Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Methods 183 (2020) 50–56

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Integrated approaches to reveal mechanisms by which RNA viruses reprogram the cellular environment

Christina Haddad[1], Jesse Davila-Calderon[1], Blanton S. Tolbert*

Department of Chemistry, Case Western Reserve University, Cleveland, OH 44106, United States

ABSTRACT

RNA viruses are major threats to global society and mass outbreaks can cause long-lasting damage to international economies. RNA and related retro viruses represent a large and diverse family that contribute to the onset of human diseases such as AIDS; certain cancers like T cell lymphoma; severe acute respiratory illnesses as seen with COVID-19; and others. The hallmark of this viral family is the storage of genetic material in the form of RNA, and upon infecting host cells, their RNA genomes reprogram the cellular environment to favor productive viral replication. RNA is a multifunctional biomolecule that not only stores and transmits heritable information, but it also has the capacity to catalyze complex biochemical reactions. It is therefore no surprise that RNA viruses use this functional diversity to their advantage to sustain chronic or lifelong infections. Efforts to subvert RNA viruses therefore requires a deep understanding of the mechanisms by which these pathogens usurp cellular machinery. Here, we briefly summarize several experimental techniques that individually inform on key physicochemical features of viral RNA genomes and their interactions with proteins. Each of these techniques provide important vantage points to understand the complexities of virus-host interactions, but we attempt to make the case that by integrating these and similar methods, more vivid descriptions of how viruses reprogram the cellular environment emerges. These vivid descriptions should expedite the identification of novel therapeutic targets.

1. Background

Mammalian Ribonucleic acid (RNA) viruses persist to pose serious threats to human health and global economies. As this article is being prepared, the world is living through a viral pandemic (COVID-19) caused by the Severe Acute Respiratory Syndrome-Related Coronavirus (SARS-CoV-2). SAR-CoV-2 is a positive-sense RNA virus that was first reported December of 2019 in the city of Wuhan in China. There are currently no vaccines or antivirals to prevent the spread of SARS-CoV-2. Knowledge of the structures, dynamics, and interactions of viral RNA genomes, like SARS-CoV-2, informs on essential mechanisms by which this large family of viruses reprogram the cellular environment to cause chronic or lifelong infections. When we understand the processes by which RNA viruses usurp their hosts, we are better positioned to develop novel strategies for therapeutic intervention. That level of understanding requires integrating multiple approaches and collaborating across scientific disciplines.

RNA is a diverse, multifunctional biomolecule that is involved in both the transfer and storage of genetic information as well as the modulation of a myriad of biological processes by virtue of its capacity to fold into complex structures and to catalyze biochemical reactions [1]. It is therefore no surprise that RNA viruses take advantage of the unique physicochemical properties of their viral genomes to assemble functional complexes, which in turn drives almost every aspect of their replication cycles within host cells (Fig. 1). Many of these complexes are formed through the recruitment of cognate RNA-binding proteins (RBPs) to specific genomic (or sub-genomic) loci, and the nature of these interactions manifest as signals that regulate each step of viral gene expression [2,3]. Thus, studying viral RNA structures and their interactions with cognate RBPs are essential to understanding the pathogenesis of RNA viruses and to further assist the design of novel antivirals. In this article, we attempt to describe how integrating methods that probe RNA structures and its interactions can inform on mechanisms that regulate viral gene expression for two representative positive-sense RNA viral families, namely Enteroviruses and Coronaviruses. Members from both of these families have caused widespread outbreaks in recent history.

1.1. Enteroviruses

Non-polio human enteroviruses (EV) are persistent pathogens that cause millions of infections in the United States and globally each year [4,5]. Infections typically manifest with mild illness; however, protracted infections in the immunocompromised (mostly infants, children and teenagers) can lead to severe neurological disorders, morbidity, paralysis, respiratory failure and death [6,7]. The National Institutes for

* Corresponding author.
E-mail address: bst18@case.edu (B.S. Tolbert).
[1] Contributing equally to this work.

https://doi.org/10.1016/j.ymeth.2020.06.013
Received 21 April 2020; Received in revised form 16 June 2020; Accepted 19 June 2020
Available online 02 July 2020

