Dougherty, J. P. & Peltz, S. W. The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J. Virol*. **78**, 2082–2087 (2004).
6. Shehu-Xhilaga, M., Crowe, S. M. & Mak, J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol*. **75**, 1834–1841 (2001).
7. Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. Signals for ribosomal frameshifting in the Rous sarcoma virus Gag-Pol region. *Cell* **55**, 447–458 (1988).
This article reports the discovery of the frameshift site and stem-loop structure as requirements for PRFs in the synthesis of Gag–Pol protein.
8. van Knippenberg, I., Carlton-Smith, C. & Elliott, R. M. The N-terminus of Bunyamwera orthobunyavirus NSs protein is essential for interferon antagonism. *J. Gen. Virol*. **91**, 2002–2006 (2010).
9. McFadden, N. et al. Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4. *PLoS Pathog*. **7**, e1002413 (2011).
10. Scholthof, H. B. The Tombusvirus-encoded P19: from irrelevance to elegance. *Nat. Rev. Microbiol*. **4**, 405–411 (2006).
11. Chen, W. et al. A novel influenza A virus mitochondrial protein that induces cell death. *Nat. Med*. **7**, 1306–1312 (2001).
12. Rancurel, C., Khosravi, M., Dunker, A. K., Romero, P. R. & Karlin, D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J. Virol*. **83**, 10719–10736 (2009).
This paper was one of the first to suggest that overprinting is a mechanism for de novo gene and protein creation, and uses viruses as a model system to characterize the properties and the phylogenetic distributions of such proteins.
13. Neme, R. & Tautz, D. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics* **14**, 117 (2013).
14. Penno, C., Kumari, R., Baranov, P. V., van Sinderen, D. & Atkins, J. F. Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: potential implications for modulation of GagPol synthesis. *Nucleic Acids Res*. **45**, 10156–10167 (2017).
15. Hausmann, S., Garcin, D., Delenda, C. & Kolakofsky, D. The versatility of paramyxovirus RNA polymerase stuttering. *J. Virol*. **73**, 5568–5576 (1999).
16. Barr, J. N. & Wertz, G. W. Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3′-AUAC-5′ tetranucleotide: implications for the mechanism of transcription termination. *J. Virol*. **75**, 6901–6913 (2001).
17. Thomas, S. M., Lamb, R. A. & Paterson, R. G. Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* **54**, 891–902 (1988).
A seminal paper that first identified transcriptional slippage as a mechanism of generating multiple transcripts from one gene in paramyxovirus SV5.
18. Olspert, A., Chung, B. Y., Atkins, J. F., Carr, J. P. & Firth, A. E. Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO Rep*. **16**, 995–1004 (2015).
19. Ratinier, M. et al. Transcriptional slippage prompts recoding in alternate reading frames in the hepatitis C virus (HCV) core sequence from strain HCV-1. *J. Gen. Virol*. **89**, 1569–1578 (2008).
20. Kolakofsky, D., Roux, L., Garcin, D. & Ruigrok, R. W. H. Paramyxovirus mRNA editing, the "rule of six" and error catastrophe: a hypothesis. *J. Gen. Virol*. **86**, 1869–1877 (2005).
21. Dillon, P. J. & Parks, G. D. Role for the phosphoprotein P subunit of the paramyxovirus polymerase in limiting induction of host cell antiviral responses. *J. Virol*. **81**, 11116–11127 (2007).
22. Gainey, M. D., Dillon, P. J., Clark, K. M., Manuse, M. J. & Parks, G. D. Paramyxovirus-induced shutoff of host and viral protein synthesis: role of the P and V proteins in limiting PKR activation. *J. Virol*. **82**, 828–839 (2008).
23. Didcock, L., Young, D. F., Goodbourn, S. & Randall, R. E. The V protein of simian virus 5 inhibits interferon signalling by targeting STAT1 for proteasome-mediated degradation. *J. Virol*. **73**, 9928–9933 (1999).
24. Kato, A., Kiyotani, K., Sakai, Y., Yoshida, T. & Nagai, Y. The paramyxovirus, Sendai virus, V protein encodes a luxury function required for viral pathogenesis. *EMBO J*. **16**, 578–587 (1997).
