# ESSAY

# SAMUEL LAZZERI S3811433

Examiner: Dr. Danny Incarnato

# Antiviral strategies against RNA viruses

# An overview from classical therapies to innovative approaches targeting RNA structures



# ABSTRACT

RNA viruses occupy a relevant position among the human pathogens, counting a continuously increasing number of different species, which can cause more or less severe diseases in humans. Impressive research efforts over the last decades have led to the development of a variety of different antiviral strategies, but the biological diversity and rapid adaptive rates of RNA viruses have proven to be difficult to overcome. This implies a continuous research for new antiviral compounds and strategies. In this respect, RNA structural elements have been recently recognized to participate to virus functioning and could therefore be appealing and innovative antiviral pharmaceutical targets.

Here we report the most common antiviral strategies currently applied, along with the factors underlying the exceptional ability of RNA viruses in developing resistance. We then analyse the possibility to target viral RNA structural elements with antiviral purposes, highlighting the most problematic issues concerning this approach and reporting the most advanced progresses in this direction. Although the field can be still considered on its early development and several crucial challenges still need to be addressed, the antiviral strategy of targeting RNA structural elements has an enormous potential and could dramatically improve our antiviral arsenal, which is of critical importance considering the emerging crisis of viral drug resistance.

# TABLE OF CONTENTS

ABSTRACT	2
INTRODUCTION TO VIRUSES	4
RNA VIRUSES	5
PATHOGENESIS OF RNA VIRUSES	6
NOTABLE RNA VIRUSES INFECTING HUMANS	7
CURRENT STRATEGIES IN TREATING RNA VIRUSES	9
RNA STRUCTURES	13
HOW TO STUDY RNA STRUCTURES	16
VIRAL RNA STRUCTURES	21
RNA STRUCTURES AS TARGET OF SMALL MOLECULES	31
IDENTIFYING SMALL MOLECULES BINDING RNA	33
ADVANCES IN TARGETING VIRAL RNA STRUCTURES	38
DISCUSSION AND CONCLUSION	44
REFERENCES	46
FIGURES AND TABLES SOURCES	65

### **INTRODUCTION TO VIRUSES**

Viruses are defined as infectious agents that can only replicate within a host organism, namely they are obligate intracellular parasites<sup>1</sup>. In fact, as viruses do not possess the machinery required to replicate, they need to usurp the host one for producing progeny and propagating. Virus particles are typically composed of a nucleic acid genome or core, which is the genetic material of the virus, surrounded by a capsid made up of virus-encoded proteins<sup>2</sup>. Viral genetic material encodes also other viral proteins involved in virus replication. In some viruses, the protein shell is enclosed in a lipid membrane called envelope, which is usually derived from the cell in which the virus replicates. The virus life cycle refers to the multiple steps involved in the virus propagation and can be divided into three main stages: entry, genome replication and exit<sup>3</sup>. Entry is the first stage and involves attachment, in which a virus particle encounters the host cell and attaches to the cell surface, penetration, in which a virus particle reaches the cytoplasm, and uncoating, in which the virus sheds its capsid. Following the uncoating, the naked viral genome is utilized for gene expression and viral genome replication. Distinct replication strategies characterize different viral families, but all viruses share the feature of entirely relying on the host translation machinery for the protein synthesis. Finally, when the viral proteins and viral genomes have accumulated, they are assembled to form progeny virion particles and then released extracellularly. Virion assembly and the release from the cell constitute the exit, which is the last stage of the virus life cycle.

To date, more than 5000 different genotypes of viruses have been identified, which can infect a variety of living organisms, from bacteria to plants to animals, including humans<sup>4</sup>. In particular, more than 200 different viruses are known to be capable of breaking into human cells and cause disease in humans<sup>5</sup>. The development of self-consistent classification schemes for this plethora of entities is a major challenge for virologists. The International Committee on Taxonomy of Viruses (ICTV) identified a limited number of viral features that can be used for classification, among which there is the nature of the viral genome<sup>6</sup>. Viral genomes are very diverse, since they can be DNA or RNA, single or double-stranded, linear or circular, and vary in length and in the number of molecules. Nevertheless, a specific type of genome is always the same for any given type of virus.

#### **RNA VIRUSES**

RNA viruses are by definition viruses that have RNA as genetic material<sup>7</sup>. While DNA viruses share many common patterns of gene expression and genome replication with the host cell, viruses using RNA as genetic material have devised some strategies to replicate such material, since the cell does not have machinery for RNA-directed RNA replication. In fact, the replication of RNA viruses requires specific enzymes that are not present in the uninfected host cell.

The genetic material of RNA viruses can be either single-stranded (ssRNA), as observed in the majority of the cases, or double-stranded (dsRNA). An important classification of ssRNA viruses is based on the sense or polarity of the RNA, namely on whether the viral genome can be directly utilized as mRNA or whether it must first be transcribed into mRNA. Positive-sense RNA viruses have the genome in the same sense as mRNA and thus can be immediately translated by the host cell. Negative-sense viruses carry a genome with the opposite sense, namely complementary, compared to mRNA, which must therefore be converted to positive-sense RNA by an RNAdependent RNA polymerase before translation. There are then ambisense RNA viruses, whose genome is characterized by RNA that is in part of positive and in part of negative polarity<sup>8</sup>. RNA viruses with double-stranded genomes have obviously both senses of the nucleic acid, and the mRNAs coding viral proteins are transcribed by an RNA-dependent RNA polymerase using the negative strand as template. These different classes of RNA viruses show unique features in replication and gene expression in relation to the nature of their genome<sup>9</sup>. However, they are all united by replicating their genomes using a virally encoded RNA-dependent RNA polymerase, with the RNA genome functioning as template for the synthesis of additional RNA strands<sup>10</sup>. It is interesting to note that the error frequency, namely the frequency of incorporating an incorrect base, of RNA-directed RNA replication is quite high compared to that for DNA replication<sup>11,12</sup>. Typically, DNA-directed DNA replication leads to incorporation of one mismatched base per 107 to 109 base pairs, while RNA-directed RNA synthesis typically results in one error per 10<sup>5</sup> or 10<sup>4</sup> nucleotides.

There is another group of RNA viruses, called retroviruses, characterized by including DNA intermediates in their replication cycle, which makes them particularly unique compared to the RNA viruses classes mentioned so far<sup>13</sup>. Retroviruses have a positive-sense ssRNA genome that serve as mRNA coding for viral proteins and enzymes, included the reverse transcriptase (RT). This enzyme possesses the ability to convert genomic viral RNA into cellular DNA, which is then integrated into the host cell chromosomal DNA<sup>14</sup>. The host cell then treats the viral DNA as part of its own genome, transcribing and translating the viral genes along with the cell own genes. In this way, the proteins required to assemble new copies of the virus are produced and the virus replication occurs by the simple process of transcription.

#### **PATHOGENESIS OF RNA VIRUSES**

Studies from the last decades have placed RNA viruses as primary aetiological agents of human emerging pathogens, occupying up to 40% of all emerging infectious diseases<sup>15–17</sup>. RNA viruses are indeed frequently highlighted as the most common class of pathogens behind new human diseases, with a rate of 2 to 3 novel viruses being discovered each year<sup>18</sup>. Moreover, it is believed that these data are underestimated as a consequence of the inadequate surveillance in tropical and subtropical countries, where even established endemic pathogens are often misdiagnosed<sup>18</sup>.

Because of their exceptionally short generation time and their fast evolutionary rate, RNA viruses have great chances to infect new host species. In fact, RNA viruses show remarkable capabilities to adapt to new environments and confront the different selective pressures they encounter. Selective pressures not only include their host immune system and defence mechanisms, but also the continuously evolving antiviral treatments. Their peculiar evolutionary rate arises from their surprisingly high mutation rate<sup>19,20</sup>. Mutation rates of RNA viruses can occur roughly at rates of six orders of magnitude greater than those of their cellular hosts<sup>20</sup>. Moreover, their mutability can surpass the one of DNA viruses by five orders of magnitude<sup>21</sup>, even though it is important to remember that mutation rates can dramatically vary among different viruses within the same taxonomic group<sup>22</sup>. The main reason for the high mutation rate of RNA viruses is the enhanced error frequency observed during the replication cycles, which is caused by the single protein present in all RNA viruses, namely the RNA-dependent polymerase<sup>23</sup>, either the RNA-dependent RNA polymerase or the RNA-dependent DNA polymerase, i.e. the reverse transcriptase. In fact, these enzymes have an inherent higher error frequency than those utilizing DNA as a template and lack of proofreading activity, which increases dramatically the rate of mutation as error correction during the replication process is omitted. RNA viral populations are considered to form quasispecies, which mean basically a swarm of genetic mutants revolving around a consensus sequence, and the enhanced variability appears to be beneficial increasing the probabilities to continue replicating inside the host<sup>24,25</sup>. The RNA viral error rate is at the limit of mutation tolerability, and small increases in this rate generate what is known as mutational meltdown or error catastrophe, in which the viral fitness plummets down, leading to viral extinction<sup>25,26</sup>.

Beside the high mutation rate caused by the lack of proofreading activity, other mechanisms, such as recombination and reassortment, play key roles in RNA viral evolution. Recombination is defined as the synthesis of chimeric RNA molecules from two different progeny genomes<sup>27,28</sup>. It can be intra-genomic when the two segments come from the same origin, namely from the same infecting virus, or inter-genomic when the two segments come from different origins, namely from different viruses infecting the same cell. Reassortment is typical of segmented viruses and implies the mixing of the genetic material, which occurs when segments from different progeny viruses are packaged within a single virion<sup>29</sup>. These phenomena, recombination and reassortment, increase the rate at which beneficial genetic variants are obtained, allowing the emergence of new combinations from previously existent mutants<sup>30</sup>.

# NOTABLE RNA VIRUSES INFECTING HUMANS

More than 200 human-infective RNA virus species have been identified to date and this number keeps increasing, as well as the knowledge in the field, thanks to enormous research effort which counts tens of thousands of published papers per year<sup>31,32</sup> The following table reports some of the most notable RNA viruses infecting humans along with a brief description of the disease they cause.

RNA virus	Description	Ref.
Human immunodeficiency viruses (HIV)	HIV is a retrovirus that infects immune system cells causing the acquired immunodeficiency syndrome (AIDS), which is a condition characterized by the progressive failure of the immune system	
Hepatitis C virus (HCV)	HCV is a positive-sense ssRNA virus that primarily causes hepatitis C, which is a live disease that can lead to liver fibrosis and cirrhosis, liver cancer and general liver failur	
Influenza A virus (IAV)	IAV is a negative-sense ssRNA virus responsible to cause influenza and respiratory diseases	
Respiratory syncytial virus (RSV)	RSV is a negative-sense ssRNA virus that causes infections of the respiratory tract, leading to common colds, bronchiolitis, and sometimes more serious respiratory disorders such as pneumonia	
Severe acute respiratory syndrome coronavirus (SARS-CoV)	SARS-CoV is a positive-sense ssRNA virus that causes the severe acute respiratory syndrome (SARS), which is a respiratory disease characterized by flu-like symptoms, such as fever, muscle pain, lethargy, cough, sore throat, that can also lead to shortness of breath and pneumonia	37
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)	SARS-CoV-2 is a positive-sense ssRNA virus responsible to cause the coronavirus disease 2019 (COVID-19), which is a respiratory and vascular disease characterized by fever, cough, fatigue, breathing difficulties, that can also lead to acute respiratory distress syndrome (ARDS)	
Middle East respiratory syndrome–related coronavirus (MERS-CoV)	MERS-CoV is a positive-sense ssRNA virus responsible to cause the respiratory infection known as the Middle East respiratory syndrome (MERS), characterized by fever, cough, diarrhoea, and shortness of breath	
Ebola virus (EBOV)	EBOV is a negative-sense ssRNA virus responsible to cause the ebola virus disease (EVD), which is a viral haemorrhagic fever with high mortality rate	
Rhinovirus (RV)	RV is a positive-sense ssRNA that causes the common cold	
Poliovirus (PV)	PV is a positive-sense ssRNA virus causative of poliomyelitis, a disease characterized by muscle weakness that can result in flaccid paralysis and inability to move	42
Measles virus (MV)	MV is a negative-sense ssRNA virus responsible for causing measles, typically associated with skin rash, fever and inflamed eyes	43
Rotavirus A (RVA)	RVA is a double-stranded RNA virus, which causes gastroenteritis and diarrhoea	44
Dengue virus (DENV)	DENV is a positive-sense ssRNA virus responsible for causing the dengue fever, characterized by high fever, headache, vomiting, arthralgia, muscle pains, and skin rash	45

Zika virus (ZIKV)	ZIKV is a positive-sense ssRNA virus that causes the zika fever, typically associated with mild symptoms including fever, red eyes, arthralgia, headache, and maculopapular rash		
Yellow fever virus (YFV)	YFV is a positive-sense ssRNA virus responsible for causing yellow fever, which is a disease typically associated with fever, headache, nausea, and muscle pains, but more rarely also liver damage, bleeding and kidney problems	47	
Chikungunya virus (CHIKV)	CHIKV is a positive-sense ssRNA which causes chikungunya, disease characterized by fever, arthralgia, headache, muscle pain, joint swelling, and rash	48	
Rabies virus (RV)	RV is a negative-sense ssRNA virus that causes rabies, which consists in inflammation of the brain that leads to violent movements, uncontrolled excitement, inability to move parts of the body, confusion, loss of consciousness, and often results in death	49	

Table 1. Most notable RNA viruses infecting humans

# CURRENT STRATEGIES IN TREATING RNA VIRUSES

The general principle behind the development of drug treatments against pathogens is to identify targets specifically involved in the pathogen replication so that it can be inhibited without harming the host<sup>50</sup>. The specificity for the pathogen is a crucial aspect that requires careful consideration in the development of effective and safe treatments. Compounds showing inhibiting ability against the pathogen may indeed cause undesirable side effects in the patient leading to harmful and unsuccessful therapies. In addition, while developing therapeutic compounds it is important to search for drug-like properties, such as bioavailability, solubility, permeability, metabolic stability and effective transportation, which are of critical importance for the success of drug candidates<sup>51</sup>. The therapeutic index is a parameter calculated comparing the beneficial therapeutic effect of a given compound and its toxicity and therefore reflects the safety of a drug<sup>52</sup>.