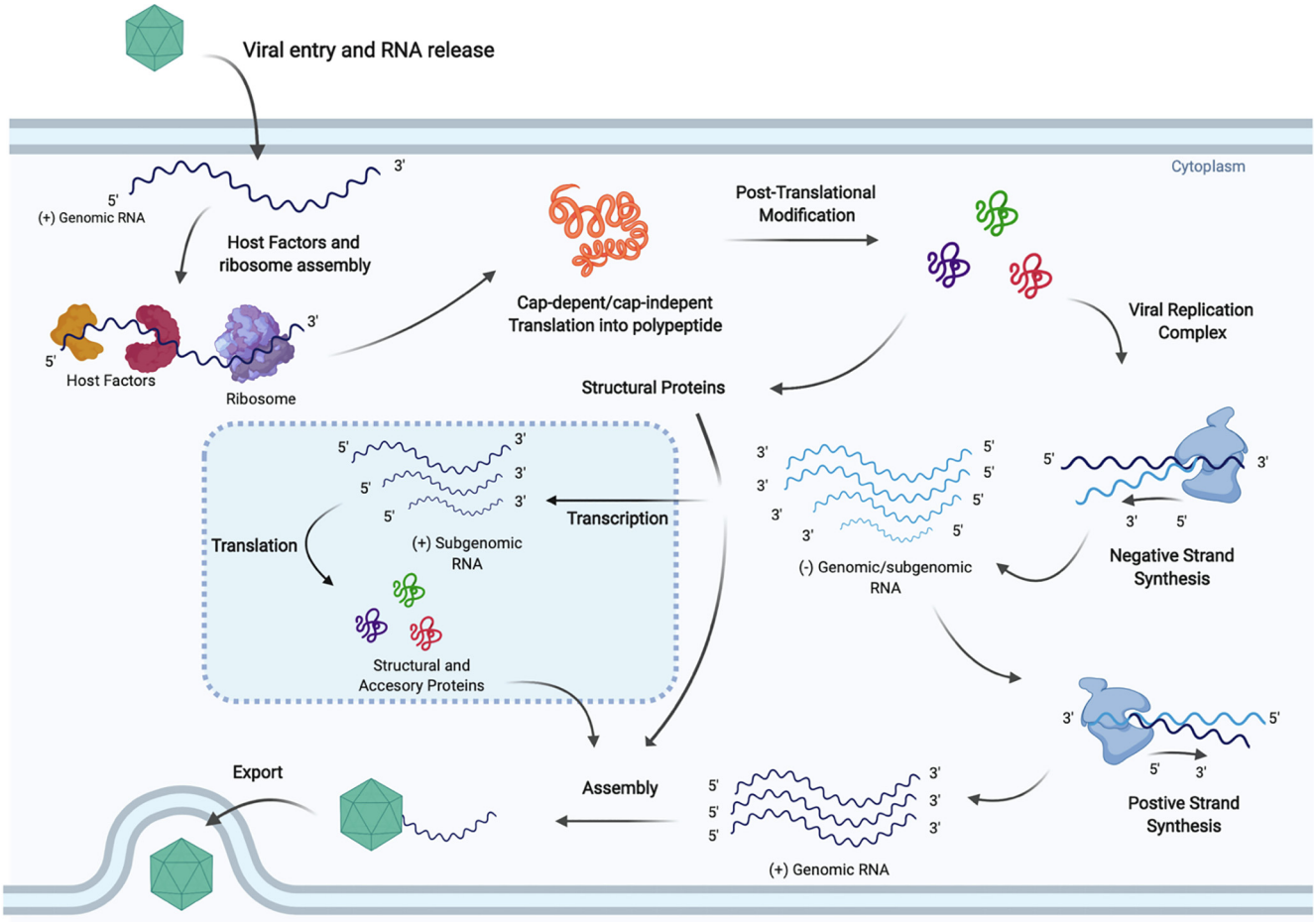

Fig. 1. Generalized replication cycle for positive-sense RNA viruses depicting viral pathways by which these viruses reprogram the cellular environment.

Allergies and Infectious Diseases identified EV-A71 and EVD68 as emerging infectious pathogens [8], and the World Health Organization discussed including both viruses in its Blueprint List of Priority Diseases [9]; emphasizing the serious threat that these viruses represent to public health. In a 2018 EV-A71 outbreak in Vietnam, 53,000 children were hospitalized and six died [10]. Similar cases with significant mortality rates have been reported in Taiwan and other parts of Asia-Pacific, thus, reiterating the urgency to develop antivirals or vaccines, and the necessity to better understand the molecular mechanisms involved in host-virus interactions [11].

EV-A71 is a non-enveloped single-stranded RNA virus that contains a 7500 nucleotide (nt) positive sense genome; a dual-purpose RNA element that must serve as template for both viral translation and genome replication [12,13]. Cellular entry is initiated through interactions between the viral capsid and host membrane receptors such as the scavenger receptor B2 (SCARB2), P-selectin glycoprotein ligand-1 (PSGL-1), heparan sulfate and annexin II (Anx2) and sialic acid-linked glycan [14]. Upon cellular entry, the single strand positive-sense viral RNA genome is released into the host cytoplasm. Given its limited coding capacity, EV-A71 uses multiple strategies to usurp host factors and modulate viral protein synthesis and replication. Particularly, the virus takes advantage of its highly structured 5′ untranslated region (5′ UTR) to initiate translation in a cap-independent pathway [13]. The 5′UTR is predicted to fold into six stem loops. Stem loop (SL) I adopts a 'cloverleaf' structure known to interact with the viral 3C protease to promote genome replication, whereas, stem loops II-VI form the active type I Internal ribosome entry site (IRES) involved in the recruitment of the ribosome [15]. The viral genome is translated in a cap-independent pathway, such that a single polyprotein is synthesized, which is further processed by the viral-encoded proteases 2A and 3C into structural and non-structural proteins [12]. Additionally, the viral-encoded proteases facilitate the shutdown of the host translation and transcription machinery, which produces the ideal environment for viral translation and replication, and ultimately apoptosis [16].

The EV-A71 genome cannot undergo translation and replication concurrently on the same genomic RNA, as the ribosome blocks the 3′-5′ progression of the elongating viral RNA polymerase [17]. Thus, the virus coordinates complex processes to transition between these two particular stages of its replication cycle. Specifically, genomes undergoing replication are translocated to virus-induced vesicles, allowing for spatial separation from those undergoing translation in the cytoplasm [18]. The EV-A71 genome is subsequently replicated through a negative strand intermediate and packaged into the viral capsid [12] (Fig. 1).

A myriad of distinct host RBPs, generally termed IRES trans-acting factors (ITAFs) are recruited to the EV-A71 IRES to regulate translation. Detailed descriptions of both their regulatory sites, and effects on IRES activity and replication have been reported elsewhere [15,19–23]. In brief, these include the positive regulators: polypyrimidine tract binding protein (PTB), poly(rC)-binding protein 1 (PCBP1), poly(rC)-binding protein 2 (PCBP2), far upstream element binding protein 1 (FBP1), Src-associated protein in mitosis (Sam68), human antigen R (HuR), Argonaute 2 (Ago2), heteronuclear ribonucleoprotein K (hnRNP K) and heteronuclear ribonucleoprotein A1 (hnRNP A1). Depletion of these factors in infected cells results in poor IRES activity and low viral titers in replication assays, indicating their connectedness to viral replication. Likewise, the far upstream element binding protein 2 (FBP2) and the heteronuclear ribonucleoprotein D (hnRNP D, also known as