25. Vulliémoz, D. & Roux, L. "Rule of six": how does the Sendai virus RNA polymerase keep count? *J. Virol*. **75**, 4506–4518 (2001).
26. Iseni, F. et al. Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA* **8**, 1056–1067 (2002).
27. Shabman, R. S. et al. Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. *MBio* **5**, e02011-14 (2014).
28. Mehedi, M. et al. A new Ebola virus nonstructural glycoprotein expressed through RNA editing. *J. Virol*. **85**, 5406–5414 (2011).
29. Sanchez, A., Trappier, S. G., Mahy, B. W., Peters, C. J. & Nichol, S. T. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl Acad. Sci. USA* **93**, 3602–3607 (1996).
30. Sanchez, A. et al. Biochemical analysis of the secreted and virion glycoproteins of Ebola virus. *J. Virol*. **72**, 6442–6447 (1998).
31. Volchkov, V. E. et al. GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **214**, 421–430 (1995).
32. Selman, M., Dankar, S. K., Forbes, N. E., Jia, J. J. & Brown, E. G. Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerg. Microbes Infect*. **1**, e42 (2012).
33. Neumann, G., Hughes, M. T. & Kawaoka, Y. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J*. **19**, 6751–6758 (2000).
34. Tsai, P. L. et al. Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. *PLoS Pathog*. **9**, e1003460 (2013).
35. Shih, S. R. & Krug, R. M. Novel exploitation of a nuclear function by influenza virus: the cellular SF2/ASF splicing factor controls the amount of the essential viral M2 ion channel protein in infected cells. *EMBO J*. **15**, 5415–5427 (1996).
36. Artarini, A. et al. Regulation of influenza A virus mRNA splicing by CLK1. *Antiviral Res*. **168**, 187–196 (2019).
37. Bogdanow, B. et al. The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. *Nat. Commun*. **10**, 5518 (2019).
38. Jiang, T., Nogales, A., Baker, S. F., Martinez-Sobrido, L. & Turner, D. H. Mutations designed by ensemble defect to misfold conserved RNA structures of influenza A segments 7 and 8 affect splicing and attenuate viral replication in cell culture. *PLoS ONE* **11**, e0156906 (2016).
39. Cai, Z. et al. VirusCircBase: a database of virus circular RNAs. *Brief Bioinform*. **22**, 2182–2190 (2021).
40. Zhao, J. et al. Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus. *Nat. Commun*. **10**, 2300 (2019).
This paper reports how oncogenic strains of human papillomavirus use back-splicing to generate circular RNAs of oncogene *E7*, which have an essential role in the malignant transformation of infected cells.
41. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. & Baranov, P. V. Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res*. **44**, 7007–7078 (2016).
42. Loughran, G., Firth, A. E. & Atkins, J. F. Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. *Proc. Natl Acad. Sci. USA* **108**, E1111–E1119 (2011).
43. Harger, J. W., Meskauskas, A. & Dinman, J. D. An "integrated model" of programmed ribosomal frameshifting. *Trends Biochem. Sci*. **27**, 448–454 (2002).
44. Dinman, J. D. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip. Rev. RNA* **3**, 661–673 (2012).
45. Jacks, T. et al. Characterization of ribosomal frameshifting in HIV-1 Gag-Pol expression. *Nature* **331**, 280–283 (1988).
46. Dulude, D., Baril, M. & Brakier-Gingras, L. Characterization of the frameshift stimulatory signal controlling a programmed −1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res*. **30**, 5094–5102 (2002).
47. Cassan, M., Delaunay, N., Vaquero, C. & Rousset, J. P. Translational frameshifting at the Gag-Pol junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells. *J. Virol*. **68**, 1501–1508 (1994).
48. Brierley, I., Meredith, M. R., Bloys, A. J. & Hagervall, T. G. Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *J. Mol. Biol*. **270**, 360–373 (1997).
49. Firth, A. E., Blitvich, B. J., Wills, N. M., Miller, C. L. & Atkins, J. F. Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. *Virology* **399**, 153–166 (2010).
50. Baranov, P. V. et al. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* **332**, 498–510 (2005).
51. Plant, E. P., Rakauskaite, R., Taylor, D. R. & Dinman, J. D. Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J. Virol*. **84**, 4330–4340 (2010).