Given the fact that viruses are obligate intracellular parasites, it is easy to understand that the identification of antiviral compounds with good therapeutic index is difficult to achieve. Unlike bacterial cells, which are free-living organisms, viruses utilize the host cell environment for much of their life cycle. Therefore, chemical agents that inhibit both virus and host functions cannot be good choices for therapies. The preferred strategy is to identify viral functions that differ significantly from or are not found within the host and are therefore unique. To achieve this is necessary to study and understand the life cycle of viruses of clinical interest, as all the essential steps are potential sites for antiviral intervention<sup>53</sup>. The following passage is thus to develop compounds able to specifically block these steps, ideally without interfering with host functions, so that the viral infection can be defeated without damaging the patient.

The field of antiviral research has taken on a new dimension since the global spread of human immunodeficiency virus (HIV) caused the acquired immune deficiency syndrome (AIDS) epidemic in the 1980s, with unprecedented efforts in academic and pharmaceutical laboratories to develop new effective antiviral therapies<sup>54</sup>. These efforts have led to remarkable advances in developing innovative strategies against different viruses, with more than 180 antiviral medications approved by the Food and Drug Administration (FDA) in the last 30 years<sup>55,56</sup>. In addition, it is important to mention that many novel antiviral therapeutics is currently in clinical-stage evaluation. However, HIV still remains the most intensively studied and characterized virus, with the largest number of specific antiviral agent approved. In second position there is HCV and, remarkably, HIV and HCV therapeutics combined account for more than two-thirds of all the specific antiviral drugs approved so far<sup>56</sup>.

According to their mechanism of action, antiviral drugs can be classified in different classes, among which the most common are inhibitors of the entrance in the host cell and inhibitors of viral enzymes<sup>57</sup>, even though other viral features offer appealing opportunities, such as disrupting the capsid formation<sup>58</sup> or inhibiting the release from the infected cell<sup>59</sup>.

### Inhibitors of the entrance in the host cell

In the first step of the infection, the virus attaches to specific receptor molecules expressed on the surface of the host cell and subsequently infiltrate the target cell. There are different strategies for entering the host cell, such as through membrane fusion in viruses with the envelope, or through endocytosis, or simply by injecting the viral capsid or genome into the host cytoplasm<sup>60</sup>. An antiviral strategy is to interfere with the interaction between the virus and its binding sites in order to prevent the viral infection upstream. In the case of HIV, the virus expresses on its surface a specific glycoprotein, which is able to interact with the receptor CD4 and the co-receptor (CXCR4 or CCR5) expressed on the surface of the targets cells, in particular CD4<sup>+</sup> T-cells, and in this way can enter<sup>61,62</sup>. HIV antivirals, some of which have been approved while other are still under experimentation, can act in different ways during the attachment and fusion, such as binding and inactivating the viral glycoprotein<sup>63</sup>, or acting as antagonist for the co-receptors CXCR4<sup>64</sup> or CCR5<sup>65</sup>. Similar strategies have been followed for other RNA viruses, such as IAV<sup>66</sup> and RSV<sup>67</sup>, while in other cases, as for HCV<sup>68</sup>, the mechanisms concerning the attachment to the target cells still need to be elucidated.

#### Inhibitors of viral enzymes

Virus encoded enzymes are very attractive targets, because they are not present in uninfected cells, and inhibitors of viral enzymes account for more than two-thirds of all antivirals<sup>56</sup>. Notable targets of these drugs are the viral polymerase, protease and integrase.

#### Polymerase inhibitors

RNA viruses are characterized by the presence of an RNA-dependent polymerase, either an RNAdependent RNA polymerase or a reverse transcriptase in the case of retroviruses, which is required for synthesizing viral nucleic acids and can the target of antiviral drugs. In general, there are two distinct categories of polymerase inhibitors, namely the nucleoside and the non-nucleoside analogs. Nucleoside analogs mimic natural nucleosides and are used as substrates by the viral polymerase causing termination in the synthesis of the viral nucleic acid molecules thanks to their peculiar features<sup>69</sup>. For instance, some nucleoside analogs, called dideoxynucleosides, miss the 3' hydroxyl group (OH) compared to natural nucleosides and thereby cause chain termination in the nascent molecule, as the reaction of elongation cannot take place<sup>70</sup>. Antiviral nucleoside analogs have been successfully developed against HIV, HCV, IAV and RSV, although sometimes selectivity for the viral polymerase is difficult to achieve and thereby these compounds show low tolerability<sup>71–73</sup>. As far as non-nucleoside polymerase inhibitors are concerned, they typically bind to polymerase allosteric pockets distinct from the enzyme active site causing conformational alterations that inhibit the enzyme<sup>74</sup>. Effective drugs belonging to this class have been approved or are under clinical trials as treatments for HIV<sup>75</sup>, HCV<sup>76</sup>, IAV<sup>77</sup>, RSV<sup>78</sup>, and EBOV<sup>79</sup>.

# Protease inhibitors

Viral proteases are important for processing viral proteins so that they can reach their final and functional configuration. Almost all RNA viruses use the strategy of translating a large precursor polyprotein that is further cleaved by viral proteases in a highly regulated manner<sup>80</sup>. This strategy leads to several advantages for the virus, such as having a more compact genome, regulating proteins activity by differential cleavage site usage, and allowing alternative functions for proteins in their precursor forms versus their mature forms. Effective antivirals protease inhibitors have been identified for several RNA viruses, including HIV, HCV, PV, DENV, CHIKV, and SARS-CoV<sup>81–83</sup>.

# Integrase inhibitors

Integrase inhibitors offer an effective strategy against retroviruses, which use the enzyme integrase to integrate their retrotranscribed genome into the host chromosomal DNA<sup>84</sup>. Several integrase inhibitors have been approved as treatments against HIV, while others are currently in clinical trials<sup>85</sup>.

One of the most problematic issues concerning therapies based on viral inhibitor compounds is the extraordinary ability of RNA viruses to develop resistance<sup>86</sup>. For instance, it was shown that HIV acquires significant resistance after a brief exposure to some non-nucleoside reverse transcriptase inhibitors thanks to just one point mutation<sup>87</sup>. Multidrug therapies, also known as highly active antiretroviral therapies (HAART), emerged as a viable strategy to inhibit the generation of viral resistance against HIV<sup>88</sup>, HCV<sup>89</sup>, and IAV<sup>90,91</sup>. They consist in applying more drugs simultaneously, usually 3 to 5, and are based on a probabilistic principle, namely if a virus has a random probability to carry the genetic resistance against one single drug, then its probability to carry several combined resistances should decrease geometrically as the number of drug substances increases in the therapeutic regimen. In the case of HIV, HAART consists in the combination of both nucleoside and non-nucleoside reverse transcriptase inhibitors with protease or integrase inhibitors and is the most common therapy currently used<sup>92</sup>. However, simply increasing the number of drug substances is not always a practical solution, because patients under these treatments can develop severe side effects in the long term, besides the fact that these therapies are quite complicated and expensive<sup>93</sup>. Furthermore, multidrug resistance can also emerge<sup>94</sup>, especially if optimal dosages are missed or in the case of sequential therapies, in which drugs administered sequentially in time<sup>95</sup>. Although some authors argue that the viral genome cannot mutate indefinitely, and that mutational resistance must have a cost in terms of reduced replicative fitness for the virus<sup>96</sup>, virus resistance to drugs is considered an emergent crisis of the last decades<sup>97</sup>.

The extraordinary evolutionary capabilities of RNA viruses have stimulated the development of an entire diversity of pharmaceutical alternative strategies. Fist of all there are vaccines, which are considered the best prophylactic measure against viruses and microbial pathogens in general. Efficient vaccines are available for some RNA viruses, but in many instances and for different reasons, including technologic and economic restrictions, they are scarcely used in the field<sup>98</sup>. In

addition, for several RNA viruses of clinical interest, including HIV, HCV, and RSV, no licensed vaccines exist despite decades of research effort<sup>99,100</sup>. An alternative and attractive antiviral approach consists in RNA interference, which is based on the use of antisense oligonucleotides complementary to specific portions of viral mRNAs so that double-stranded RNA structures are formed leading to viral mRNA silencing and degradation<sup>101,102</sup>. This approach appears particularly promising and powerful, although still needs to be improved and one of the main obstacles is the lack of safe and effective delivery of the interfering RNA molecules<sup>103,104</sup>. In addition, the use of ribozymes, which are RNA molecules with cleavage activity that can be engineered to specifically target RNA molecules of interest, such as viral mRNAs, has also been proposed as an antiviral strategy<sup>105,106</sup>. Another intriguingly antiviral approach is known as lethal mutagenesis and consists in the use of mutagenic nucleotide analogues that have alternate base pairing properties and lead to the induction of mutations<sup>107</sup>. The principle behind is that an excessive mutation rate results to be detrimental and thereby leads to the viral elimination<sup>108</sup>. Other recently developed strategies are based on the use of monoclonal antibodies, which can neutralize the virus acting as entry inhibitors<sup>109</sup> and recruiting and activating the immune system<sup>110,111</sup>. Finally, rather than targeting viruses directly, an interesting antiviral tactic can be to stimulate the host immune system to attack them, which can be done through immunomodulatory substances as interferons<sup>112</sup> and defensins<sup>113</sup>. resulting very effective in some instances.

The advances achieved in antiviral therapies are undoubtedly remarkable and our antiviral arsenal has grown impressively over the last decades, but the biological diversity and rapid adaptive rates of RNA viruses have proven to be difficult to overcome and continuously require the search for new antiviral compounds and strategies. In this regard, it has been recently recognized that viral RNA genomes and transcripts tend to fold forming complex structural elements, which are of critical importance for the viral infection and fitness<sup>114</sup>. Consequently, it has hypothesized to target them with small molecules in order to develop innovative antiviral approaches. The realization of this strategy implies at first instance a deep understanding of the molecular logics that underlie RNA structuring, which correlates with the development of sophisticated tools to study this phenomenon. On the other hand, it is necessary to understand to what extent it is possible to target RNA structures following the logics of protein targeting, and in particular in the context of RNA viruses.

#### **RNA STRUCTURES**

In order to develop a potential antiviral strategy based on targeting viral RNA structural elements, it is first necessary to understand the logics behind the formation of RNA structures. While RNA was firstly identified as the carrier of the genetic information coded by the DNA and necessary to produce proteins, as stated in the central dogma of molecular biology DNA to RNA to proteins, it was soon recognized that it does much more than<sup>115</sup>. RNA is indeed able to cover a surprisingly high number of different biological functions, with the emerging theme that much of RNA functional complexity is rooted in its ability to form intricate and dynamic structures, following the logic according to which specific structures allow interacting with specific molecular components so that a biological function can be explicated<sup>116,117</sup>.

RNA is a biopolymer that consists of ribose nucleotides, namely nitrogenous bases appended to a ribose sugar, attached by phosphodiester bonds forming strands of varying lengths. The nitrogenous bases in RNA are adenine, guanine, cytosine, and uracil. Although RNA is typically a single-stranded molecule, the presence of self-complementary sequences in the RNA strand leads to intrachain base pairing and folding of the ribonucleotide chain into complex structural forms<sup>118</sup>. It is convenient to describe RNA structure in hierarchical terms, comparable to those used in describing protein structure: primary, secondary, tertiary, and quaternary structures.



#### Figure 1. Hierarchy of RNA folding

(a) The primary structure corresponds to the RNA sequence. (b, c) The secondary structure, which in this case corresponds to two stem-loops, forms by Watson-Crick base pairing between complementary nucleotides. (d) Tertiary interactions, in this case the coaxial stacking of double helices, lead to the final three-dimensional structure, which in this case occurs forming a pseudoknot

The **primary structure** simply refers to the nucleotides sequence of the RNA molecule (**Figure 1a**). Some RNAs function as unstructured single-stranded species, such as the messenger RNA (mRNA) that must be unfolded for the genetic message to be translated.

RNA secondary structure is dominated by Watson-Crick (WC) base pairing, often through very long-range interactions, leading to the formation of double-helical structures of varying size, which however seldom exceed 8 to 10 base pairs in length<sup>119,120</sup>. RNA helices show antiparallel righthanded conformation and adopt the A-form structure, which is characterized by the displacement of the bases from the helical axis. Isolated base pairs are not thermodynamically stable, but formation of several consecutive base pairs readily occurs, resulting in a variety of possible arrangements. Interestingly, in many RNAs more than half of all nucleotides are incorporated into helices and G-U pairs are almost as common as the canonical G-C or A-U base pairs, introducing slight distortions in double-helical structure that participate in creating specific surface conformation that can be recognized by binders<sup>121</sup>. The double-stranded helices are interrupted by single-stranded portions, which can form specific loop elements such as hairpins, bulges, and internal loops<sup>122</sup>. The hairpin or stem-loop is the most common and most studied element of RNA secondary structure and is formed when the phosphodiester backbone folds back on itself to form a double-helical tract, namely the stem, leaving unpaired nucleotides to form a single-stranded region, namely the loop<sup>123</sup>. The RNA stem-loop architecture is represented in Figures 1b and 1c. There is a fundamental difference between RNA and protein secondary structures. Protein secondary structure is generally only marginally stable in the absence of stabilizing tertiary structure interactions, whereas RNA secondary structure is often stable on its own. The secondary structure motifs (Figure 2) represent the building blocks through which most complex RNA three-dimensional structures are constructed.