the AU-rich element binding factor 1, (AUF1) have been identified as negative regulators since viral protein synthesis and IRES activity were significantly attenuated upon siRNA knockdown of these factors in infected cells. These ITAFs are nuclear proteins that are relocated to the cytoplasm upon viral infection, thus signaling changes to normal cellular functions [15,22]. Competition between the positive and negative regulators suggests a mechanism by which the virus fine-tunes its protein synthesis while simultaneously coordinating the reduction of the host cells translation levels and overall physiological homeostasis [19,24]. Interestingly, almost all of these ITAFs are known to interact with the 5′UTR of other picornaviruses to regulate IRES activity and replication, suggesting an evolutionary preference to conserve RNA structural features that drive specific RBP recognition [15,19]. The current dogma supports a model in which ITAFs cycle the IRES through different conformational states to modulate ribosome assembly; however, the molecular mechanisms by which ITAFs interact with the IRES to regulate this process remain poorly understood. Therefore, knowledge of the IRES structures and how ITAFs remodel it to assemble functional complexes is essential to understand how EV-A71 promotes viral protein synthesis to produce progeny virions.

1.2. Coronaviruses

December of 2019, the city of Wuhan in China witnessed an outbreak of the novel coronavirus disease (COVID-19), spreading to 185 countries and regions in months [25]. According to the Johns Hopkins Coronavirus Research Center, globally more than 7 million cases and more than 400,000 deaths were recorded as of June 12, 2020. The rapid spread of the infection is attributed to its ability to target the respiratory system [26]. This virus demonstrates similar symptoms to previously known coronaviruses, such as dry cough, dyspnea, and fever; however, it has the ability to infect lower respiratory airways leading to multiple organ failure in severe cases [27]. No treatment or vaccine are available to date.

COVID-19 is an infection caused by the severe acute respiratory syndrome-related coronavirus, namely SARS-CoV-2 and formally known as 2019-nCoV [25,28]. The SARS-CoV-2 virus belongs to the coronaviridae family, subfamily orthocoronavirinae, and genera betacoronaviruses, similar to MERS-CoV and SARS-CoV [25,28,29]. The SARS-CoV-2 genome displays 79.6% sequence identity to SARS-CoV, and 96% similarity to the bat-related coronavirus [30]. The enveloped coronavirus (CoV) genome is a positive-sense, single-strand RNA, which varies in size from 27 to 32 Kb, specifically 29.9 Kb in SARS- CoV-2 [29,31]. The enveloped virion carries a surface spike protein which binds to the host cell surface receptor, angiotensin converting enzyme 2 (ACE2).[32] This interaction promotes fusion of viral and cellular membranes, subsequently releasing the viral contents into the host cytoplasm. Upon cellular entry, the viral RNA genome is uncoated and released into the cytoplasm where it serves as a template for cap-dependent viral protein synthesis [33,34] (Fig. 1).

The open reading frames, ORF1a and ORF1b, covers two-thirds of the viral genome to encode two large polyproteins (pp1a and pp1ab), which are post-translationally cleaved into 16 nonstructural proteins (nsps) [29,34]. Polyprotein translation utilizes a ribosomal frameshifting mechanism that requires 5′ cap formation and 3′ polyadenylation of the viral genome [35,36]. Nsp3 or papain-like protease and mainly nsp5 or 3C-like protease ($3CL^{pro}$) are responsible for processing the polyprotein into mature nsps [34,37,38]. Structural and accessory proteins are encoded by the remaining one-third of the viral genome [33].

In CoVs, 5′ and 3′-untranslated regions (UTRs) consisting of phylogenetically conserved stem loops are required for replication and transcription [37,39]. Multiple nsps and a number of host factors assemble to form the replicase-transcriptase complex (RTC) at the 3′-UTR in order to synthesize genomic and subgenomic RNA (sgRNA) through negative strand template intermediates [34,39] (Fig. 1). RNA-dependent RNA polymerase (RdRp) or nsp12 replicates genomic RNA and transcribes sgRNA. Transcriptional regulatory sequences (TRSs) guide nsp12 and are present at the 5′-leader sequence (TRS-Leader or TRS-L) and at the genomes encoding for accessory and structural proteins (TRS-Body or TRS-B) [39]. RdRp continues transcription after encountering TRS-B sequences and switches to the TRS-L to transcribe the 5′ leader sequence; however, this mechanism is poorly understood [40]. Having common 5′-ends, sgRNAs translate only the 5′ segments of their ORFs into accessory and structural proteins and recognize the rest of the sequence as an untranslated region [33,41].

Comparatively to other coronaviruses, nsp1 promotes degradation of host mRNA and inhibits host cell translation [42]; a common strategy by which positive-sense RNA viruses reprogram the cellular environment [3]. In Murine Hepatitis Virus (MHV), a betacoronavirus, hnRNP A1 and polypyrimidine tract-binding protein (PTB) bind to the 5′-UTR, specifically at the TRS, and play a role in RNA synthesis [43,44]. In addition, eukaryotic initiation factors, eIFs, 3i, 3f, and 3e along with other host proteins assemble in the microenvironment of the replicase-transcriptase complex (RTC) [45]. The siRNA-mediated knockdown of these initiating factors showed a reduction in RNA replication, indicating their involvement in virus-dependent reprogramming of the cellular environment [45]. These aforementioned virus-host interactions and their mechanisms of action are yet to be understood.

The novelty of SARS-CoV-2 raises many questions concerning the mechanisms by which the virus regulates its gene expression. Obtaining structural details of the UTRs and identifying functional binding sites of RBPs will be deeply insightful in elucidating how this virus replicates within host cells. Focusing on essential RNA-RNA and RNA-protein interactions, such as RdRp, $3CL^{pro}$ [46,47], or cellular RBPs will inform on novel targets to therapeutically inhibit SARS-CoV-2, while simultaneously shedding light on the cellular pathways hijacked by the virus.