52. Su, M. C., Chang, C. T., Chu, C. H., Tsai, C. H. & Chang, K. Y. An atypical RNA pseudoknot stimulator and an upstream attenuation signal for −1 ribosomal frameshifting of SARS coronavirus. *Nucleic Acids Res*. **33**, 4265–4275 (2005).
53. Dulude, D., Berchiche, Y. A., Gendron, K., Brakier-Gingras, L. & Heveker, N. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology* **345**, 127–136 (2006).
54. Garcia-Miranda, P. et al. Stability of HIV frameshift site RNA correlates with frameshift efficiency and decreased virus infectivity. *J. Virol*. **90**, 6906–6917 (2016).
55. Karacostas, V., Wolffe, E. J., Nagashima, K., Gonda, M. A. & Moss, B. Overexpression of the HIV-1 Gag-Pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* **193**, 661–671 (1993).
56. Suzuki, Y. et al. Characterization of RyDEN (C19orf66) as an interferon-stimulated cellular inhibitor against dengue virus replication. *PLoS Pathog*. **12**, e1005357 (2016).
57. Wang, X. et al. Regulation of HIV-1 Gag-Pol expression by shiftless, an inhibitor of programmed −1 ribosomal frameshifting. *Cell* **176**, 625–635.e14 (2019).
This article describes how the host factor shiftless functions as a universal inhibitor of −1 PRFs by causing premature translational termination in HIV-1 and several other viruses.
58. Schwartz, S., Felber, B. K., Fenyö, E. M. & Pavlakis, G. N. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J. Virol*. **64**, 5448–5456 (1990).
The authors report the discovery of leaky scanning of an upstream viral protein as the mechanism of generating the envelope protein in HIV-1.
59. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology* **318**, 463–473 (2004).

60. Stacey, S. N. et al. Leaky scanning is the predominant mechanism for translation of human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA. *J. Virol*. **74**, 7284–7297 (2000).
61. Fuller, F., Bhown, A. S. & Bishop, D. H. Bunyavirus nucleoprotein, N, and a non-structural protein, NSS, are coded by overlapping reading frames in the S RNA. *J. Gen. Virol*. **64**, 1705–1714 (1983).
62. Vera-Otarola, J. et al. The Andes hantavirus NSs protein is expressed from the viral small mRNA by a leaky scanning mechanism. *J. Virol*. **86**, 2176–2187 (2012).
63. Strebel, K., Klimkait, T. & Martin, M. A. A novel gene of HIV-1, *vpu*, and its 16-kilodalton product. *Science* **241**, 1221–1223 (1988).
64. Willey, R. L., Maldarelli, F., Martin, M. A. & Strebel, K. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *J. Virol*. **66**, 7193–7200 (1992).
65. Terwilliger, E. F., Cohen, E. A., Lu, Y. C., Sodroski, J. G. & Haseltine, W. A. Functional role of human immunodeficiency virus type 1 Vpu. *Proc. Natl Acad. Sci. USA* **86**, 5163–5167 (1989).
66. Sauter, D. et al. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe* **6**, 409–421 (2009).
67. Wise, H. M. et al. Overlapping signals for translational regulation and packaging of influenza A virus segment 2. *Nucleic Acids Res*. **39**, 7775–7790 (2011).
68. Zell, R. et al. Prevalence of PB1-F2 of influenza A viruses. *J. Gen. Virol*. **88**, 536–546 (2007).
69. McAuley, J. L. et al. PB1-F2 proteins from H5N1 and 20th century pandemic influenza viruses cause immunopathology. *PLoS Pathog*. **6**, e1001014 (2010).
70. Zamarin, D., García-Sastre, A., Xiao, X., Wang, R. & Palese, P. Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS Pathog*. **1**, e4 (2005).
71. Lulla, V. et al. An upstream protein-coding region in enteroviruses modulates virus infection in gut epithelial cells. *Nat. Microbiol*. **4**, 280–292 (2019).
72. Irigoyen, N. et al. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog*. **12**, e1005473 (2016).
73. Wu, H. Y., Guan, B. J., Su, Y. P., Fan, Y. H. & Brian, D. A. Reselection of a genomic upstream open reading frame in mouse hepatitis coronavirus 5′-untranslated-region mutants. *J. Virol*. **88**, 846–858 (2014).