Figure 2. Schematic representation of RNA structural motifs

The main folding motifs are highlighted: three-way junction (green), internal loops (purple and light blue), bulge (red), apical loop (pink), single-stranded region (dark pink), and pseudoknot (gold)

The tertiary structure rises indeed from interactions between two or more secondary structure elements and defines the overall folding of the RNA molecules<sup>124</sup>. Tertiary interactions consist of base stacking, hydrogen bonds, intercalation, base triplet formation, and base pairing between complementary loop sequences. In addition, non-canonical base pairs, unpaired bases, and the backbone functional groups are very important in the context of RNA tertiary folding<sup>125</sup>. In particular, unpaired bases can twist or flip out of a helical patch to define unique surfaces for recognition by other RNA portions during the formation of the tertiary architectures. A pretty common RNA structure motif is the pseudoknot, represented in Figure 1d, which forms when complementary primary sequences of a hairpin or internal loop and a single-stranded region interact with each other by WC base pairing<sup>126</sup>. The formation of a pseudoknot creates an extended helical region through helical stacking of the hairpin double-helical stem and the newly formed loop-loop interaction helix. Although the pseudoknot is only marginally more stable than the two hairpins, tertiary interactions between unpaired nucleotides in the bridging loops and between base pairs within the extended helix can increase the stability of this structure. Interestingly, it is commonly observed that divalent metal ions, especially magnesium ions, are used to screen the negatively charged phosphate groups along the helical backbone in order to build a compactly folded structure with close, which otherwise would not be possible because of electrostatic repulsion<sup>127,128</sup>.

Finally, **quaternary structure** arises from the association of multiple RNA molecules to form supramolecular structures<sup>129</sup>. There are relatively few well-characterized examples of RNA quaternary structure, but these are relatively important. For example, during splicing mRNAs associate with small nuclear ribonucleoproteins (snRNPs) that interact with each other forming RNA-RNA quaternary structures, which are indeed essential for RNA splicing to occur<sup>130</sup>. In most examples characterized thus far, the quaternary association of RNA molecules mainly occurs by conventional WC base pairing, which however can be characterized by particular arrangements, as in the case of the so-called kissing stem loop, which forms between self-complementary loop nucleotides of two different stem-loop structures<sup>131</sup>.

#### HOW TO STUDY RNA STRUCTURES

Understanding the molecular logics of RNA structuring strongly correlates with the ability to study this phenomenon. Consequently, it results of critical importance to develop sophisticated tools to study and predict the structure of RNA molecules. There are alternative approaches to investigate RNA structure, which are continuously improved to give more and more precise outcomes<sup>132,133</sup>. On one hand there are in silico approaches, which rely on the use of specific software and algorithms that apply the theoretical RNA structuring rules with the aim of predicting the actual RNA structure<sup>134–136</sup>. These algorithms receive as input the RNA sequence and, based on thermodynamics, predict the RNA structure considering the interactions that minimize the free energy<sup>137,138</sup>. However, in silico predictions usually fall short of predicting the actual structures, especially for long and complex RNA molecules<sup>139</sup>. This is due to the fact thermodynamic parameters alone are not enough, as the actual RNA folding is affected by many other variables, such as the presence of ions, temperature, pH, interactions with other molecules, and RNA modifications<sup>140,141</sup>. In addition, there is evidence that often RNAs fold to biologically relevant structures that are not the minimum free energy configuration<sup>142,143</sup>. This results in the need to develop experimental approaches aimed at investigating the actual RNA structure, so that our knowledge on the underlying logics could be expanded and applied to improve the accuracy of the predicting algorithms<sup>144,145</sup>. To date, approximately 90% of the WC interactions, typical of secondary structuring, are predicted correctly in the best prediction models, while it results much more complicated to predict the non-WC interactions, which are however essential for determining the three-dimensional fold of an RNA molecule<sup>146</sup>.

Several experimental approaches to investigate the RNA structure have been developed, which can be divided in two main categories, namely biophysical and biochemical approaches. Each category is characterized by peculiar advantages and drawbacks and the different methods can be applied in parallel to reveal various aspects of RNA structuring.

Biophysical methods are based on the classical techniques applied in structural biology research, which include X-ray crystallography (XRC)<sup>147</sup>, nuclear magnetic resonance spectroscopy (NMR)<sup>148</sup>, and cryogenic electron microscopy (cryo-EM)<sup>149</sup>. Following different principles, these techniques offer extremely powerful tools to construct photographs at atomic or near atomic resolution of the studied RNA molecule, although they are typically limited by technical complications<sup>150,151</sup>. For instance, XRC has difficulty resolving the structure of non-compact RNAs or unstructured and flexible RNA regions, lacking therefore suitability for studying conformationally heterogeneous RNAs that exist in multiple functional states, while NMR can only be applied to the study of small RNA motifs and is highly restricted by buffer composition, which implies that it cannot be performed under physiological conditions. In addition, biophysical methods have low-throughput nature, as they imply the study of a single RNA molecule at time.

Biochemical methods rely on the use of enzymatic or chemical structural probes that react differently with RNA in relation to its structural features and differences in reactivity can therefore serve as a footprint of the structure along the sequence<sup>152</sup>. The enzymatic structural probing implies that the investigated RNA molecule is treated with specific nuclease enzymes, which clave RNA into fragments with structure-specific and base-specific patterns<sup>153</sup>. In fact, some RNases are known to specifically cleave RNA in base-paired regions, while others specifically cleave RNA in unstructured regions<sup>154,155</sup>. Combining the results obtained by different RNases treatments, the basepaired and unstructured regions of the RNA can thus be determined<sup>156</sup>. RNA chemical probing is based on the use of specific reagents that modify the RNA functional groups forming covalent adducts on the RNA at the site of reaction<sup>157,158</sup>. The principle behind relies on the fact that the reactivity of the reagents depends on the local accessibility of RNA, with folded and structured regions resulting less accessible. For example, strong hairpins often result in a low-high-low pattern, where low reactivity regions correspond to the stem, and the high reactivity pattern indicates the loop. Numerous chemical probes showing different features can be applied<sup>159,160</sup>. There are base-specific chemical probes, which include dimethylsulfate (DMS) that methylates unpaired adenine and cytosine, 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metha-p-toluene sulfonate (CMCT) that reacts primarily with unpaired uracil and guanine, and diethylpyrocarbonate (DEPC) that reacts specifically with adenine. There are also non-base-specific chemical probes, such as the selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) reagents<sup>161</sup>, which react with the ribose 2'-OH of unpaired nucleotides and are thus able to interrogate all the four different nucleotides at the same time. SHAPE reagents include 1-methyl-7-nitroisatoic anhydride (1M7), N-propanone isatoic anhydride (NPIA), N-methylisatoic anhydride (NMIA), 2methylnicotinic acid imidazolide (NAI), and 2-methyl-3-furoic acid imidazolide (FAI).

After the enzymatic or chemical treatment it is necessary to extract the RNA structural information. This is typically done analysing the cDNA generated by reverse transcription of the investigated RNA molecule. In fact, the reverse transcription reaction is blocked when the enzyme encounters a strand scission (enzymatic probing) or an RNA-chemical adduct (chemical probing), and therefore generates a population of truncated cDNAs whose 3' end correspond either to the nucleotide of the cleavage site or to the nucleotide before the site of chemical modification. Traditionally, the generated cDNAs population was analysed by gel electrophoresis, with labour-intensive and hazardous experimental readouts to map back the RNA structure. These drawbacks were partially overcome analysing cDNAs by capillary electrophoresis, which allowed developing semiautomatically readouts methods based on bioinformatics tools, which were however still characterized by low-throughput outcomes. The field of RNA structure determination has been recently revolutionized by the development of innovative approaches that couple the traditional structural probing techniques with next-generation sequencing (NGS) technologies<sup>162,163</sup>. The application of these methods allows the analysis of several samples within single experiments and guarantee more facile and accurate high-throughput structure prediction, including the low abundance RNA species.

This has led to the possibility of studying the RNA structurome, namely the RNA structures on a genome-wide scale<sup>164</sup>. Following are described some of the most advanced high-throughput methods currently used for determining the RNA structure, which are summarized in the following table.

Technique	Probe class	Probe	Read out	Ref.
PARS	Enzymatic	Nuclease P1	Cleavage	165
Frag-seq	Enzymatic	RNase V1/S1	Cleavage	166
ds/ssRNA-seq	Enzymatic	RNase V1/RNase I	Cleavage	167
Structure-seq	Chemical	DMS	RT-stop	171
DMS-seq	Chemical	DMS	RT-stop	172
Mod-seq	Chemical	DMS	RT-stop	173
CIRS-seq	Chemical	DMS/CMCT	RT-stop	174
SHAPE-Seq	Chemical	SHAPE (1M7)	RT-stop	175, 176
SHAPES	Chemical	SHAPE (NPIA)	RT-stop	177
icSHAPE	Chemical	SHAPE (NAI-N3/FAI-N3)	RT-stop	178
SHAPE-MaP-seq	Chemical	SHAPE (1M7, NMIA, 1M6)	MaP	182
DMS-MaP-seq	Chemical	DMS	MaP	183

 Table 2. Methods for high-throughput RNA structure determination

Despite some differences, these techniques share several core steps. As with the traditional methods, RNA structure is interrogated using enzymatic or chemical probes and for the read-out RNA molecules are reverse transcribed into cDNA so that the probing information is contained within the cDNA fragments. Sequence libraries are then prepared by the addition of two adapters flanking the cDNA and barcodes for sample multiplexing. After sequencing, bioinformatics processing is used to predict the RNA structure. To do this, NGS reads are converted into reactivity values, broadly defined as a measure of the flexibility of a given nucleotide position, and reactivities can then be used to generate RNA structural models that account for the tendency of more reactive nucleotides to be unpaired. A schematic representation of the process is given in **Figure 3**.



**Figure 3. Schematic representation of the bioinformatic processing to predict RNA structures** NGS reads are converted into reactivity values which are then used by specific algorithms that predict the fold of the investigated RNA molecule accounting for the tendency of more reactive nucleotides to be unpaired as high reactivities correspond to flexible nucleotide positions that are not participating in RNA structures

High-throughput methods based on the use of enzymatic probes include parallel analysis of RNA structure (PARS)<sup>165</sup>, fragmentation sequencing (Frag-seq)<sup>166</sup>, and ds/ssRNA-seq<sup>167</sup>, which are similar approaches that differ for the type of nuclease applied and other technical aspects. A major drawback of these strategies is that the use of enzymes as structural probes restricts them to in vitro studies as these enzymes have membrane-impermeant nature. Chemical probes, on the other hand, can diffuse across the cell membrane and thereby probe RNA structures in their native environment<sup>168,169</sup>, besides the fact that these probes allow the interrogation of RNA structures at higher resolution because of their smaller size. The possibility of conducting in vivo analysis is of critical importance as RNA structures often differ in vitro versus in vivo<sup>170</sup>. There are alternative strategies relying on the use of different chemical probes. DMS is applied in Structure-seq<sup>171</sup>, DMS-seq<sup>172</sup>, and Mod-seq<sup>173</sup>, which are similar strategies that consist of slightly different technical passages. Chemical inference of RNA structures (CIRS-seg<sup>174</sup>) combines the use of DMS and CMCT, while SHAPE reagents are used in the techniques as SHAPE-Seq<sup>175,176</sup>, SHAPES<sup>177</sup> and icSHAPE<sup>178</sup>. The association of chemical probing with high-throughput sequencing paved the way for genome-wide in vivo RNA structure probing, even if the analysis of specific low abundance RNA targets remained technically challenging.

The chemical probes based high-throughput methods mentioned so far rely on the identification of reverse transcriptase truncation products. An interesting alternative strategy, called mutational profiling (**MaP-seq**), has been recently developed. This approach takes advantage of RT enzymes that, under specific conditions, induce mutations in the nascent cDNA when they encounter nucleotides modified by the chemical probe, instead of generating truncated cDNAs. The mutational profiling strategy is simpler to implement and more sensitive, allowing rare RNAs to be

effectively examined<sup>179,180</sup>. In addition, Map-seq offers the possibility of distinguishing heterogeneous RNA structure subpopulations from one another, while in truncation approaches the structure signal corresponds to a population average<sup>181</sup>. MaP-seq has been coupled with SHAPE and DMS chemical probes, in methodologies known respectively as **SHAPE-MaP-seq**<sup>182</sup> and **DMS-MaP-seq**<sup>183</sup>. Another benefit of MaP-seq is that it permits the analysis of multiple modified nucleotide positions in a single RNA molecule. Following this principle, the **RING-MaP**<sup>184</sup> methodology allows the direct detection of nucleotide-nucleotide interactions seen as correlated positions of RNA modification.

Adaptations and improvements of the RNA structure investigation strategies allow getting more sophisticated and precise measurements, principally exploiting the flexibility of structural probing experiments in comparative reactivity analysis. For instance, parallel analysis of RNA structure (PARS) with temperature elevation (**PARTE**<sup>185</sup>) involves the study of RNA structures at different temperatures, giving a more detailed information as gradually increasing the temperature structural motifs unfold based on their stability<sup>186</sup>. Alternatively, the concentration of ions, involved in tertiary structures, can be gradually modified while assessing the RNA structure, giving insights on the three-dimensional folding<sup>187,188</sup>. In addition, it is interesting to note that positions on RNA can be protected from enzymes and chemicals or not only by local structure but also by binding proteins or other ligands over that position. This leads to the possibility of using these strategies also to evaluate RNA-binding properties<sup>189,190</sup>.