Despite differences in the life cycles of entero- and coronaviruses, sufficient similarities allow for the parallel discussion of shared features of their biology (Fig. 1). Positive-sense RNA viruses use their genomes as a template for both translation and replication; carry genes encoding for RNA-dependent RNA polymerases; utilize viral encoded proteases to shut-down cellular protein synthesis; and require interactions with specific RBPs to modulate multiple viral processes to coordinate shuttling between translation and replication stages. The molecular, biochemical and structural determinants governing these processes remain poorly understood, however. While it is plausible that a single technique can illuminate mechanistic aspects of stages within the cellular replication cycles of these viruses, their complexities necessitate a more integrative approach to account for the large number of variables.

The goal of this article is to highlight the effectiveness of integrating several techniques (as discussed in this edition) to overcome major barriers to understanding viral RNA structures and their interactions with cognate proteins. Secondary structural models of viral RNAs can be constructed using computational predictions (Moss et al), chemical probing (Rouskin et al), or antisense oligonucleotide based hybridization (Contreras et al); and in many cases, even higher-resolution atomic details of viral RNAs can be revealed by NMR spectroscopy (Keane et al). Since most RBPs interact with RNAs through unpaired regions, CLIP-seq (Kutluay et al) provides a global map of sites occupied by proteins along viral RNA structures. Ideally, these techniques, and others like them [21,48–52], complement each other such that the collective results provide a comprehensive description of the biochemical mechanisms by which viral RNA structures and their interactions contribute to viral replication.

2. Computational and biochemical approaches to study viral RNA structures and its interactions

2.1. ScanFold: computational insight into RNA secondary structure

Advancements in high-throughput sequencing technologies have

allowed for quick, affordable and reliable determination of DNA and RNA sequences. This has further expedited the study of viral genomes and the potential of having a better understanding of their biology; however, much of these sequences require further structural and functional characterization to truly understand mechanisms. Thus, computational methods have been developed to complement experimental RNA analysis and to serve as benchmarks to test for functional relevance of structural elements.

As described by Moss et al, ScanFold is a single-sequence computational method implemented to identify functional RNA structural motifs unlike traditional methods, such as align-and-fold or fold-and-align, which rely on previous sequence alignments [53]. ScanFold decouples these steps, which minimizes computational time and aids researchers studying systems with poor sequence alignments [54]. Potential functional regions on the RNA are identified by analyzing the thermodynamic parameter z-score from which a single base pair arrangement is assigned to each nucleotide in the input sequence and a structural model is built [54].

The ScanFold pipeline was previously benchmarked against experimentally supported models of the well-studied HIV-1 genome [53], and it has been used recently to identify thermodynamically stable RNA structures throughout the SARS-CoV-2 genome (https://www.biorxiv.org/content/10.1101/2020.04.17.045161v1). A similar approach can be employed to identify probable functional regions along the genome of other RNA viruses, which are thought to coordinate multiple aspects of their replication cycles through co-opting RBPs (Fig. 1). Furthermore, ScanFold results can be complemented with sequence alignment data to identify functional regions that have been evolutionarily conserved in RNA viruses.

Although ScanFold can identify RNA sequences with potential to form stable structures, the obvious “limitation” is that the structures are predicted and therefore need to be validated experimentally by other structural methods described in this article (Fig. 2). Particularly, these structures predicted *in silico* can be further confirmed experimentally using chemical probing approaches such as DMS-MaPseq.

2.2. Probing RNA secondary structure via Dimethyl sulfate mutational profiling and sequencing (DMS-MaPseq)

Viral RNA structures can regulate host protein interactions and gene expression. Solving RNA structures can help narrow down the function of different structural elements that contribute to replication, transcription, or translation. The complex structure formation of single stranded RNA can range from Watson-crick base pairing to other non-canonical elements such as bulges, internal and apical loops. Even more complicated structures that involve long-range interactions such as pseudoknots play key roles in viral replication [40]. Chemical probing of RNA is a widely used technique to study RNA structures. As described by Rouskin et al, Dimethyl sulfate (DMS) probing coupled with mutational profiling and high throughput sequencing (DMS-MaP seq) has proven to provide robust RNA secondary structures either *in vitro*, *in vivo*, or *in virion* [55].

DMS modifies non-base paired adenosine and cytosine nucleotides present in bulges, loops, and other regions where the Watson-Crick edges of these bases are exposed. Using thermostable group II intron reverse transcriptase (TGIR-II), the modified RNA is reverse transcribed by creating a mutation when it comes across a methylated nucleotide. High throughput sequencing of the products will generate DMS driven RNA secondary structures that will highlight open regions, such as bulges or loops. Of note, these DMS reactivity patterns are typically used as pseudo-energy restraints in RNA structure prediction algorithms. *In vivo* DMS-Map seq allows the analysis of the entire viral genome including host genes. The latter feature allows for a comparative analysis as to how virus infection changes host RNA structures. Solving viral RNA secondary structures *in vivo* renders a clearer understanding of cellular environmental effects, structural changes as infection progresses, mutational effects on RNA structures, and how structural changes impact function.