74. Hofmann, M. A., Senanayake, S. D. & Brian, D. A. A translation-attenuating intraleader open reading frame is selected on coronavirus mRNAs during persistent infection. *Proc. Natl Acad. Sci. USA* **90**, 11733–11737 (1993).
75. Shabman, R. S. et al. An upstream open reading frame modulates ebola virus polymerase translation and virus replication. *PLoS Pathog*. **9**, e1003147 (2013).
This paper relates how an upstream ORF suppresses the translation of the downstream canonical L protein as a way to maintain protein expression level in ebolavirus.
76. Degnin, C. R., Schleiss, M. R., Cao, J. & Geballe, A. P. Translational inhibition mediated by a short upstream open reading frame in the human cytomegalovirus gpUL4 (gp48) transcript. *J. Virol*. **67**, 5514–5521 (1993).
77. Chen, A., Kao, Y. F. & Brown, C. M. Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons. *Nucleic Acids Res*. **33**, 1169–1181 (2005).
78. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl Acad. Sci. USA* **106**, 7507–7512 (2009).
79. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890 (2015).
80. Jin, X., Turcott, E., Englehardt, S., Mize, G. J. & Morris, D. R. The two upstream open reading frames of oncogene *mdm2* have different translational regulatory properties. *J. Biol. Chem*. **278**, 25716–25721 (2003).
81. Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *J. Gen. Virol*. **93**, 1385–1409 (2012).
82. Young, S. K. & Wek, R. C. Upstream open reading frames differentially regulate gene-specific translation in the integrated stress response. *J. Biol. Chem*. **291**, 16927–16935 (2016).
83. Gupta, K. C. & Patwardhan, S. ACG, the initiator codon for a Sendai virus protein. *J. Biol. Chem*. **263**, 8553–8556 (1988).
This paper represents one of the first observations of near-cognate start codon use in viral protein expression.
84. Boeck, R., Curran, J., Matsuoka, Y., Compans, R. & Kolakofsky, D. The parainfluenza virus type 1 P/C gene uses a very efficient GUG codon to start its C' protein. *J. Virol*. **66**, 1765–1768 (1992).
85. Prats, A. C., De Billy, G., Wang, P. & Darlix, J. L. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. *J. Mol. Biol*. **205**, 363–372 (1989).
86. Corcelette, S., Massé, T. & Madjar, J. J. Initiation of translation by non-AUG codons in human T-cell lymphotropic virus type I mRNA encoding both Rex and Tax regulatory proteins. *Nucleic Acids Res*. **28**, 1625–1634 (2000).
87. Machkovech, H. M., Bloom, J. D. & Subramaniam, A. R. Comprehensive profiling of translation initiation in influenza virus infected cells. *PLoS Pathog*. **15**, e1007518 (2019).
88. Shirako, Y. Non-AUG translation initiation in a plant RNA virus: a forty-amino-acid extension is added to the N terminus of the soil-borne wheat mosaic virus capsid protein. *J. Virol*. **72**, 1677–1682 (1998).
89. Carroll, R. & Derse, D. Translation of equine infectious anemia virus bicistronic *tat-rev* mRNA requires leaky ribosome scanning of the *tat* CTG initiation codon. *J. Virol*. **67**, 1433–1440 (1993).
90. Sasaki, J. & Nakashima, N. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc. Natl Acad. Sci. USA* **97**, 1512–1515 (2000).
This paper reports the discovery that translation initiation of an insect virus mRNA is independent of both AUG and the methionine initiator tRNA.
91. Wilson, J. E., Powell, M. J., Hoover, S. E. & Sarnow, P. Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol. Cell. Biol*. **20**, 4990–4999 (2000).
92. Rialdi, A. et al. The RNA exosome syncs IAV-RNAPII transcription to promote viral ribogenesis and infectivity. *Cell* **169**, 679–692.e14 (2017).
93. Dias, A. et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**, 914–918 (2009).
94. Reich, S. et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361–366 (2014).
95. Koppstein, D., Ashour, J. & Bartel, D. P. Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. *Nucleic Acids Res*. **43**, 5052–5064 (2015).