In conclusion, remarkable advances in RNA structural probing methodologies have reached a point where it is feasible to characterize the RNA structure on a genome-wide scale, although important problems remain despite the tremendous progress of the past years<sup>191</sup>. In particular, it still results challenging to determine higher order tertiary three-dimensional structures, as conventional RNA structural probing experiments provide only one-dimensional information on whether a nucleotide is base-paired, but not the base-pairing partner. In this regard, recently developed multidimensional chemical mapping (MCM<sup>192</sup>) methods supplement one-dimensional information through the systematic perturbation of each nucleotide position by mutation or chemical modification and provide promising tools for modelling tertiary structures, although there is still room for improvement. These methods include the already mentioned RING-MaP, and others such as mutate-and-map (M2-seq<sup>193,194</sup>), multiplexed OH cleavage analysis (MOHCA<sup>195</sup>), MaP-2D<sup>196</sup>, and RNA proximity ligation (RPL<sup>197</sup>). Another problematic issue concerns the computational analysis of the vast and complex RNA structural profiling data generated by NGS, which results quite challenging from the bioinformatics and algorithmic points of view also considering that the validation of the predicted structures may be complicated, as the actual structures of most RNAs remain unknown<sup>198</sup>. SeqFold<sup>199</sup>, RNAstructure<sup>200</sup>, and ViennaRNA package<sup>201</sup> are among the most applied software and algorithms to predict RNA structures using NGS data.

#### VIRAL RNA STRUCTURES

The development and improvement of high-throughput and genome-wide methods for the study of the RNA structurome allowed to investigate its role in the context of RNA viruses, revealing that RNA structure constitutes a full-fledged component of the viral genetic code. The presence of structural elements in the genome and transcripts of RNA viruses has been investigated in a variety of both simplified and complex biologically relevant states. These states include RNA transcribed in vitro and refolded, RNA gently extracted from virus particles or from infected cells, as well as RNA directly in native virus particles or in infected cells<sup>202</sup>. These analyses revealed a wealth of novel RNA structures across coding and non-coding regions linked by relatively unstructured regions. The major challenge remains to identify which of them are involved in viral functioning and could thus be potential targets for small molecules with antiviral purposes. This can be addressed considering the features of the investigated RNA structures. For instance, RNAs with relevant biological function are usually highly structured with architectures associated to low Shannon entropy<sup>203,204</sup>. In addition, these structures are likely to show evolutionary conservation<sup>205-207</sup>, as regions with functional importance are supposed to be preserved. An interesting approach can be to compare the behaviour of RNA structural elements in different biological states or contexts of a virus, such as in packaged virions versus in absence of viral proteins<sup>208,209</sup>. Differences in the reactivity profiles may indicate that particular structural arrangements are involved in specific functions (Figure 4). Another informative approach consists of identifying portions that if mutated disrupt the RNA architectures and reduce the viral fitness<sup>210,211</sup>, which are thus likely to contain functional motifs.



#### Figure 4. Strategy to identify RNA structures involved in virus functioning

Differences in the reactivity profiles of distinct viral biological states can be compared to identify sites of reactivity protections and enhancements, revealing state-specific RNA conformations, which may indicate a correlation with specific functions. In the figure, black, orange, and red bars indicate respectively low, medium, and high nucleotide reactivities.

Efforts in these directions have led to the identification of several viral RNA architectures that participate in different steps of viral life cycle. These include key functions in replication, reverse transcription, transcriptional regulation, viral protein translation, nucleocytoplasmic transport, virion packaging, and evasion of host immune responses<sup>212</sup>. Following are reported some notable functional RNA structures identified for HIV, which remains the most characterized RNA virus in this context, HCV, IAV, ZIKV and coronaviruses SARS-CoV and SARS-CoV-2. Relevant information regarding functional RNA structural elements is available in the literature for the other RNA viruses as well, such for EBOV<sup>213,214</sup>, DENV<sup>215</sup> and YFV<sup>216</sup>.

#### HIV

The HIV RNA structurome remains the most extensively characterized with several RNA structural elements identified to participate in different steps of viral life cycle, including activating transcription, initiating reverse transcription, facilitating genomic dimerization, directing virion packaging, manipulating reading frames, and interacting with viral and host proteins<sup>217–219</sup>. To mention some of the most relevant examples, the 5' leader of the HIV genome is arranged in specific structural elements, which constitute the packaging signal ( $\Psi$ ). These structures are recognized by the viral Gag protein, which guarantees that the viral genome is selectively encapsidated<sup>220</sup>. The RNA adopts a tandem three-way junction structure, in which guanosines essential for both packaging and high-affinity binding to the Gag protein are exposed in helical junctions<sup>221</sup>, as shown in **Figure 5a**. Interestingly, small nucleotide perturbations of the packaging signal sequence cause catastrophic effects on viral infectivity<sup>222</sup>. Another interesting example is the dimerization signal. HIV contains two copies of genomic RNA that are non-covalently linked via interactions between specific sequences located near their 5' ends, which are named dimer initiation sequences (DIS). These elements are structured in hairpin loops and are characterized by the presence of short palindromic sequence enabling intermolecular base-pairing, thus forming kissingloop structures<sup>223</sup>, as shown in Figure 5b. Genome dimerization is crucial for the retroviral life cycle, being involved in the selective packaging of the genome and regulating translation and reverse transcription, besides conferring the great advantage of allowing genetic recombination during reverse transcription, which increases genetic diversity<sup>224</sup>.



#### Figure 5. HIV packaging signal (Ψ) and dimer initiation sequences (DIS)

(a) The packaging signal adopts a tandem three-way junction structure and contains 17 unpaired or weakly paired guanosines (in red) that serve as binding sites for the Gag protein. (b) Location and mechanism of HIV RNA dimerization, which involves the DIS of two homologous stands (represented in red and green)

Another extensively characterized HIV RNA element is the trans-activation response (TAR) element. It is a repetitive RNA structural element located at the 5' UTR, which serves as binding site for the Tat protein that consequently stimulates transcription by a complex mechanism that also involves host cell factors<sup>225</sup>. The Tat-binding site in TAR consists of a conserved RNA stem-loop with a pyrimidine-rich bulge (**Figure 6a**), which is conformationally dynamic but adopts a stable and ordered structure in complex with Tat peptide<sup>226</sup>. In a study it was shown that viruses containing mutations in the TAR RNA structure had dramatically reduced levels of gene expression compared to the wild-type virus<sup>227</sup>. Brief mention should be given also to the HIV Rev response element (RRE), which is a highly structured RNA element (**Figure 6b**) located in the coding region of the viral genome and functions as high-affinity binding site for the viral transcripts from the nucleus to the cytoplasm, where they are translated to produce essential viral proteins or packaged as genomes for new virions<sup>230</sup>.



#### Figure 6. HIV trans-activation response (TAR) and Rev response (RRE) elements

(a) Secondary structure of the TAR element. (b) Predicted secondary structure of a minimal RRE with the major stem loops labeled. Stem IIB is a well-characterized high-affinity site for Rev necessary but not sufficient for RRE function. (c) Schematic representation of how an export-competent Rev-RRE complex might form. Rev molecules assemble onto the RRE scaffold to form an oligomeric assembly

# <u>HCV</u>

The genome of HCV folds into complex structural elements scattered throughout all regions of the genomic RNA, as represented in Figure 7. These structures include highly conserved base-pairings, which are involved in elaborate secondary structures and long-range tertiary interactions. Recent evidence indicates that these RNA architectures represent discrete folded units that contribute directly to numerous aspects of viral lifecycle and infectivity<sup>231</sup>. For instance, the set of stem-loops located within the region encoding NS5B were shown to interact with RNA motifs in the 3' UTR, resulting in a network of RNA elements essential for replication as disrupting these interactions caused impairment in viral replication<sup>232</sup>. In particular, these architectures are believed to direct the RNA polymerase for replication promoting the formation of a particularly robust replication complex<sup>233</sup>. In a study it was shown that the introduction of mutations aimed at disrupting the longrange kissing-loop interaction between SL427 and SL588 abolished viral replication and infectivity<sup>234</sup>. T he same study showed that the disruption of the proper base-pairing of a stem loop within the domain SL1412 lead to the production of non-infectious virus, and it was proposed that this structures serve as an RNA packaging element necessary to interact with capsid proteins during particle assembly<sup>234</sup>. One of the most interesting HCV structural elements investigated to date is a large and complex motif within the domain SL6038. This region of RNA toggles between two distinct structural states: a long stem-loop and a cloverleaf conformation. Genetic and functional analysis suggests that this region switches between different conformations during different phases of the HCV lifecycle, as locking the structure into the stem-loop conformation abolishes viral replication<sup>234</sup>. Interestingly, it was proposed that the overall secondary structural organization of the HCV genome has implications for evasion of the human immune system, as 90% of helices are limited to seven base pairs or less, perhaps to avoid recognition by innate immune sensors that detect double-stranded RNA235.



#### Figure 7. RNA structural elements in the HCV genome

Representation of the secondary and tertiary structural elements of the HCV genome. Labels at the bottom indicate the region of the genome where structures are located. Long-range tertiary interactions are depicted in dark grey. Alternate stem-loop and a cloverleaf structures of SL6038 are depicted

One of the most notable and characterized HCV RNA elements is the internal ribosome entry site (IRES), which is composed by domains II–IV and located at the 5' UTR<sup>236,237</sup>. The IRES element recruits ribosomes directly at the viral start codon and directs translation of the coding sequence without the need for most host cell initiation factors. Although the secondary structure of the IRES has been defined for two decades (**Figure 8a**), only recently the three-dimensional tertiary structure has been revealed thanks to high-resolution structural studies (**Figure 8b**). This helped to understand how IRES recruits and stabilizes the cellular translation machinery<sup>238</sup>. Basically, domains II and III of the IRES form large stem-loops that adopt extended structures, while domain IV forms a short and unstable stem that encompasses the start codon. The flexibility of these domains is critical for function, as they enable the IRES to recruit the 40S ribosomal subunit and initiate translation of viral genes. After the 40S subunit has been recruited, domain II reaches across the head of the 40S subunit, wedging open the mRNA binding tunnel to allow the HCV coding sequence to bind. For this to occur, the weak stem-loop in domain IV must unfold, explaining why stabilizing mutations within this stem-loop are detrimental to HCV translation<sup>239</sup>.



#### Figure 8. Structural features of HCV IRES element

(a) Secondary structure of the HCV IRES element. Domains of the IRES are colour-coded according to the legend. (b) Tertiary structure of the IRES element interacting with the 40S ribosomal subunit

IAV

IAV transcription and replication imply that the RNA-dependent RNA polymerase recognizes a partial duplex structure that forms through circularization of the segments that compose IAV segmented ssRNA genome, which occurs by hybridization between conserved nucleotides at the 5' and 3' end of each segment<sup>240,241</sup>, as shown in **Figure 9a**. The resulting RNA duplex serves as promoter for transcription and replication, playing a crucial for the viral infection. In fact, the disruption of this structure leads to a non-infective form of IAV<sup>242</sup>. **Figure 9b** shows the structure of the IAV RNA promoter bound to the polymerase complex. Intriguingly, the peculiar RNA structures of the IAV genome are believed to direct reassortment between different IAV strains, which recurrently leads to the emergence of new pandemic influenza strains<sup>243</sup>.





(a) Terminal sequence of the IAV genomic RNA that functions as promoter for the RNA polymerase. The conserved 13 nucleotides at the 5' end and 12 nucleotides at the 3' end are boxed. (b) X-ray structure of the IAV promoter bound to the polymerase complex. PA, PB1, and PB2 are the polymerase subunits. The 5' and 3' ends are drawn in cyan and yellow, respectively. The right side shows the general view of the complex, while on the left are reported closer views.

It was recently shown that also IAV mRNAs adopt specific structures in vivo<sup>244</sup>. In the same study, the authors designed mutations aimed at disrupting mRNAs structural domains and obtained viruses with a remarkably attenuated replicative capacity, demonstrating that IAV relies on these structural elements to promote efficient viral replication and infectivity. **Figure 10** shows an example concerning the structural disruption of an IAV mRNA obtained by targeted mutations.



# Figure 10. Targeted disruption of IAV mRNA structural motifs

The example reported refers to the structural disruption of the IAV mRNA segment 4 (HA), nucleotides 1607-1670. The red triangles indicate the sites of the disrupting mutations. On the left site is shown the wild-type structure and on the right the mutant one

# <u>ZIKV</u>

Recent studies investigating RNA structures in the ZIKV genome identified a wealth of functional secondary structural elements as well as specific long-range intramolecular interactions involved in the viral infection<sup>210,245</sup>. For instance, the interaction between the SLA domain in the 5' UTR and the E protein coding region, shown in Figure 11a, was proposed to a play role in virus replication and translation regulation. Interestingly, mutations disrupting this extended RNA interaction resulted in reduced virus infectivity, which was partially rescued by compensatory mutations aimed at restoring the RNA-RNA interaction<sup>210</sup>. A peculiar feature of ZIKV infection, which is a typical signature of flavivirus (ZIKV, DENV, YFV, CHIKV) infections, is the accumulation of abundant virus-derived non-coding RNA named flaviviral subgenomic RNA (sfRNA) in infected cells<sup>246</sup>. This RNA represents a product of incomplete degradation of viral genomic RNA by the cellular nucleases that stall at the conserved highly structured elements of sfRNA. In particular, the threedimensional architecture is mainly directed by two stem-loops that fold into intricate knot-like structures, which are stabilized by the formation of a pseudoknot and other long-range interactions<sup>247</sup>, as shown in Figure 11b. Although most of the underlying mechanisms remain unknown, sfRNAs have been shown to be responsible for viral pathogenicity and host adaptation, and mutations disrupting sfRNAs architecture compromise these functions<sup>248</sup>.



#### Figure 11. ZIKV RNA structural elements

(a) Representation of the long-range interaction between the SLA domain in the 5' UTR and the E protein coding region in the ZIKV genome. (b) Structure of the ZIKV sfRNA. On the left side is shown the secondary structure with stem-loops colour-coded and pseudoknot depicted. On the right side the crystal structure from different perspectives. Shown in orange and blue, the 3' end of the first stem-loop wraps around the 5' end of the RNA, inhibiting unwinding from the 5' end

# SARS-CoV and SARS-CoV-2

Coronaviruses, including SARS-CoV and SARS-CoV-2, bear the largest single-stranded RNA genome in nature, which is characterized by a highly conserved architecture consisting of numerous RNA structural elements (**Figure 12**) involved in the viral life cycle and infection<sup>249–251</sup>.