As discussed above, EV-A71 controls viral translation and replication through its 5′-UTR, and the UTRs of CoV are also used to modulate its gene expression. Despite their distinctive expression pathways, mapping their 5′ and 3′ RNA secondary structures using DMS-MaP seq and identifying unpaired RNA regions can be insightful in identifying

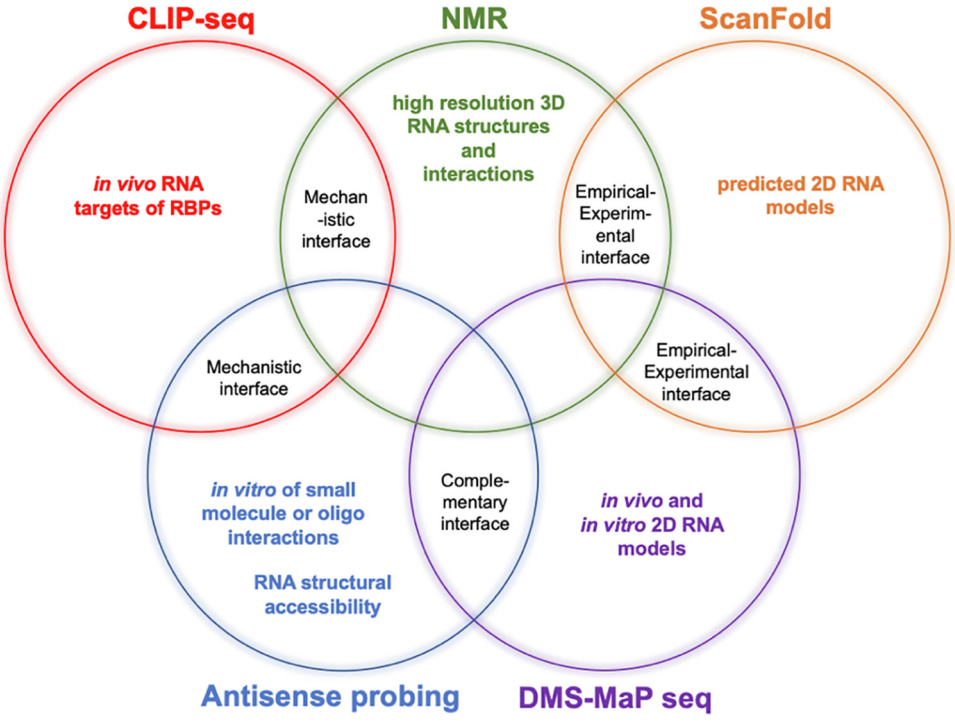

Fig. 2. Integrated approaches to develop comprehensive descriptions of virus-host interactions essential for viral replication.

RNA-protein binding sites and exposed surfaces to facilitate long-range interactions. Assessment of structural changes under the influence of RBPs, such as ITAFs and viral proteins, could identify RNA structural elements involved in protein interactions during translation or replication. Moreover, this approach can also inform on protein-induced RNA conformational changes that might be important for viral function. Highlighting RNA secondary structures and possible RNA-protein binding sites can help characterize the 3′ end RTC for CoV and the 5′-UTR recruitment of ITAFs in EV-A71.

DMS-MaP seq has proven to be a reliable quantitative high throughput probing method to generate quantitative analysis of RNA secondary structures *in vivo*, which produces a higher degree of insight when coupled with NMR or other advanced structural or computational methods. Because of the ability to probe RNA structure within native biological contexts, DMS-Map seq and related technologies has the advantage of providing insights into viral RNA structures under physiological conditions. One limitation of DMS-Map seq is that only the Watson-Crick edge of exposed A and C bases are typically detected as modified, therefore the information content harvested is most reliable for determining which bases are not involved in stable Watson-Crick base pairs. Therefore, antisense RNA probing can fill in gaps missed by DMS-MaP seq (Fig. 2).

2.3. Antisense probing of RNA structure accessibility

RNA-RNA and RNA-protein interactions are crucial for viral RNA function and regulation. The propensity of any interaction taking place depends on the structure of the RNA. The degree of accessible regions on the RNA can facilitate these interactions. Alterations in accessible regions are not only due to structural rearrangement and long-range RNA interactions, but also due to other protein or nucleic acid binding events that can block certain RNA regions. Antisense probing techniques (as described by Contreras et al) can identify the accessibility profile of RNA regions that are exposed and this information can aid in the determination of the structural changes that accompany RNA-RNA and/or RNA-protein interactions [56].

In the aim of studying structural RNA accessibility, antisense RNA probing utilizes a Structural Sensing System (iRS^3), a previously designed *in vivo* sensor, that utilizes an antisense RNA probe that hybridizes to the RNA of interest. This probe constitutes of a 9 to 16 nucleotide RNA complementary to its target, a stem loop structure that blocks the ribosome binding site (RBS) by binding to the cis-acting region, and a GFP encoding region to generate a fluorescent output signal [56].

The *in vitro* antisense probing technique enables the systematic study of environmental factors. This cell-free iRS^3 (CF- iRS^3) method gives the advantage of studying the influence of individual elements on the accessibility of the RNA. A change in accessible regions can confirm binding sites of proteins, oligomers, or small molecules and identify RNA conformational changes that accompany these interactions. Its structural accessibility profiles can reveal the impact of RBPs on rearranging RNA conformations such as those changes that stimulate IRES-dependent translation.

In EV-A71 and CoV, translation of viral proteins is one of the initial steps following cellular entry. Profiling RNA accessible regions for the 5′-UTR of EV-A71 and understanding RBPs or ITAF alterations of these regions will generate information on binding sites and their influence on folding. Studying the 3′-UTR along with the 5′-UTR, can relay information on replication pathways for both viruses (Fig. 1). For CoV, the RTC at the 3′ end is an interesting target to assess, whether in studying structural accessibility or RNA-protein binding sites or determining how TRS interactions complete CoV genome replication. The transcription and translation of sgRNA depend on a common 5′ leader sequence; this makes it a valuable RNA target to comprehend its structure and protein recruitment. That being said, the antisense RNA probing technique requires *a priori* knowledge on RBPs to efficiently study these interactions; this shortcoming can be addresses through the use of CLIP-seq.

2.4. Crosslinking immunoprecipitation coupled with sequencing (CLIP-seq)

Common to positive-sense RNA viruses, RNA-binding proteins and other factors are recruited to specific regions across the genome to regulate essential processes of the viral life cycle. As such, identifying these RNA-protein interactions are crucial to understanding viral pathogenesis. As described by Kutluay et al, the utility of CLIP-seq variants relies on determining RBP binding sites on the RNA at nucleotide resolution, while experiments are routinely performed under physiological conditions within cells and yield low background signal [57].