96. Gu, W. et al. Influenza A virus preferentially snatches noncoding RNA caps. *RNA* **21**, 2067–2075 (2015).
97. Sikora, D., Rocheleau, L., Brown, E. G. & Pelchat, M. Influenza A virus cap-snatches host RNAs based on their abundance early after infection. *Virology* **509**, 167–177 (2017).
98. Ho, J. S. Y. et al. Hybrid gene origination creates human-virus chimeric proteins during infection. *Cell* **181**, 1502–1517 (2020).
This article describes the discovery of upstream AUGs in cap-snatched host sequences being transcribed and translated to generate host–virus hybrid proteins.
99. Wise, H. M. et al. An alternative AUG codon that produces an N-terminally extended form of the influenza A virus NP is a virulence factor for a swine-derived virus. Preprint at https://doi.org/10.1101/738427 (2019).
100. Kim, H., Webster, R. G. & Webby, R. J. Influenza virus: dealing with a drifting and shifting pathogen. *Viral Immunol*. **31**, 174–183 (2018).
101. Firth, A. E. & Brown, C. M. Detecting overlapping coding sequences in virus genomes. *BMC Bioinformatics* **7**, 75 (2006).
102. Firth, A. E. & Brown, C. M. Detecting overlapping coding sequences with pairwise alignments. *Bioinformatics* **21**, 282–292 (2005).
103. Jagger, B. W. et al. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* **337**, 199–204 (2012).
104. Mohamadi, M. et al. Hepatitis C virus alternative reading frame protein (ARFP): production, features, and pathogenesis. *J. Med. Virol*. **92**, 2930–2937 (2020).
105. Zanker, D. J. et al. Influenza A virus infection induces viral and cellular defective ribosomal products encoded by alternative reading frames. *J. Immunol*. **202**, 3370–3380 (2019).
106. Depledge, D. P. et al. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat. Commun*. **10**, 754 (2019).
107. Di, H. et al. Expanded subgenomic mRNA transcriptome and coding capacity of a nidovirus. *Proc. Natl Acad. Sci. USA* **114**, E8895–E8904 (2017).
108. Bansal, A. et al. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J. Exp. Med*. **207**, 51–59 (2010).
This paper describes how an ARF in HIV encodes cryptic epitopes that contribute to the majority of the CD8 T cell response during infection.
109. Vetsika, E. K. et al. Sequential administration of the native $TERT_{572}$ cryptic peptide enhances the immune response initiated by its optimized variant $TERT_{572Y}$ in cancer patients. *J. Immunother*. **34**, 641–650 (2011).
110. Dolan, B. P., Li, L., Takeda, K., Bennink, J. R. & Yewdell, J. W. Defective ribosomal products are the major source of antigenic peptides endogenously generated from influenza A virus neuraminidase. *J. Immunol*. **184**, 1419–1424 (2010).
111. Bansal, A. et al. Enhanced recognition of HIV-1 cryptic epitopes restricted by HLA class I alleles associated with a favorable clinical outcome. *J. Acquir. Immune Defic. Syndr*. **70**, 1–8 (2015).
112. Bet, A., Sterrett, S., Sato, A., Bansal, A. & Goepfert, P. A. Characterization of T-cell responses to cryptic epitopes in recipients of a noncodon-optimized HIV-1 vaccine. *J. Acquir. Immune Defic. Syndr*. **65**, 142–150 (2014).
113. Peng, B. J. et al. Antisense-derived HIV-1 cryptic epitopes are not major drivers of viral evolution during the acute phase of infection. *J. Virol*. **92**, e00711-18 (2018).
114. Maness, N. J. et al. Robust, vaccine-induced $CD8^+$ T lymphocyte response against an out-of-frame epitope. *J. Immunol*. **184**, 67–72 (2010).
Four articles[111–114] demonstrate how the T cell response to cryptic epitopes produced from ARFs in HIV is robust, associated with improved clinical outcomes and associated with minimal viral escape.
115. Seif, M., Einsele, H. & Löffler, J. CAR T cells beyond cancer: hope for immunomodulatory therapy of infectious diseases. *Front. Immunol*. **10**, 2711 (2019).