The 5' UTR region is highly structured and three major conserved stem-loops (named SL1, SL2, and SL4) have been identified in all coronaviruses groups<sup>252</sup>. Besides these, other SLs are present in SARS-CoV and SARS-CoV-2 reaching the total number of eight<sup>253,254</sup>, despite the fact that SL6-SL8 are located downstream of the start codon of ORF1, but nevertheless they are usually regarded as part of the 5' UTR (Figure 12). The SL1 has a functionally bipartite structure. On the one hand, the upper part of the SL1 must be folded and, indeed, destabilizing mutations introduced in the SL1 of mouse hepatitis virus (MHV), a widely used model of coronaviruses including SARS-CoV and SARS-CoV-2, lead to defects in viral replication that can be rescued by compensatory mutations restoring the base-pairing<sup>255</sup>. On the other hand, the lower part of the SL1 needs to be structurally dynamic to establish a transient long-range interaction with the 3' UTR, enabling the synthesis of subgenomic mRNAs (sgRNAs). Accordingly, studies showed that mutations in the lower part of the SL1 lead to defective sgRNAs synthesis<sup>256</sup>. The specific function of SL2 still remains to be determined, but it was shown that mutations disrupting the SL2 stem are not tolerated and lead to impaired sgRNAs synthesis in MHV<sup>257</sup>. The SL4 consists of a long bipartite hairpin (SL4a and SL4b separated by an internal loop) located right downstream of the transcription regulatory sequence at the leader (TRS-L). In MHV, the complete disruption of the SL4 structure, as well as the separate deletion of SL4a or SL4b, is tolerated, while the complete deletion of the SL4 is lethal<sup>258</sup>. These observations suggest that SL4 might function as a structural spacer, determining the proper orientation of the SL1, SL2, and TRS-L, possibly regulating the synthesis of sgRNAs.





Secondary structure of SARS-CoV-2 genome obtained through NMR spectroscopy. The principal RNA elements are highlighted, including the eight stem-loops (SLs) located at 5' UTR, the frameshifting element (FSE) in the coding region, and the BSL and s2m located at 3' UTR. In black are shown the untranslated regions and in orange the coding regions. Stop represents the end of ORF9

Also the 3' UTR region of coronaviruses genome contains functional RNA architectures. The most proximal structural element is the bulged stem-loop (BSL), which was proven to be essential for viral replication in MHV<sup>259,260</sup>. The terminal portion of the BSL has been proposed to mediate the formation of an alternative mutually-exclusive pseudoknotted conformation by direct base-pairing with the loop of a hairpin located down-stream of the BSL<sup>261</sup>, as shown in **Figure 12**. The interconversion between the BSL and the pseudoknot structure has been proposed to act as a molecular switch, regulating viral RNA synthesis and modulating the transition between different steps of the negative strand synthesis<sup>262</sup>. Another highly conserved structure element of the 3' UTR is the stem-loop II-like motif (s<sup>2</sup>m), which has a very particular geometry forming a sharp 90° kink in the helix axis that suggests this structure element might be involved in hijacking the host cell protein synthesis machinery<sup>263,264</sup>.

The most well studied and characterized structural element in coronaviruses genome is probably the ribosomal frameshifting element (FSE)<sup>265,266</sup>. Programmed ribosomal frameshifting is an evolutionary strategy found in many viruses, which allows maximizing the coding content of genomes by providing a mechanism for the translation of overlapping reading frames<sup>267</sup>. Basically, the transition from highly expressed structural proteins to viral enzymes expressed at low levels occurs by a -1 frameshift during translation. Translational frameshifting is triggered by the interplay of two RNA motifs that compose the frameshift signal, which include a slippery sequence where the reading frame change occurs, followed by a structured region that stalls the ribosome. In the case of coronaviruses, the RNA frameshifting element consists of a slippery sequence, followed by a single-stranded spacer and a three-stemmed RNA pseudoknot<sup>268,269</sup>, which is shown in **Figure 13**. The threaded conformation has been proposed to be responsible for the ribosomal frameshifting via a torsional restraint mechanism<sup>270</sup>. Interesting studies showed that alternating the FSE structure in SARS-CoV and SARS-CoV-2 leads to impaired viral functions<sup>271–273</sup>.



Figure 13. SARS-CoV-2 frameshifting element (FSE)

Representation of the secondary structure of the SARS-CoV-2 frameshifting element (FSE), which is characterized by a three-stem pseudoknot motif. The FSE in SARS-CoV differs for a single nucleotide (C instead of A) as highlighted

Besides the well-characterized structures within the 5' and 3' UTRs and the FSE, the coronaviruses genome is predicted to have an exceptionally high propensity to form stable RNA structures, significantly higher compared with that of other RNA viruses, including HCV that is one of the most highly structured viral RNAs to date<sup>274</sup>. The majority of transcription regulatory sequences at the body (TRS-Bs) reside within stem-loop structures, characterized by different degrees of exposure of the TRS core sequence. It is conceivable that these structures regulate the accessibility of the TRS-Bs to finely tune the synthesis of the different sgRNAs and in this way regulating the expression levels of viral proteins. It was shown that coronaviruses genome is more prone to form locally-stable secondary structures than long-range interactions and this preference might be involved in safeguarding the genome stability and ensuring translation fidelity<sup>275</sup>. Nonetheless, the existence of several long-range interactions has also been proven, such as interactions between 5' and 3' UTRs, which have been hypothesized to drive genome circularization<sup>276</sup>. In addition, some studies revealed the presence of RNA G-quadruplexes in coronaviruses genomes<sup>277,278</sup>.

### RNA STRUCTURES AS TARGET OF SMALL MOLECULES

The experimental results reported before indicate that disrupting viral RNA structural elements through mutations lead to impairment of viral functioning and fitness, supporting a potential antiviral therapy based on targeting viral RNA structures. A crucial aspect to consider consists in assessing to what extent it is in general possible to target RNA structures by small-molecules, following the logics of targeting proteins. The key point is whether RNA can form specific binding sites capable of binding small molecules, which can alter its architecture and functionality. Recent evidence suggests that this is the case, encouraging the idea that targeting RNA structures by small therapeutics molecule could be actually developed<sup>279–281</sup>. For example, many RNA riboswitch regulatory elements are known to bind diverse small-molecule metabolites, which alter their function<sup>282</sup>, as well as small-molecule antibiotics are known to bind bacterial ribosomal RNA and interfere with translation<sup>283</sup>. There is currently only one approved drug class that works by binding RNA, which consists in the linezolid antibiotics. They are broad-spectrum antibacterial agents that bind to the large subunit RNA of the bacterial ribosome and interfere with correct tRNA positioning, preventing the bacterial protein synthesis<sup>284</sup>.

Although these encouraging insights, key challenges still need to be addressed in developing RNAtargeting therapies<sup>285</sup>. After the identification of structured RNA involved in biological functions of interest, it is necessary to target them. In order to develop effective and safe treatments, the interaction between the target RNA and binding molecule must be as specific as possible. This implies that the target RNA contains RNA motifs with sufficient structural sophistication to guarantee the achievement of high affinity and specificity binding by small-molecules<sup>286</sup>. This is a crucial aspect that has been largely neglected in the early research in the field of RNA-targeted ligand discovery. In fact, past works identified mostly small-molecule ligands positively charged and with planar configuration capable of intercalating RNA in a nonspecific way, which bind to RNA with high affinity but low selectivity<sup>287,288</sup>. The approaches are however evolving rapidly, and the identification of targets that allow selective bindings in now a priority<sup>285,289</sup>. The structural complexity of a RNA molecule can be counted as the amount of information required to define it, which can be measured in bits<sup>290</sup>. One bit corresponds to the information required to distinguish between two possibilities, for example between purine and pyrimidine bases. The total information content for an RNA motif is given by summing the information content for every position in the target element. To have an idea about the entity of this type of description of the structural complexity, one should consider that 10 bits is equivalent to an expected occurrence of 1 in 1024 nucleotides randomly taken. A structural element such as a short helix requires the information of 9 bits to be described. To be a good quality target, an RNA motif should show a structural complexity of 30 bits, which corresponds to a motif that will occur roughly once in a billion nucleotides and therefore guarantee to be reasonably confident that an RNA target is unique in the human transcriptome<sup>290</sup>.

Interestingly, motifs with a complexity exceeding 30 bits commonly contain high-quality pockets and clefts<sup>291</sup>. These arrangements are simply consequences of having a complex structure and can bind efficiently ligands, such as potential drugs (Figure 14). Therefore, there is a correlation between the structural complexity of an RNA structure and the possibility to target it with high affinity and selective, which makes the investigation and the choice of the target structural motives of remarkable importance and lead to the general preference of higher-order secondary and tertiary structures. It is important to note a potential RNA target not necessarily possesses adequate features to be amenable targets and this can be particularly challenging to assess due to the limitations of the current tools in identifying complex RNA structures with accuracy<sup>292,293</sup>. Besides the identification of an appropriate target, it is necessary to identify small molecules able to interact with it. Screening approaches and discovery strategies can be applied for identifying suitable molecules, including high-throughput screening and the use of molecules libraries, as well as structure-inspired rational design and computational modelling, which however require deep structural knowledge of the target that can be hard to achieve in the case of complex RNA structures<sup>294-296</sup>. Affinity and selectivity for the target cover the main focus, but pharmacological properties and therapeutic index must be considered as well in order to develop safe therapeutic molecules<sup>297</sup>. It is important to note that finding a functional balance between affinity for the target and pharmacological properties can be extremely challenging. The following section reports some screening methods currently used to identify molecules able to interact with RNA, which contribute in uncovering and elucidating the molecular logics underlying RNA-binding.





RNA structural motifs, including helical regions, hairpin loops, bulges, multi-branched loops, internal loops, and long-range interactions like pseudoknots, fold in three-dimensional architectures creating potential pockets and clefts that enable the binding of small molecules. These arrangements are consequences of having complex structures, which leads to the preference in the target choice for high-order secondary and tertiary structures in order to reach selective bindings

### **IDENTIFYING SMALL MOLECULES BINDING RNA**

Discovering compounds that target and modulate RNA still remains challenging, although over the past years innovative approaches have significantly improved our chemical knowledge about RNA binders<sup>298,299</sup>. In particular, many methods typically used for studying protein-binding compounds have been adapted for RNA, while others have been specifically developed exploiting the peculiar RNA features. Following are reported some of the most notable ones.

### Fluorescence-based methods

Many methods for studying and assessing small molecules binding RNA involve fluorescent labelling of the RNA molecule. RNA can be labelled at its terminus with specific probes or internally with modified nucleotides and the fluorescence is then used to monitor physical changes that occur as consequences of binding<sup>300</sup>. For instance, the probe fluorescence emission can increase or decrease as result of the conformational change due to the RNA-ligand interaction<sup>301,302</sup>, as shown in Figure 15a. Other widely used methods for identifying small-molecule binding to RNA rely on conjugating RNA with a donor fluorochrome and an acceptor fluorochrome to form a donoracceptor pair<sup>303</sup>. The transfer of energy between the two fluorochromes occurs only if they are in close proximity, which can result from the conformational changes upon a ligand binding (Figure 15c). Alternatively, other methods imply the presence of a quencher, which possesses the ability of quench the emission of the fluorescent probe<sup>304</sup>. The conformational changes related to the binding of a ligand displace the quencher and fluorescence emission can be detected (Figure 15b). Also, the alterations in molecular rotation through fluorescence polarization (FP) can be applied to investigate RNA-ligand interactions<sup>305</sup>. Basically, the sample containing RNA and small molecules is invested by plane polarized light and the interactions can be detected because the binding influences the molecular rotation of the species and cause retention of polarization (Figure 15d). Other physical changes that occur during binding and that can thus be considered for investigating RNA binding molecules include changes in molecular movement in a thermal gradient (microscale thermophoresis<sup>306</sup>) and alterations of RNA thermal stability (differential scanning fluorimetry<sup>307</sup>).





#### Figure 15. Fluorescence-based methods to identify small molecules binding RNA

(a) A single dye can probe local changes in the RNA structure induced by ligand binding and as result experience an increase or decrease in the fluorescence emission. (b) Quenching-based detection methods rely on the fact that conformational changes caused by the binding of a ligand displace the quencher and fluorescence emission can be detected. (c) The donor-acceptor pair allows measuring ligand binding as the two fluorochromes are in close proximity as consequence of the conformational changes that occur upon ligand binding. (d) Fluorescence polarization (FP) assays use plane polarized light to measure the molecular rotation of the fluorescently labelled sample. If the fluorescent binder is not bound to the dye, then its rapid molecular rotation results in depolarization. If the molecule does bind, then polarization is retained and that change is measured using a polarimeter

#### Small-molecule microarrays

A powerful approach to identify RNA-binding ligands consists of small-molecule microarrays (SMM). This method involves the printing of small molecules on specific surfaces, which usually are included in wells of plates<sup>308</sup>. Labelled RNAs are then hybridized to the printed molecules, while unbound species are washed away. Hits are identified by accumulation of the label in correspondence of spots where the printed molecule and RNA interact, as shown in **Figure 16**, and are then further investigated revealing features concerning the small molecule-RNA interaction. Interestingly, this method can include libraries of small molecules compounds and libraries of RNAs and thus allow the assessment of millions of binding interactions simultaneously<sup>309</sup>.



#### Figure 16. Small-molecule microarrays (SMM) to identify small molecules binding RNA

Small molecules can be directly absorbed onto the surface of agarose-coated microarrays, which are then hybridized with a radiolabelled RNA library. Hits are identified by signal accumulation in the correspondent spot and further investigated

This method was recently applied in a screening approach named two-dimensional combinatorial screening (2DCS) whereby a library of array-immobilized small molecules was incubated with libraries of RNA secondary structures motifs<sup>310</sup>. The most avid interactions were achieved including excess competitor oligonucleotides. The bound RNA was then excised and sequenced, building an encyclopedia of privileged RNA secondary structure motif-small molecule interactions, which was then included in the online database Inforna<sup>311</sup>. Following this approach it is possible to define a chemical code for recognizing RNAs with small molecules, which can inform drug design.