The CLIP-seq framework was used in a recent study to identify binding sites for the splicing regulators hnRNP A1 and hnRNP H1 along the HIV-1 genome [58]. That study revealed RBP binding sites proximal to splice acceptor and donor signals that control HIV splicing. Mutations of select binding sites resulted in changes in HIV splicing patterns that had impacts on viral replication. NMR spectroscopy experiments carried out on protein-RNA interaction identified from CLIP-seq offered additional mechanistic insights into sequence specific recognition of the hnRNP H protein for its HIV targets. Given the large number of RBPs known to interact with genomic and subgenomic viral RNAs to modulate translation, replication and the shift between these two stages, CLIP-seq can be employed to understand virology at the molecular level. Altogether, this methodology provides the experimental scaffold to study host-viruses interactions more comprehensively and design novel strategies to design antivirals.

As demonstrated for the HIV CLIP-seq study, it is useful to couple CLIP-seq with other structural approaches to provide more mechanistic details on the RNA physicochemical features that contribute to form functional RNP complexes (Fig. 2). A drawback to CLIP-seq is that no information of the surrounding RNA structural environment is obtained, which thus compromises interpretations of the influence of RNA-RNA interactions on RNA-protein binding events. This limitation can be overcome by incorporating DMS-Map seq or other *in vivo* chemical probing techniques into the protocol. Once *in vivo* binding sites are determined, NMR spectroscopy can complement the study of RNA-protein interactions by mapping binding interfaces and defining the three-dimensional structures of both the RNA and interacting proteins.

2.5. Visualizing RNA secondary and three dimensional structure through nuclear magnetic resonance (NMR) spectroscopy

Among the various approaches to study biomolecules at the atomic level, Nuclear Magnetic Resonance (NMR) spectroscopy remains the go to analytical tool to study in-solution RNA structure, its dynamics and, interactions with other molecules which provides insightful information on the function and activity of RNA elements and molecular recognition events [59,60]. NMR has proven to be a valuable technique to determine RNA structure and study relevant conformational dynamics. Multiple advances in RNA sample preparation, data acquisition and analysis have been developed and reviewed elsewhere [52,60].

As described by Keane et al, advancements in chemical and enzymatic synthesis have allowed the enrichment of the RNA with a second NMR-active nuclei, which can significantly remove spectral crowding and improve spectral resolution to both facilitate resonance assignments and allow for the development of innovative and selective pulse sequences. Local distance restraints can be derived from nuclear Overhauser effect spectroscopy (NOESY) which informs on the particular geometry of the sugar moiety, base stacking and base pairing types [59]. Furthermore, global structural restraints can be obtained from residual dipolar coupling (RDC), paramagnetic relaxation enhancement (PRE) and by employing hybrid approaches such as small angle X-ray scattering (SAXS). Altogether, both experimentally derived constraints (NOE, RDC, PRE derived distance restraints, and H bond

information) and empirically-derived restraints are weighted and loaded into software packages, such as Xplor-NIH, that calculate initial structures followed by refinement in extended molecular dynamics simulations in AMBER.

Collectively, gaining insight into dynamic features of important RNA motifs has been associated with the function and activity of such RNAs. Reconstructing high-resolution three-dimensional structure ensembles of functional RNA elements is key to understand viral pathogenesis at near-atomic resolution. Furthermore, NMR titration experiments can be performed to map RBPs binding sites along an RNA surface and study conformational dynamics induced after binding. The obvious limitation of NMR is that it is usually restricted to studying molecular systems less than ~50 kDa and in buffer conditions that approximate the cellular environment. Nevertheless, NMR is a powerful analytical technique that complements the aforementioned methods to provide a more in depth understanding of molecular mechanisms.

3. Discussion

Positive sense RNA viruses still remain a threat to humanity, especially due to the recent outbreak of the COVID-19 pandemic. No treatments or vaccine are available for SARS-CoV-2 as well as for EV-A71 or EVD68. The first step to understanding their mechanisms of action is to elucidate how the viral RNA elements regulate essential functions and takes over the host cellular machinery. Each of these viruses enter the cell, replicate, and translate in distinctive pathways, indicating that there are multiple targetable RNA structures to pursue for drug discovery efforts. Coupling ScanFold, DMS-MaP seq, antisense, CLIP-seq, NMR or related technologies can guide the path to understanding these processes and to expediting the discovery of novel targets for therapeutic intervention. Structures alone are just the start, it will be critical to simultaneously validate these structures and to determine at one points in the life cycles of these viruses these structures function.

For SARS-CoV-2, replication and transcription are initiated through the assembly of viral and host RBPs at the 3′-UTR forming the RTC [40]. The primary involvement of the 3′-UTR makes this complex RNA a valuable target to determine its structure so as to consider inhibiting SARS-CoV-2 at the early stage of viral genome replication. At the time of writing this article, Moss et al reported a ScanFold 2D RNA model (https://www.biorxiv.org/content/10.1101/2020.04.17.045161v1) of the SARS-CoV-2 RNA genome and independently the Das Lab used related technologies to arrive at similar models (https://doi.org/10.1101/2020.03.27.012906). To validate and complement these initial models, *in vitro* and *in vivo* secondary structure determination can be achieved through DMS-MaP seq. Since routine applications of DMS-MaP seq reports only on exposed adenosines and cytosines, antisense oligo probing should provide complementary information on all nucleotide types exposed within unpaired regions and those likely involved in long-range tertiary interactions. Once secondary structural models are known, high-resolution 3D NMR structures can be obtained in association with computational methods (https://doi.org/10.1101/2020.04.14.041962), producing models based on integrated data sets. By adapting this integrated approach, viral RNA structural models, which are cross-validated, can be determined for the 3′UTR of SARS-CoV-2 as well as others.