116. Park, S. J., Kim, Y. G. & Park, H. J. Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J. Am. Chem. Soc*. **133**, 10094–10100 (2011).
This paper describes how an RNA structure-based inhibitor screening revealed a candidate that targets and reduces the efficiency of −1 PRFs in SARS-CoV.
117. Ritchie, D. B., Soong, J., Sikkema, W. K. & Woodside, M. T. Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. *J. Am. Chem. Soc*. **136**, 2196–2199 (2014).
118. Ritchie, D. B., Foster, D. A. & Woodside, M. T. Programmed −1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl Acad. Sci. USA* **109**, 16167–16172 (2012).
119. Seth, P. P. et al. SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J. Med. Chem*. **48**, 7099–7102 (2005).
120. Parsons, J. et al. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol*. **5**, 823–825 (2009).
121. Dibrov, S. M. et al. Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proc. Natl Acad. Sci. USA* **109**, 5223–5228 (2012).
122. Yamamoto, H. et al. Molecular architecture of the ribosome-bound hepatitis C virus internal ribosomal entry site RNA. *EMBO J*. **34**, 3042–3058 (2015).
123. Juliano, R., Alam, M. R., Dixit, V. & Kang, H. Mechanisms and strategies for effective delivery of antisense and siRNA oligonucleotides. *Nucleic Acids Res*. **36**, 4158–4171 (2008).
124. Soler, M., McHutchison, J. G., Kwoh, T. J., Dorr, F. A. & Pawlotsky, J. M. Virological effects of ISIS 14803, an antisense oligonucleotide inhibitor of hepatitis C virus (HCV) internal ribosome entry site (IRES), on HCV IRES in chronic hepatitis C patients and examination of the potential role of primary and secondary HCV resistance in the outcome of treatment. *Antivir. Ther*. **9**, 953–968 (2004).

125. Zhang, H. et al. Antisense oligonucleotide inhibition of hepatitis C virus (HCV) gene expression in livers of mice infected with an HCV–vaccinia virus recombinant. *Antimicrob. Agents Chemother*. **43**, 347–353 (1999).
126. Hayden, F. G. & Shindo, N. Influenza virus polymerase inhibitors in clinical development. *Curr. Opin. Infect. Dis*. **32**, 176–186 (2019).
127. Noshi, T. et al. In vitro characterization of baloxavir acid, a first-in-class cap-dependent endonuclease inhibitor of the influenza virus polymerase PA subunit. *Antiviral Res*. **160**, 109–117 (2018).
128. Fukao, K. et al. Combination treatment with the cap-dependent endonuclease inhibitor baloxavir marboxil and a neuraminidase inhibitor in a mouse model of influenza A virus infection. *J. Antimicrob. Chemother*. **74**, 654–662 (2019).
129. Mishin, V. P. et al. Susceptibility of influenza A, B, C, and D viruses to baloxavir. *Emerg. Infect. Dis*. **25**, 1969–1972 (2019).
130. de Farias, S. T., Dos Santos Junior, A. P., Rêgo, T. G. & José, M. V. Origin and evolution of RNA-dependent RNA polymerase. *Front. Genet*. **8**, 125 (2017).
131. Guedj, J. et al. Antiviral efficacy of favipiravir against ebola virus: a translational study in cynomolgus macaques. *PLoS Med*. **15**, e1002535 (2018).
132. Omoto, S. et al. Characterization of influenza virus variants induced by treatment with the endonuclease inhibitor baloxavir marboxil. *Sci. Rep*. **8**, 9633 (2018).
133. Hofmann, W. P., Soriano, V. & Zeuzem, S. in *Antiviral Stategies: Handbook of Experimental Pharmacology* (eds Kräusslich H. G. & Bartenschlager, R.) 321–346 (2009).
134. Arts, E. J. & Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med*. **2**, a007161 (2012).
135. Bayat Mokhtari, R. et al. Combination therapy in combating cancer. *Oncotarget* **8**, 38022–38043 (2017).
136. Worthington, R. J. & Melander, C. Combination approaches to combat multidrug-resistant bacteria. *Trends Biotechnol*. **31**, 177–184 (2013).