For instance, in the study molecules composed of the scaffold structure shown in **Figure 17** were identified to avidly bind 5-nucleotide hairpins with preferential sequence composition.



#### Figure 17. Scaffold structure identified to bind 5-nucleotide hairpins

The scaffold structure highlighted in the blue boxes was identified to avidly bind 5-nucleotide RNA hairpins with the respective preferential sequence composition indicated beneath

# Surface plasmon resonance

Surface plasmon resonance (SPR) is a technique typically used for studying protein-binding, which has been recently adapted for RNA. It consists on hitting with polarized light a metal surface at the interface of materials with different refractive indices. At the appropriate angle of incidence, surface electrons of the metal become excited and oscillate, absorbing some light while reflecting the rest. If one of the surface compositions is changed, its refractive index also changes, thereby shifting the resonance angle required to absorb light<sup>312</sup>. To apply this principle for investigating RNA-ligand interactions, RNA of interest is conjugated to the metal and the binding of molecules is detected as resonance shifting<sup>313,314</sup>.

#### Mass spectrometry-based approaches

Mass spectrometry (MS) offers an effective tool to study small molecules-RNA interactions, as it allows identifying complexes identity with high precision based on their mass. Dated approaches focused on direct MS detection of RNA-small molecule complexes<sup>315</sup>, while more sophisticated methods have evolved to apply the MS principle in a wider range of contexts<sup>316</sup>. This is the case, for instance, of the tool named automated ligand detection system (ALIS), which is an affinity-selection mass spectrometry platform for the high-throughput screening of small molecules binding to macromolecules<sup>317,318</sup>. This approach, schematized in **Figure 18**, implies that RNA is incubated with mass-encoded small-molecule mixtures, then the complexes are purified by size exclusion chromatography (SEC), and subsequently dissociated through reverse-phase chromatography. Finally, the bound ligands are detected by high-resolution MS. ALIS has been recently applied to screen compounds libraries in binding RNAs allowing to identify selective binders<sup>319</sup>. In addition, in the same study cheminformatics evaluation of the ALIS generated data provided insight into the physicochemical properties and chemical substructures that direct general and specific RNA binding.



**Figure 18.** Schematic representation of the automated ligand detection system (ALIS) procedure ALIS involves equilibration of the RNA target with a mixture of compounds (1), SEC purifying the RNAsmall molecule complexes (2), reverse-phase HPLC dissociating the RNA-small molecule complex (3), and MS that identifies small-molecule hits (4)

#### NMR and XRC-based approaches

Nuclear magnetic resonance (NMR) is a sensitive technique suitable for finding RNA binders, as it allows to identify the interaction by monitoring changes in the spectra of the RNA target upon addition of the ligand, or in the spectra of the ligand upon addition of the RNA target<sup>320,321</sup>. NMR based methods can be used to identify binding sites and dynamics and to determine binding affinities. X-ray crystallography (XRC) is commonly used to rapidly generate high-resolution structures of RNA-ligand complexes, but it implies some technical issues often difficult to overcome, such as the fact that crystal contacts may block ligand-binding sites<sup>322,323</sup>. However, it provides a powerful tool in many instances.

# Virtual Screening

An approach that can be used in parallel to experimental procedures for the study of RNA-ligand interaction is the virtual screening<sup>324,325</sup>. This approach allows sampling the chemical space on the RNA target 3D structure and predict the docking of large virtual libraries of compounds based on computer modelling. Molecular docking estimates the binding mode as well as the binding affinity. Although the numerous advantages in terms of saving time and costs related to in silico approaches, challenges with modelling the three-dimensional structure and conformational flexibility of RNA targets limit the virtual screening success, especially considering that RNA can undergo large conformational changes upon ligand binding<sup>326</sup>. In addition, still relatively few experimental RNA-ligand structures are available to develop scoring functions calculating the free energy of ligand binding<sup>305</sup>. Nevertheless, docking algorithms are continuously improved and refined to overcome these issues, including more parameters describing the interacting behaviour of flexible structures and considering experimental observations. Some of the most reliable docking algorithms currently available are MORDOR<sup>326</sup>, DOCK<sup>327</sup>, DrugScoreRNA<sup>328</sup>, LigandRNA<sup>329</sup>, and RiboDOCK<sup>330</sup>.

The previously described experimental procedures aimed at identifying small molecules binding RNA usually imply the use of compounds libraries to identify novel ligands. The screening of large compound libraries can result technically problematic against challenging targets such as some RNAs<sup>331</sup>. Fragment-based drug discovery can be applied to overcome this issue, which consists of screening smaller sets of fragments for chemical starting points<sup>332,333</sup>. Fragments are defined as having less than 20 non-hydrogen atoms and allow for efficient sampling of the chemical space. More complex compounds can then be built starting from the identified small compounds and screened to obtain the most suitable ones. Another interesting device in the context of screening compounds libraries rely on the use of dynamic combinatorial libraries generated through dynamic combinatorial chemistry<sup>334,335</sup>. This concept hinges on the use of ligand binding fragments that can undergo reversible chemistry to form more complex binders and whose reaction equilibrium shifts in response to the interaction with the RNA target. In other words, upon binding, the ligand drives the reaction equilibrium to make more of the high-affinity species and depletes low-affinity species from the pool. This approach has been shown to be really effective in identifying selective RNA binding molecules. Furthermore, as the final aim is to found small molecules targeting RNA with suitable pharmaceutical properties, it may be appropriate to focus on drug-like molecules libraries for the screenings<sup>297,336</sup>.

In order to catalog the continuously growing knowledge about RNA-ligand interactions, several online searchable databases have been developed, such as the already mentioned Inforna, SMMRNA<sup>337</sup>, R-BIND<sup>338</sup>, G4LDB<sup>339</sup>, and NALDB<sup>340</sup>. They include the available small molecules known to bind RNA as well as the screening approaches applied to identify them and offer an extremely powerful tool in the RNA drug discovery research, providing insights about the molecular logics underling RNA binding.

### ADVANCES IN TARGETING VIRAL RNA STRUCTURES

Although the field of targeting RNA structural elements with antiviral purposes can be still considered on early development, some studies explored the inhibition of viral RNA function by small molecule ligands. HIV and HCV remain the most extensively studied and characterized in this context, but notable evidences are emerging also for other viral pathogens such as IAV and coronaviruses SARS-CoV and SARS-CoV-2<sup>341</sup>.

# <u>HIV</u>

Several HIV RNA regulatory elements, including the previously described trans-activation response (TAR), Rev response (RRE), dimer initiation sequence (DIS), and packaging signal ( $\Psi$ ), have been investigated for small molecules binding. The TAR element serves as binding site for the Tat protein, which consequently stimulates viral transcription, and it was early recognized that the disruption of the Tat/TAR complex by competing RNA ligands blocks the viral replication<sup>342,343</sup>. Therefore, numerous studies have focused on the discovery and design of TAR RNA-binding inhibitors following different approaches, making the TAR element one of the most characterized example in the field<sup>344–349</sup>. For instance, the knowledge about the molecular determinants allowing Tat to bind the TAR element were applied following a structure-based design methodology to design TAR-binding compounds, which were then confirmed through NMR<sup>350</sup>. Figure 19a shows the TAR element in complex with a synthetic small molecule ligand designed with this strategy. In a more recent study, a small molecule microarray screening of TAR hairpin RNA has been used to identify selective ligands from a library of more than 20000 drug-like molecules<sup>351</sup>. The thienopyridine derivative compound shown in Figure 19b was identified following this approach and its antiviral activity was positively confirmed in cell-based assays. Other recent studies identified promising TAR-binding compounds following different strategies, such as fragmentbased screenings<sup>352</sup>, fluorescent-based screenings<sup>353</sup> and virtual screenings<sup>354</sup>, and the antiviral activity of some of them was positively confirmed in vitro assays.



#### Figure 19. The HIV TAR element as RNA target

(a) Model of TAR element in complex with a synthetic small molecule ligand, shown in yellow, identified following a structure-based design methodology. (b) Representation of the TAR selective ligand thienopyridine derivative identified through a small molecule microarray screening

The RRE element, which serves as binding site for the Rev protein required for the nucleocytoplasmic export of HIV transcripts, has been also largely explored in the context of small molecules inhibitors<sup>355–358</sup>. A study identified compounds able to disrupt the Rev–RRE interaction in a fluorescent-based binding competition screen of 1120 FDA-approved drugs and their antiviral activity was confirmed in cell-based assays<sup>359</sup>. Further NMR spectroscopic analyses revealed that these compounds interact with the RRE stem-loop IIB (SL-IIB), which overlaps with the Rev binding region. **Figure 20** shows the complex between RRE SL-IIB and clomiphene, which is one of the most promising compound identified in this study. In another interesting study, surface plasmon resonance (SPR) was applied to identify compounds that bind RRE with submicromolar affinities<sup>313</sup>.



#### Figure 20. The HIV RRE as RNA target

Best-scored model of a complex between the RRE internal loop IIB (gray) and clomiphene (green carbon atoms), built by docking the ligand into the structure of subdomain IIB

Other interesting studies aimed at targeting HIV RNA structural elements applied different strategies, from libraries screenings to rational design approaches, to identify small RNA-binding molecules interfering with the recognition of the packaging signal ( $\Psi$ ) by the viral Gag protein<sup>360,361</sup>, necessary for the viral genome to be encapsidated, and disrupting the specific structure of the dimer initiation sequences (DIS)<sup>362,363</sup>, which is crucial for several HIV functions. The antiviral activity was also positively confirmed in vitro for some of these compounds.



**Figure 21** shows the kissing loop dimer of the dimer initiation sequence (DIS) in complex with an inhibitory compound.

**Figure 21. The HIV DIS as RNA target** Model of the kissing loop dimer of the dimer initiation sequence (DIS) in complex with a small inhibitory ligand shown in yellow

#### <u>HCV</u>

The most appealing and explored HCV RNA target is the internal ribosome entry site (IRES), which recruits ribosomes directly at the viral start codon so that the HCV genome can be translated<sup>364–370</sup>. Early studies screened molecules libraries for inhibition of IRES function in cell-based reporter translation assays and led to the discovery of some compound with moderate selectivity for the viral target<sup>371,372</sup>. More recently, a mass spectrometry-based screening of 180,000 compounds identified aminobenzimidazole derivatives as ligands binding the internal loop RNA of subdomain IIa of the IRES element<sup>373</sup>. In the same study, these compounds were shown to inhibit HCV in cell culture assays without causing cytotoxicity. Mechanism of action studies of aminobenzimidazole derivatives revealed that these compounds act as allosteric inhibitors of an RNA conformational switch in the IRES subdomain IIa and capture this subdomain in an extended conformation which blocks viral translation initiation<sup>374</sup>. X-ray crystallography has provided structural insight into the conformational states of the subdomain IIa switch in the absence<sup>375</sup> and presence<sup>376</sup> of inhibitor ligands, showing that in the extended conformation of the switch the bound inhibitor is encapsulated by a deep solvent-excluded RNA pocket as shown in Figure 22. Depth structural complexity and physicochemical properties of the ligand-binding pocket make the IRES subdomain Ha an ideal target for drug-like inhibitors development. Different studies applied fluorescent-based methods to investigate ligands obtained optimizing the previously identified aminobenzimidazole derivatives and recognized promising compounds with anti-HCV activity shown in cell culture assays<sup>377,378</sup>. Other studies identified interesting compounds following rational design approaches focused on the physical and chemical properties of the binding pocket of the subdomain IIa<sup>379,380</sup>. In addition, brief mention should be made of copper-binding metallopeptides that have been developed to target domains of the HCV IRES and damage the RNA through metal catalyzed cleavage<sup>381–383</sup>. These metallopeptides have been reported to effectively inhibit HCV in cell culture assays, although they hardly possess drug-like properties because of their large sizes.





(a) Representation of the IRES subdomain IIa in the absence of inhibitors. (b) Crystal structure of the subdomain IIa in complex with an aminobenzimidazole derivative IRES inhibitor. The ligand is shown in yellow, while positions of two magnesium ions are indicated by green spheres. (c) Target recognition by hydrogen bonging and partial intercalation of an aminobenzimidazole derivative in the subdomain IIa

IAV

The most appealing structural RNA element in IAV is the partial duplex that forms through genome circularization and functions as promoter recognized by the viral RNA-dependent RNA polymerase. In an earlier study, a fluorescent-based competition assay was used to identify compounds binding to the IAV promoter, but the impact of ligand binding on the promoter function was not investigated<sup>384</sup>. More recently, a study applied an NMR-based fragment binding screen on more than 4000 compounds (**Figure 23a**) identifying among the hit compounds the drug-like amino-quinazoline<sup>385</sup>, which is represented in **Figures 23b** and **23c** in complex with the IAV promoter. In the same study, a cell-based assay demonstrated the antiviral activity of the RNA promoter-targeting amino-quinazoline against different strains of IAV and no cytotoxicity was detected. In a follow-up study, amino-quinazoline analogs and derivatives were synthesized and tested showing improved binding affinity to the target and antiviral activity<sup>386</sup>. However, the investigators noted that direct inhibition of the viral RNA polymerase might have contributed to the antiviral activity of these compounds.





(a) Schematic representation of the procedure followed in the NMR-based fragment binding screen, which identified the compound amino-quinazoline as IAV promoter inhibitor. (b) Model of the ligand-target complex derived by NMR spectroscopy of the IAV promoter bound to amino-quinazoline (shown in yellow). (c) Detail view of the ligand binding site in the NMR model

# SARS-CoV and SARS-CoV-2

The coronaviruses translational frameshifting element (FSE), which is the most studied and characterized RNA structure in the context of coronaviruses, has been explored for small molecules inhibitors binding. An interesting study used an in silico docking screening approach to identify SARS-CoV FSE ligands and revealed a diazepane derivative among the hit compounds, which was confirmed as an inhibitor of translational frameshifting in cell-based assays<sup>387</sup>. **Figure 24** shows the proposed interaction modelling between the SARS-CoV FSE pseudoknotted structure and the diazepane derivative identified in the study, while the direct interaction was demonstrated measuring the ligand binding by surface plasmon resonance (SPR)<sup>388</sup>. In the same study, mechanistic analysis revealed that ligand binding reduces the conformational plasticity of the FSE pseudoknot RNA fold and thereby affects ribosomal frameshifting. In addition, it was proposed that enhanced frameshifting inhibitor ability of the diazepane derivative might be due to interaction with multiple targets besides FSE.