In addition to structure, the RNA-protein interactions and their effect on RNA folding are yet to be studied in the SARS-CoV-2 RTC. CLIP-seq can identify the RBP RNA-targets at the nucleotide level *in vivo*. With prior knowledge of the binding sites, antisense RNA probing can confirm these interactions and identify their effects on RNA folding. Coupling with NMR, binding sites for sub-domains of the RTC can be mapped onto the RNA structure. Since NMR assignments will be available, it should be straightforward to assess the extent of binding induced conformational changes. Although the RTC is used as one example, these integrated technologies and others like them should provide mechanistic insights into a wide-range of host-virus pathways (Fig. 2). For instance, RdRp plays a primary role in viral transcription and replication [40]. Uncovering its binding site on the 3′-UTR can aid in designing a site-specific drug to prevent this interaction. Also, monitoring the 3′-pseudo-knot's structural interactions under the effect of RdRp can give insight on the molecular switching mechanism for RNA synthesis. This can be tested under the influence of different viral or cellular factors at different environmental conditions. In addition, RdRp transcribes sgRNA through the guidance of TRSs and host proteins [40]. Identifying the binding sites of specific host proteins to the TRS will help investigate the mechanism of protein recruitment for transcriptional purposes. Several other RNA involved processes can be studied under the effect of viral or host factors.

Similar to SARS-CoV-2, EV-A71 and other viruses can be effectively studied by integrating structural techniques. This blueprint can be adopted for any RNA virus. Studying RNA structural interactions and the effects of viral-host RBPs on RNA structure and function are essential for understanding translation, replication, and transcription processes in order to better understand how viruses reprogram the cellular environment. The effectiveness of integrating these approaches can help design and test more effective and site-specific small molecules.

As we deal with the COVID-19 pandemic and prepare for the next viral outbreak, we hope that this article will encourage adopting integrative (collaborative) approaches that will enhance our understanding of viral RNA structures, their interactions with cognate proteins, and in turn the mechanisms by which viruses reprogram the cellular environment. With this comprehensive level of knowledge, we expect that the discovery of novel therapeutic agents will be accelerated.

Acknowledgement

This work was funded by National Institutes of Health grants U54AI50470 (the Center for HIV RNA Studies) and R01GM126833 (BST).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2020.06.013.

References

[1] S.A. Mortimer, M.A. Kidwell, J.A. Doudna, Insights into RNA structure and function from genome-wide studies, Nat. Rev. Genet. 15 (7) (2014) 469–479.
[2] Z. Li, P.D. Nagy, Diverse roles of host RNA binding proteins in RNA virus replication, RNA Biol. 8 (2) (2011) 305–315.
[3] Strauss, J.H.a.S., Ellen G, Virus and Human Disease. 2 ed. 2008: Elseveir. 480.
[4] N.J. Schmidt, E.H. Lennette, H.H. Ho, An apparently new Enterovirus isolated from patients with disease of the central nervous system, J. Infect. Dis. 129 (3) (1974) 304–309.
[5] Center for Disease, C. and Prevention, About Non-Polio Enteroviruses.
[6] H.I. Huang, S.R. Shih, Neurotropic enterovirus infections in the central nervous system. 2015.
[7] K.C. Ong, K.T. Wong, Understanding enterovirus 71 neuropathogenesis and its impact on other neurotropic enteroviruses, Brain Pathol. (2015).
[8] National Institute of, A. and D. Infectious, NIAID Emerging Infectious Diseases/Pathogens.
[9] World Health, O., List of Blueprint priority diseases.
[10] L.N.T. Nhan, et al., Severe enterovirus A71 associated hand, foot and mouth disease, Vietnam, 2018: Preliminary report of an impending outbreak. Eurosurveillance, 2018.
[11] P.C. McMinn, Recent advances in the molecular epidemiology and control of human enterovirus 71 infection, Curr. Opin. Virol. 2 (2) (2012) 199–205.
[12] J.L. Whitton, C.T. Cornell, and R. Feuer, Host and virus determinants of picornavirus pathogenesis and tropism. 2005.
[13] S.R. Thompson, P. Sarnow, Enterovirus 71 contains a type I IRES element that functions when eukaryotic initiation factor eIF4G is cleaved, Virology (2003).
[14] S. Yamayoshi, K. Fujii, S. Koike, Receptors for enterovirus 71, Emerging Microb. Infect. 3 (7) (2014) e53-e53.
[15] S.R. Shih, V. Stollar, M.L. Li, Host factors in Enterovirus 71 replication, J. Virol. 85