137. Simmonds, P., Aiewsakun, P. & Katzourakis, A. Prisoners of war – host adaptation and its constraints on virus evolution. *Nat. Rev. Microbiol*. **17**, 321–328 (2019).
138. Krause-Kyora, B. et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife* **7**, e36666 (2018).
139. Mühlemann, B. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
Two articles[138,139] show that hepatitis B virus isolated from samples from the Bronze and Neolithic Age is minimally divergent from modern strains, which suggests that viral adaptation to the host may confer an upper limit to the mutations that can be accumulated over time.
140. Majzoub, K. et al. RACK1 controls IRES-mediated translation of viruses. *Cell* **159**, 1086–1095 (2014).
141. Ullah, H., Hou, W., Dakshanamurthy, S. & Tang, Q. Host targeted antiviral (HTA): functional inhibitor compounds of scaffold protein RACK1 inhibit herpes simplex virus proliferation. *Oncotarget* **10**, 3209–3226 (2019).
142. Heaton, N. S. et al. Targeting viral proteostasis limits influenza virus, HIV, and dengue virus infection. *Immunity* **44**, 46–58 (2016).
This paper reports that viruses exhibit host dependency of SEC61-mediated protein translocation and folding, and that inhibition of SEC61 provides a broad-spectrum inhibition on growth and infectivity in several viruses.
143. Shah, P. S. et al. Comparative flavivirus–host protein interaction mapping reveals mechanisms of dengue and Zika virus pathogenesis. *Cell* **175**, 1931–1945 (2018).
144. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
145. Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N. & Dwek, R. A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* **343**, 1235681 (2014).
146. Watanabe, Y., Bowden, T. A., Wilson, I. A. & Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta, Gen. Subj*. **1863**, 1480–1497 (2019).
147. Bojkova, D. et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).
148. Dhillon, P. et al. Cytoplasmic relocalization and colocalization with viroplasms of host cell proteins, and their role in rotavirus infection. *J. Virol*. **92**, e00612-18 (2018).
149. Zhao, N. et al. Influenza virus infection causes global RNAPII termination defects. *Nat. Struct. Mol. Biol*. **25**, 885–893 (2018).
150. Boudreault, S., Roy, P., Lemay, G. & Bisaillon, M. Viral modulation of cellular RNA alternative splicing: a new key player in virus-host interactions? *Wiley Interdiscip. Rev. RNA* **10**, e1543 (2019).
151. Bonenfant, G. et al. Zika virus subverts stress granules to promote and restrict viral gene expression. *J. Virol*. **93**, e00520-19 (2019).
152. Batra, R. et al. RNA-binding protein CPEB1 remodels host and viral RNA landscapes. *Nat. Struct. Mol. Biol*. **23**, 1101–1110 (2016).
153. Tremblay, M. P. et al. Global profiling of alternative RNA splicing events provides insights into molecular differences between various types of hepatocellular carcinoma. *BMC Genomics* **17**, 683 (2016).
154. Lin, J. C. Therapeutic applications of targeted alternative splicing to cancer treatment. *Int. J. Mol. Sci*. **19**, E75 (2017).
155. Urbanski, L. M., Leclair, N. & Anczuków, O. Alternative-splicing defects in cancer: splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip. Rev. RNA* **9**, e1476 (2018).
156. Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325 (1988).
157. Jang, S. K. et al. A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol*. **62**, 2636–2643 (1988).
This paper represents one of the first observations of IRESs as an alternative cap-independent translation mechanism that is used by encephalomyocarditis virus.
158. Griffiths, A. & Coen, D. M. An unusual internal ribosome entry site in the herpes simplex virus thymidine kinase gene. *Proc. Natl Acad. Sci. USA* **102**, 9667–9672 (2005).
159. Kang, S. T., Wang, H. C., Yang, Y. T., Kou, G. H. & Lo, C. F. The DNA virus white spot syndrome virus uses an internal ribosome entry site for translation of the highly expressed nonstructural protein ICP35. *J. Virol*. **87**, 13263–13278 (2013).
160. Zhao, J. et al. IRESbase: a comprehensive database of experimentally validated internal ribosome entry sites. *Genomics Proteomics Bioinformatics* **18**, 129–139 (2020).