**Figure 24. The SARS-CoV frameshifting element (FSE) as RNA target** Proposed binding pose of the diazepane derivative in the active site of SARS-CoV pseudoknotted structure generated by computational docking

More recently, efforts have been made to target the SARS-CoV-2 frameshifting element. Recent evidence has led to revise the FSE structural model indicating that besides the slippery sequence and the three-stemmed pseudoknot, it comprises an attenuator hairpin (AH), which can be targeted in order to impair frameshifting efficiency<sup>389,390</sup>. An interesting study applied a small molecule microarray screening to identify selective binders to the SARS-CoV-2 FSE attenuator hairpin<sup>391</sup>. A promising drug-like small molecule was identified among the hit compounds, which avidly bound to the FSE attenuator hairpin structure stabilizing the hairpin folded state. In the same study, the identified molecule was shown to reduce the frameshifting efficiency in cell-based assays. Intriguingly, the ligand was further elaborated by equipping with an RNase recruiting molecule in order to be able to recruit a cellular ribonuclease to destroy the viral genome, thus forming a so-called ribonuclease targeting chimera (RIBOTAC)<sup>392,393</sup>. The RIBOTAC optimization strategy improved the bioactivity of the compound of 10-fold and a schematic representation of the working mechanism is given in **Figure 25**.



SARS-CoV-2 AH

#### Figure 25. Working mechanism of RIBOTAC degrading SARS-CoV-2 RNA

The RIBOTAC strategy implies that the SARS-CoV-2 AH binding molecule (shown in blue) is equipped with an RNase recruiting molecule (shown in green). Upon the binding, an RNase is recruited and gets activated by dimerizing, so that the SARS-CoV-2 RNA is cleaved

Besides the frameshifting element, other SARS-CoV-2 RNA structural elements have been proposed to be the target of small molecules with antiviral purposes<sup>394</sup>. A study screened a small molecules library identifying amiloride derivative compounds as inhibitors of SARS-CoV-2 replication<sup>395</sup>. Further NMR-based analysis revealed that these compounds specifically interact with structural elements located in the 5' UTR, in particular with stem loops 4, 5 and 6. Another study focused on targeting a specific RNA G-quadruplex located in the coding sequence region of SARS-CoV-2<sup>396</sup>. The authors identified compounds that inhibit viral translation both in vitro and in vivo by acting as stabilizer of the G-quadruplex structure.

The examples here reported (namely compounds targeting HIV TAR, RRE, packaging signal and DIS, HCV IRES, IAV promoter, and SARS-CoV/SARS-CoV-2 FSE) represent some of the most advanced and promising RNA-targeting antiviral molecules identified so far. However, besides the remarkable amount of effort spent in developing them, some of their features still need improvement before hypothesizing to apply them in therapies. In particular, the most challenging aspect remains to find a functional compromise between high specificity for the target and pharmaceutical properties. Among the mentioned examples, the HCV IRES element possesses the most desirable physicochemical properties pursued in an RNA target, which make the related binding molecules the most promising compounds currently available.

#### **DISCUSSION AND CONCLUSION**

RNA viruses are among the most relevant human pathogens, counting an impressive high number of different species, which can cause more or less severe disorders and diseases. Given the fact that they are obligate intracellular parasites that utilize the cellular components to replicate and propagate, it results challenging to identify specific pharmaceutical targets for defeating them without harming the host. Impressive research efforts over the past 30 years have led to the development of different effective antiviral strategies which focus on the disruption of crucial steps of the viral life cycle. In particular, our current antiviral arsenal mainly relies on inhibitors of viral entrance and of viral enzymes, often combining the use of multiple compounds simultaneously. Alternative approaches have also been developed focusing on the use of antisense oligonucleotides, monoclonal antibody, the stimulation of the host immune system and vaccination, which remains the most efficient preventive approach notwithstanding that it is not available for many relevant RNA viruses. Nevertheless, it is importantly to note that RNA viruses are characterized by extraordinary evolutionary and adaptive rates, which make even more challenging to counteract them as they rapidly develop resistance. This implies the need of a continuous research for new antiviral compounds and strategies.

In this regard, it has been recently recognized that viral RNA genomes and transcripts fold into complex and intricate structural elements involved in the viral infection and fitness and RNA structures are considered a full-fledged component of the viral genetic code. Several RNA architectures have been identified in different RNA viruses and have been shown to participate in different steps of the viral life cycle, including replication, reverse transcription, transcriptional regulation, viral protein translation, nucleocytoplasmic transport, virion packaging, and evasion of host immune responses. An emerging theme is that much of the RNA functional complexity is rooted in its ability to form intricate and dynamic structures, which allow interacting with other molecular components to elicit specific biological functions. Interestingly, mutations aimed at disrupting these RNA structures lead to impairment of virus functioning, and consequently it has hypothesized to target them with small molecules in order to develop innovative antiviral approaches. The realization of this strategy implies at first instance a deep understanding of the molecular logics that underlie RNA structuring, which correlates with the development of sophisticated tools to study this phenomenon. Remarkable advances have been reached in the field of RNA structure determination over the past years thanks to the development of innovative approaches that couple the traditional structural probing techniques with NGS technologies, reaching a point where it is feasible to characterize the RNA structure on a high-throughput and genome-wide scale. Nevertheless, there is still room for improvement and in particular it still remains challenging to determine RNA high-order tertiary and three-dimensional structures. On the other hand, it is necessary to understand to what extent it is possible to target RNA structures with small molecules following the logics of protein targeting. Although the recent evidence are promising and suggest that therapies based on targeting RNA could be actually developed, many

key challenges still need to be addressed. Insights in the field of targeting RNA indicate that it is crucially important to select RNA targets containing motives with sufficient structural complexity and sophistication, possibly with pockets and clefts, so that can bind ligands with enough affinity and selectivity. It is important to note that RNAs explicating biologically relevant roles, as RNAs involved in virus functioning, not necessarily possesses adequate features to be amenable targets and this can be particularly challenging to assess due to the limitations of the current tools in identifying complex RNA structures with accuracy. Another challenging point is the discovering of compounds that target and modulate RNA. Despite the fact that over the past years innovative approaches have significantly improved our chemical knowledge about RNA binding, efforts are still required to improve our ability to manipulate RNAs with small molecules. Although the field of targeting RNA structural elements with antiviral purposes can be still considered on early development, some studies explored the viral inhibition through RNA targeting and provided promising results in vitro studies. Nevertheless, these studies indicated that finding a functional balance between affinity for the target and pharmacological properties can be extremely challenging, which is however necessary to achieve in order to develop effective and safe therapies. Interestingly, most of these problematic issues are simply connected to technical limitations. The notable progresses and advances of the past years bode that currently technical obstacles in determining complex RNA structures and in identifying small molecules RNA binding possessing desired features will be soon overcome. In addition, predicting algorithms and software for computational modelling are constantly improving and their use can dramatically speed up experimental processes.

Targeting RNA structural elements in the genomes and transcripts of viral pathogens offers an extraordinary opportunity to expand our repertoire of antiviral drug targets and to open the doors to a completely innovative antiviral strategy, which is of critical importance considering the worrying development of resistance against the currently available antivirals. One of the strongest and most promising point of RNA-targeting antiviral therapies relies on the significant observation that the RNA structures involved in viral infections are characterized by high structural complexity, which is indeed necessary to explicate their function, and at the same times gives them the features to be good targets for small binding molecules. Furthermore, these structures are evolutionary conserved because of their crucial roles, which promises a high barrier to resistance development.

In my opinion, RNA-targeting therapies will be a concrete reality in the near future and not only in the antiviral context. Although it will probably take still a consistent amount on time, mainly in relation to the need of overcoming the current technical limitations, targeting RNA with small molecules will remarkably improve our therapeutic possibilities and the deriving benefits will be notable. I believe for these reasons that studying and understanding RNA biology and in particular the relation between RNA structure and function should be considered of essential interest. In the context of RNA viruses, the investigation of the function of RNA elements should be a priority, as it can reveal new potential targets and the direction for intervention. The promising early results in targeting viral RNA structures should be taken as encouragement to invest serious research efforts in the field given its enormous potential.

# REFERENCES

- 1. Flint, J., Racaniello, V. R., Rall, G. F., Skalka, A. M. & Hatziioannou, T. *Principles of virology, Volume 1: Molecular biology*. (John Wiley & Sons, 2020).
- 2. Lodish, H. *et al.* Viruses: Structure, function, and uses. in *Molecular Cell Biology. 4th edition* (WH Freeman, 2000).
- 3. Ryu, W.-S. Chapter 3 Virus Life Cycle. in (ed. Ryu, W.-S. B. T.-M. V. of H. P. V.) 31–45 (Academic Press, 2017). doi:https://doi.org/10.1016/B978-0-12-800838-6.00003-5.
- 4. Sic, B. A. & Rology, V. I. Basic Virology by Wagner.
- 5. Woolhouse, M., Scott, F., Hudson, Z., Howey, R. & Chase-Topping, M. Human viruses: Discovery and emeraence. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2864–2871 (2012).
- 6. Lefkowitz, E. J. *et al.* Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717 (2018).
- 7. Holmes, E. C. *The evolution and emergence of RNA viruses*. (Oxford University Press, 2009).
- 8. Nguyen, M. & Haenni, A. L. Expression strategies of ambisense viruses. *Virus Res.* **93**, 141–150 (2003).
- 9. Tao, Y. J. & Ye, Q. RNA virus replication complexes. *PLoS Pathog.* 6, 1–3 (2010).
- 10. Payne, S. Chapter 10 Introduction to RNA Viruses. in (ed. Payne, S. B. T.-V.) 97–105 (Academic Press, 2017). doi:https://doi.org/10.1016/B978-0-12-803109-4.00010-6.
- 11. Venkataraman, S., Prasad, B. V. L. S. & Selvarajan, R. RNA dependent RNA polymerases: Insights from structure, function and evolution. *Viruses* **10**, 1–23 (2018).
- 12. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral Mutation Rates. *J. Virol.* **84**, 9733–9748 (2010).
- 13. Weiss, R. A. Retrovirus classification and cell interactions. *J. Antimicrob. Chemother.* **37**, 1–11 (1996).
- 14. Coffin, J. M. & Fan, H. The Discovery of Reverse Transcriptase. *Annu. Rev. Virol.* **3**, 29–51 (2016).
- 15. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
- 16. Woolhouse, M. E. J. & Gowtage-Sequeria, S. Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **11**, 1842–1847 (2005).
- 17. Morens, D. M., Folkers, G. K. & Fauci, A. S. The challenge of emerging and re-emerging infectious diseases. *Nature* **430**, 242–249 (2004).
- 18. Rosenberg, R. Detecting the emergence of novel, zoonotic viruses pathogenic to humans. *Cell. Mol. Life Sci.* **72**, 1115–1125 (2015).
- 19. Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **73**, 4433–4448 (2016).
- Holmes, E. C. The Evolutionary Genetics of Emerging Viruses. *Annu. Rev. Ecol. Evol. Syst.* 40, 353–372 (2009).
- 21. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276 (2008).
- 22. Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* 73, 4433–4448 (2016).
- Jácome, R., Becerra, A., Ponce de León, S. & Lazcano, A. Structural Analysis of Monomeric RNA-Dependent Polymerases: Evolutionary and Therapeutic Implications. *PLoS One* 10, e0139001 (2015).

- 24. Holmes, E. C. The RNA virus quasispecies: fact or fiction? J. Mol. Biol. 400, 271–273 (2010).
- 25. Lauring, A. S. & Andino, R. Quasispecies Theory and the Behavior of RNA Viruses. *PLOS Pathog.* 6, e1001005 (2010).
- 26. Novella, I. S., Presloid, J. B. & Taylor, R. T. RNA replication errors and the evolution of virus pathogenicity and virulence. *Curr. Opin. Virol.* **9**, 143–147 (2014).
- 27. Chetverin, A. B. The puzzle of RNA recombination. *FEBS Lett.* **460**, 1–5 (1999).
- Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F. & González-Candelas, F. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **30**, 296–307 (2015).
- 29. McDonald, S. M., Nelson, M. I., Turner, P. E. & Patton, J. T. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nat. Rev. Microbiol.* **14**, 448–460 (2016).
- 30. Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9**, 617–626 (2011).
- 31. Woolhouse, M. E. J. & Adair, K. Europe PMC Funders Group The diversity of human RNA viruses. **8**, 159–171 (2018).
- 32. Woolhouse, M. E. J. & Brierley, L. Epidemiological characteristics of human-infective RNA viruses. *Sci. Data* **5**, 1–6 (2018).
- 33. Seitz, R. Human Immunodeficiency Virus (HIV). *Transfus. Med. Hemotherapy* **43**, 203–222 (2016).
- 34. Manns, M. P. et al. Hepatitis C virus infection. Nat. Rev. Dis. Prim. 3, (2017).
- 35. Moghadami, M. A Narrative Review of Influenza: A Seasonal and Pandemic Disease. *Iran. J. Med. Sci.* **42**, 2–13 (2017).
- 36. Borchers, A. T., Chang, C., Gershwin, M. E. & Gershwin, L. J. Respiratory syncytial virus--a comprehensive review. *Clin. Rev. Allergy Immunol.* **45**, 331–379 (2013).
- Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y. & Yuen, K. Y. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin. Microbiol. Rev.* 20, 660–694 (2007).
- 38. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* (2020) doi:10.1038/s41579-020-00459-7.
- 39. Fehr, A. R., Channappanavar, R. & Perlman, S. Middle East Respiratory Syndrome: Emergence of a Pathogenic Human Coronavirus. *Annu. Rev. Med.* **68**, 387–399 (2017).
- 40. Hasan, S., Ahmad, S. A., Masood, R. & Saeed, S. Ebola virus: A global public health menace: A narrative review. *J. Fam. Med. Prim. care* **8**, 2189–2201 (2019).
- 41. Jacobs, S. E., Lamson, D. M., St. George, K. & Walsh, T. J. Human Rhinoviruses. *Clin. Microbiol. Rev.* 26, 135 LP 162 (2013).
- 42. Mehndiratta, M. M., Mehndiratta, P. & Pande, R. Poliomyelitis: historical facts, epidemiology, and current challenges in eradication. *The Neurohospitalist* **4**, 223–229 (2014).
- 43. Laksono, B. M., de Vries, R. D., McQuaid, S., Duprex, W. P. & de Swart, R. L. Measles Virus Host Invasion and Pathogenesis. *Viruses* **8**, 210 (2016).
- 44. Sadiq, A., Bostan, N., Yinda, K. C., Naseem, S. & Sattar, S. Rotavirus: Genetics, pathogenesis and vaccine advances. *Rev. Med. Virol.* **28**, e2003 (2018).
- 45. Hasan, S., Jamdar, S. F., Alalowi, M. & Al Ageel Al Beaiji, S. M. Dengue virus: A global human threat: Review of literature. *J. Int. Soc. Prev. Community Dent.* **6**, 1–6 (2016).
- 46. Agumadu, V. C. & Ramphul, K. Zika Virus: A Review of Literature. *Cureus* 10, e3025–e3025 (2018).
- 47. Douam, F. & Ploss, A. Yellow fever virus: knowledge gaps impeding the fight against an old foe. *Trends Microbiol.* **26**, 913–928 (2018).