(19) (2011) 9658–9666.
[16] M.-L. Li, et al., EV71 3C protease induces apoptosis by cleavage of hnRNP A1 to promote apaf-1 translation, PLoS One 14 (9) (2019) e0221048-e221048.
[17] J.-Y. Lin, et al., Viral and host proteins involved in picornavirus life cycle, J. Biomed. Sci. 16 (1) (2009) 103.
[18] P. Ahlquist, et al., Host factors in positive-strand RNA virus genome replication, J. Virol. 77 (15) (2003) 8181.
[19] J.Y. Lin, M.L. Li, S.R. Shih, Far upstream element binding protein 2 interacts with enterovirus 71 internal ribosomal entry site and negatively regulates viral translation, Nucl. Acids Res. (2009).
[20] J.Y. Lin, et al., Heterogeneous nuclear ribonuclear protein K interacts with the enterovirus 71 5′ untranslated region and participates in virus replication, J. Gen. Virol. (2008).
[21] M. Tolbert, et al., HnRNP A1 alters the structure of a conserved enterovirus IRES domain to stimulate viral translation, J. Mol. Biol. (2017).
[22] K.M. Lee, C.J. Chen, S.R. Shih, Regulation Mechanisms of Viral IRES-Driven Translation, Elsevier Ltd., 2017, pp. 546–561.
[23] J.Y. Lin, G. Brewer, M.L. Li, HuR and Ago2 bind the internal ribosome entry site of enterovirus 71 and promote virus translation and replication, PLoS One 10 (10) (2015) e0140291.
[24] P.N. Huang, et al., Far upstream element binding protein 1 binds the internal ribosomal entry site of enterovirus 71 and enhances viral translation and viral growth, Nucl. Acids Res. (2011).
[25] C. Liu, et al., Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases, ACS Publications, 2020.
[26] Y.-Y. Zheng, et al., COVID-19 and the cardiovascular system, Nat. Rev. Cardiol. (2020) 1–2.
[27] H.A. Rothan, S.N. Byrareddy, The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, J. Autoimmun. (2020) 102433-102433.
[28] A.E. Gorbalenya, et al., The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, Nature Microbiol. (2020).
[29] Y. Fan, et al., Bat coronaviruses in China, Viruses 11 (3) (2019) 27–32.
[30] P. Zhou, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature (2020) 1–4.
[31] Y.-R. Guo, et al., The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status, Mil. Med. Res. 7 (1) (2020) 1–10.
[32] J. Shang, et al., Cell entry mechanisms of SARS-CoV-2, Proc. Natl. Acad. Sci. (2020) 11727–11734.
[33] H.J. Maier, E. Bickerton, P. Britton, Coronaviruses: Methods and Protocols, Springer, Berlin, 2015.
[34] E. de Wit, et al., SARS and MERS: recent insights into emerging coronaviruses, Nat. Rev. Microbiol. 14 (8) (2016) 523-523.
[35] G. Taiaroa et al., Direct RNA sequencing and early evolution of SARS-CoV-2. bioRxiv, 2020.
[36] M. Macchiagodena, M. Pagliai, and P. Procacci, Inhibition of the main protease 3cl-pro of the coronavirus disease 19 via structure-based ligand design and molecular modeling. arXiv preprint arXiv:2002.09937, 2020.
[37] J.W. Oh, D.G. Ahn, Recombinant SARS-CoV nsp12 and the use of thereof and the method for producing it. 2013, Google Patents.
[38] Y.-C. Chang et al., Potential therapeutic agents for COVID-19 based on the analysis of protease and RNA polymerase docking. 2020.
[39] A.R. Fehr, S. Perlman, Coronaviruses: An Overview of their Replication and Pathogenesis, Springer, 2015, pp. 1–23.
[40] I. Sola, et al., Continuous and discontinuous RNA synthesis in coronaviruses, Ann. Rev. Virol. 2 (2015) 265–288.
[41] K. Nakagawa, K.G. Lokugamage, S. Makino, Viral and Cellular mRNA Translation in Coronavirus-Infected Cells, Elsevier, 2016, pp. 165–192.
[42] W. Kamitani, et al., A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein, Nat. Struct. Mol. Biol. 16 (11) (2009) 1134-1134.
[43] H.-P. Li, et al., Polypyrimidine tract-binding protein binds to the leader RNA of mouse hepatitis virus and serves as a regulator of viral transcription, J. Virol. 73 (1) (1999) 772–777.
[44] H.-P. Li, et al., Heterogeneous nuclear ribonucleoprotein A1 binds to the transcription-regulatory region of mouse hepatitis virus RNA, Proc. Natl. Acad. Sci. 94 (18) (1997) 9544–9549.
[45] P. V'Kovski, et al., Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling, Elife 8 (2019) e42037-e42037.
[46] M.U. Mirza, M. Froeyen, Structural elucidation of SARS-CoV-2 vital proteins: computational methods reveal potential drug candidates against Main protease, Nsp12 RNA-dependent RNA polymerase and Nsp13 helicase. 2020.
[47] L. Kiemer, et al., Coronavirus 3CL pro proteinase cleavage sites: possible relevance to SARS virus pathology, BMC Bioinf. 5 (1) (2004) 72.
[48] K. Kappel et al., Ribosolve: rapid determination of three-dimensional RNA-only structures. bioRxiv, 2019: p. 717801.
[49] M. Kubota, C. Tran, R.C. Spitale, Progress and challenges for chemical probing of RNA structure inside living cells, Nat. Chem. Biol. 11 (12) (2015) 933–941.
[50] F.C.Y. Lee, J. Ule, Advances in CLIP technologies for studies of protein-RNA interactions, Mol. Cell 69 (3) (2018) 354–369.
[51] E.J. Strobel, A.M. Yu, J.B. Lucks, High-throughput determination of RNA structures, Nat. Rev. Genet. 19 (10) (2018) 615–634.
[52] A. Schlundt, J.N. Tants, M. Sattler, Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition, Methods 118–119 (2017) 119–136.
[53] R.J. Andrews, J. Roche, W.N. Moss, ScanFold: an approach for genome-wide discovery of local RNA structural elements—applications to Zika virus and HIV, PeerJ 6 (2018) e6136.
[54] R.J. Andrews, L. Baber, W.N. Moss, Mapping the RNA structural landscape of viral genomes, Methods (2019).
[55] P. Tomezsko, H. Swaminathan, S. Rouskin, Viral RNA structure analysis using DMS-MaPseq, Methods (2020).
[56] A.J. Lukasiewicz, L.M. Contreras, Antisense probing of dynamic RNA structures, Methods (2020).
[57] C.S. Mugisha, K. Tenneti, S.B. Kutluay, Identifying the RNA targets of RNA binding proteins with clip, Methods (2019).
[58] S.B. Kutluay, et al., Genome-wide analysis of heterogeneous nuclear ribonucleoprotein (hnRNP) binding to HIV-1 RNA reveals a key role for hnRNP H1 in alternative viral mRNA splicing, J. Virol. 93 (21) (2019) e01048–e1119.
[59] B. Fürtig, et al., NMR spectroscopy of RNA, ChemBioChem 4 (10) (2003) 936–962.
[60] R.P. Barnwal, F. Yang, G. Varani, Applications of NMR to structure determination of RNAs large and small, Arch. Biochem. Biophys. 628 (2017) 42–56.