161. Sweeney, T. R., Abaeva, I. S., Pestova, T. V. & Hellen, C. U. The mechanism of translation initiation on type 1 picornavirus IRESs. *EMBO J*. **33**, 76–92 (2014).
162. Kolupaeva, V. G., Lomakin, I. B., Pestova, T. V. & Hellen, C. U. Eukaryotic initiation factors 4G and 4A mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. *Mol. Cell. Biol*. **23**, 687–698 (2003).
163. Fernández, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823–831 (2014).
164. Deniz, N., Lenarcic, E. M., Landry, D. M. & Thompson, S. R. Translation initiation factors are not required for Dicistroviridae IRES function in vivo. *RNA* **15**, 932–946 (2009).
165. Fütterer, J., Kiss-László, Z. & Hohn, T. Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* **73**, 789–802 (1993).
166. Pooggin, M. M., Fütterer, J., Skryabin, K. G. & Hohn, T. A short open reading frame terminating in front of a stable hairpin is the conserved feature in pregenomic RNA leaders of plant pararetroviruses. *J. Gen. Virol*. **80**, 2217–2228 (1999).
167. Yueh, A. & Schneider, R. J. Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells. *Genes Dev*. **10**, 1557–1567 (1996).
168. Latorre, P., Kolakofsky, D. & Curran, J. Sendai virus Y proteins are initiated by a ribosomal shunt. *Mol. Cell. Biol*. **18**, 5021–5031 (1998).
169. Meyers, G. Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. *J. Biol. Chem*. **278**, 34051–34060 (2003).
170. Horvath, C. M., Williams, M. A. & Lamb, R. A. Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. *EMBO J*. **9**, 2639–2647 (1990).
171. Ahmadian, G., Randhawa, J. S. & Easton, A. J. Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. *EMBO J*. **19**, 2681–2689 (2000).
172. Gould, P. S. & Easton, A. J. Coupled translation of the respiratory syncytial virus M2 open reading frames requires upstream sequences. *J. Biol. Chem*. **280**, 21972–21980 (2005).
173. Jeudy, S., Abergel, C., Claverie, J. M. & Legendre, M. Translation in giant viruses: a unique mixture of bacterial and eukaryotic termination schemes. *PLoS Genet*. **8**, e1003122 (2012).
174. Schueren, F. & Thoms, S. Functional translational readthrough: a systems biology perspective. *PLoS Genet*. **12**, e1006196 (2016).
175. Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A. & Böck, A. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* **331**, 723–725 (1988).
176. Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T. C. & Hatfield, D. L. Identification of a selenocysteyl-tRNA(Ser) in mammalian cells that recognizes the nonsense codon, UGA. *J. Biol. Chem*. **264**, 9724–9727 (1989).
177. Ryan, M. D. & Drew, J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J*. **13**, 928–933 (1994).
178. Donnelly, M. L. L. et al. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol*. **82**, 1013–1025 (2001).
179. Luke, G. A. et al. Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. *J. Gen. Virol*. **89**, 1036–1042 (2008).
180. Sharma, P. et al. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res*. **40**, 3143–3151 (2012).
181. Ogino, T. & Banerjee, A. K. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell* **25**, 85–97 (2007).
182. Decroly, E., Ferron, F., Lescar, J. & Canard, B. Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol*. **10**, 51–65 (2011).
183. Ahola, T. & Kääriäinen, L. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl Acad. Sci. USA* **92**, 507–511 (1995).
184. Goodfellow, I. The genome-linked protein VPg of vertebrate viruses – a multifaceted protein. *Curr. Opin. Virol*. **1**, 355–362 (2011).
185. Sola, I., Almazán, F., Zúñiga, S. & Enjuanes, L. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu. Rev. Virol*. **2**, 265–288 (2015).
186. Kim, D. et al. The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921.e10 (2020).
187. Finkel, Y. et al. The coding capacity of SARS-CoV-2. *Nature* **589**, 125–130 (2021).

Acknowledgements I.M. is supported by Burroughs Wellcome Fund 1017892, NIH/NIAID-U01AI150748 and Chan Zuckerberg Initiative 2018-191895

Author contributions J.S.Y.H., Z.Z. and I.M. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to I.M.

Peer review information *Nature* thanks Sean Whelan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2021