- 48. Galán-Huerta, K. A., Rivas-Estilla, A. M., Fernández-Salas, I., Farfan-Ale, J. A. & Ramos-Jiménez, J. Chikungunya virus: A general overview. *Med. Univ.* **17**, 175–183 (2015).
- 49. Singh, R. *et al.* Rabies epidemiology, pathogenesis, public health concerns and advances in diagnosis and control: a comprehensive review. *Vet. Q.* **37**, 212–251 (2017).
- 50. Zhou, S. F. & Zhong, W. Z. Drug design and discovery: Principles and applications. *Molecules* **22**, 1–6 (2017).
- 51. Di, L., Kerns, E. H. & Carter, G. T. Drug-like property concepts in pharmaceutical design. *Curr. Pharm. Des.* **15**, 2184–2194 (2009).
- 52. Muller, P. Y. & Milton, M. N. The determination and interpretation of the therapeutic index in drug development. *Nat. Rev. Drug Discov.* **11**, 751–761 (2012).
- 53. Wang, X., Zou, P., Wu, F., Lu, L. & Jiang, S. Development of small-molecule viral inhibitors targeting various stages of the life cycle of emerging and re-emerging viruses. *Front. Med.* 11, 449–461 (2017).
- 54. Barré-Sinoussi, F., Ross, A. L. & Delfraissy, J. F. Past, present and future: 30 years of HIV research. *Nat. Rev. Microbiol.* **11**, 877–883 (2013).
- 55. De Clercq, E. Fifty Years in Search of Selective Antiviral Drugs. J. Med. Chem. 62, 7322–7339 (2019).
- 56. Chaudhuri, S., Symons, J. A. & Deval, J. Innovation and trends in the development and approval of antiviral medicines: 1987–2017 and beyond. *Antiviral Res.* **155**, 76–88 (2018).
- 57. Ghongade, K., Bilur, L. & Mali, A. 'A REVIEW ON CURRENT APPROACHES OF DEVELOPMENT OF ANTIVIRAL DRUGS'. *Int. J. Biol. Macromol.* 7, 1–21 (2018).
- 58. Berke, J. M. *et al.* Capsid assembly modulators have a dual mechanism of action in primary human hepatocytes infected with hepatitis B virus. *Antimicrob. Agents Chemother.* **61**, 1–14 (2017).
- 59. Bassetti, M., Castaldo, N. & Carnelutti, A. Neuraminidase inhibitors as a strategy for influenza treatment: pros, cons and future perspectives. *Expert Opin. Pharmacother.* **20**, 1711–1718 (2019).
- 60. Helenius, A. Virus Entry: Looking Back and Moving Forward. J. Mol. Biol. 430, 1853–1862 (2018).
- 61. Melikyan, G. B. HIV entry: a game of hide-and-fuse? *Curr. Opin. Virol.* 4, 1–7 (2014).
- 62. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: cell binding and entry. *Cold Spring Harb. Perspect. Med.* **2**, a006866 (2012).
- 63. Yi, H. A., Fochtman, B. C., Rizzo, R. C. & Jacobs, A. Inhibition of HIV Entry by Targeting the Envelope Transmembrane Subunit gp41. *Curr. HIV Res.* **14**, 283–294 (2016).
- 64. Steen, A., Schwartz, T. W. & Rosenkilde, M. M. Targeting CXCR4 in HIV cell-entry inhibition. *Mini Rev. Med. Chem.* 9, 1605–1621 (2009).
- Sayana, S. & Khanlou, H. Maraviroc: a new CCR5 antagonist. *Expert Rev. Anti. Infect. Ther.* 7, 9–19 (2009).
- 66. Beard, K. R., Brendish, N. J. & Clark, T. W. Treatment of influenza with neuraminidase inhibitors. *Curr. Opin. Infect. Dis.* **31**, 514–519 (2018).
- 67. Roymans, D. *et al.* Therapeutic efficacy of a respiratory syncytial virus fusion inhibitor. *Nat. Commun.* **8**, 167 (2017).
- 68. Cocquerel, L., Voisset, C. & Dubuisson, J. Hepatitis C virus entry: potential receptors and their biological functions. *J. Gen. Virol.* **87**, 1075–1084 (2006).
- Squires, K. E. An introduction to nucleoside and nucleotide analogues. *Antivir. Ther.* 6 Suppl 3, 1–14 (2001).
- Shen, C.-H. Chapter 11 Techniques in Sequencing. in (ed. Shen, C.-H. B. T.-D. M. B.) 277– 302 (Academic Press, 2019). doi:https://doi.org/10.1016/B978-0-12-802823-0.00011-0.

- 71. Seki, J. I. Nucleoside Analogues. Nippon Nogeikagaku Kaishi 66, 1349–1353 (1992).
- 72. Martin, J. C., Hitchcock, M. J. M., De Clercq, E. & Prusoff, W. H. Early nucleoside reverse transcriptase inhibitors for the treatment of HIV: A brief history of stavudine (D4T) and its comparison with other dideoxynucleosides. *Antiviral Res.* **85**, 34–38 (2010).
- 73. Jordan, P. C., Stevens, S. K. & Deval, J. Nucleosides for the treatment of respiratory RNA virus infections. *Antivir. Chem. Chemother.* **26**, 1–19 (2018).
- 74. Das, K. & Arnold, E. HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. *Curr. Opin. Virol.* **3**, 111–118 (2013).
- 75. Campiani, G. *et al.* Non-nucleoside HIV-1 reverse transcriptase (RT) inhibitors: past, present, and future perspectives. *Curr. Pharm. Des.* **8**, 615–657 (2002).
- Sarisky, R. T. Non-nucleoside inhibitors of the HCV polymerase. J. Antimicrob. Chemother. 54, 14–16 (2004).
- 77. Mifsud, E. J., Hayden, F. G. & Hurt, A. C. Antivirals targeting the polymerase complex of influenza viruses. *Antiviral Res.* **169**, 104545 (2019).
- 78. DeVincenzo, J. P. *et al.* Activity of Oral ALS-008176 in a Respiratory Syncytial Virus Challenge Study. *N. Engl. J. Med.* **373**, 2048–2058 (2015).
- 79. Warren, T. K. *et al.* Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature* **531**, 381–385 (2016).
- 80. Yost, S. A. & Marcotrigiano, J. Viral precursor polyproteins: keys of regulation from replication to maturation. *Curr. Opin. Virol.* **3**, 137–142 (2013).
- 81. Anderson, J., Schiffer, C., Lee, S.-K. & Swanstrom, R. Viral protease inhibitors. *Handb. Exp. Pharmacol.* **189**, 85–110 (2009).
- 82. Wensing, A. M. J., van Maarseveen, N. M. & Nijhuis, M. Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance. *Antiviral Res.* **85**, 59–74 (2010).
- 83. Nitsche, C. Strategies Towards Protease Inhibitors for Emerging Flaviviruses. *Adv. Exp. Med. Biol.* **1062**, 175–186 (2018).
- 84. Andrake, M. D. & Skalka, A. M. Retroviral Integrase: Then and Now. *Annu. Rev. Virol.* **2**, 241–264 (2015).
- 85. Anstett, K., Brenner, B., Mesplede, T. & Wainberg, M. A. HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* **14**, 36 (2017).
- 86. Weber, D. J., Rutala, W. A., Fischer, W. A., Kanamori, H. & Sickbert-Bennett, E. E. Emerging infectious diseases: Focus on infection control issues for novel coronaviruses (Severe Acute Respiratory Syndrome-CoV and Middle East Respiratory Syndrome-CoV), hemorrhagic fever viruses (Lassa and Ebola), and highly pathogenic avian influenza v. *Am. J. Infect. Control* **44**, e91–e100 (2016).
- 87. Lucas, G. M. Antiretroviral adherence, drug resistance, viral fitness and HIV disease progression: a tangled web is woven. *J. Antimicrob. Chemother.* **55**, 413–416 (2005).
- 88. Lu, D.-Y. et al. HAART in HIV/AIDS Treatments: Future Trends. Infect. Disord. Drug Targets 18, 15–22 (2018).
- 89. Mizokami, M. *et al.* Ledipasvir and sofosbuvir fixed-dose combination with and without ribavirin for 12 weeks in treatment-naive and previously treated Japanese patients with genotype 1 hepatitis C: an open-label, randomised, phase 3 trial. *Lancet. Infect. Dis.* **15**, 645–653 (2015).
- 90. Nguyen, J. T. *et al.* Efficacy of combined therapy with amantadine, oseltamivir, and ribavirin in vivo against susceptible and amantadine-resistant influenza A viruses. *PLoS One* 7, e31006 (2012).
- 91. Seo, S. *et al.* Combination therapy with amantadine, oseltamivir and ribavirin for influenza A infection: safety and pharmacokinetics. *Antivir. Ther.* **18**, 377–386 (2013).

- 92. Eggleton, J. S. & Nagalli, S. Highly Active Antiretroviral Therapy (HAART). in *StatPearls [Internet]* (StatPearls Publishing, 2020).
- 93. Esser, S., Helbig, D., Hillen, U., Dissemond, J. & Grabbe, S. Side effects of HIV therapy. J. der Dtsch. Dermatologischen Gesellschaft = J. Ger. Soc. Dermatology JDDG 5, 745–754 (2007).
- 94. Memoli, M. *et al.* MultiDrug-Resistant 2009 Pandemic Influenza A(H1N1) Viruses Maintain Fitness and Transmissibility in Ferrets. *J. Infect. Dis.* **203**, 348–357 (2011).
- 95. Yim, H. J. *et al.* Evolution of multi-drug resistant hepatitis B virus during sequential therapy. *Hepatology* **44**, 703–712 (2006).
- 96. Hughes, D. & Andersson, D. I. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nat. Rev. Genet.* **16**, 459–471 (2015).
- Tamura, D. *et al.* Emergence of Multidrug-Resistant Influenza A(H1N1)pdm09 Virus Variants in an Immunocompromised Child Treated With Oseltamivir and Zanamivir. *J. Infect. Dis.* 212, 1209–1213 (2015).
- 98. Saiz, J. C. Vaccines against rna viruses. Vaccines 8, 1-4 (2020).
- 99. Noor, A. & Krilov, L. R. Respiratory syncytial virus vaccine: where are we now and what comes next? *Expert Opin. Biol. Ther.* **18**, 1247–1256 (2018).
- 100. Goh, G. K.-M., Dunker, A. K., Foster, J. A. & Uversky, V. N. HIV Vaccine Mystery and Viral Shell Disorder. *Biomolecules* 9, (2019).
- 101. Qureshi, A., Tantray, V. G., Kirmani, A. R. & Ahangar, A. G. A review on current status of antiviral siRNA. *Rev. Med. Virol.* **28**, e1976 (2018).
- 102. Maillard, P. V, van der Veen, A. G., Poirier, E. Z. & Reis e Sousa, C. Slicing and dicing viruses: antiviral RNA interference in mammals. *EMBO J.* **38**, (2019).
- 103. Chen, Y., Cheng, G. & Mahato, R. I. RNAi for treating hepatitis B viral infection. *Pharm. Res.* **25**, 72–86 (2008).
- 104. Roberts, T. C., Ezzat, K., El Andaloussi, S. & Weinberg, M. S. Synthetic SiRNA Delivery: Progress and Prospects. *Methods Mol. Biol.* **1364**, 291–310 (2016).
- Bartolomé, J., Castillo, I. & Carreño, V. Ribozymes as antiviral agents. *Minerva Med.* 95, 11–24 (2004).
- 106. Menke, A. & Hobom, G. Antiviral ribozymes. Mol. Biotechnol. 8, 17–33 (1997).
- 107. Domingo, E. & Perales, C. Viral Quasispecies and Lethal Mutagenesis. *Eur. Rev.* 24, 39–48 (2016).
- Bull, J. J., Sanjuán, R. & Wilke, C. O. Theory of Lethal Mutagenesis for Viruses. J. Virol. 81, 2930 LP 2939 (2007).
- 109. Shimizu, Y. *et al.* Monoclonal Antibodies against Occludin Completely Prevented Hepatitis C Virus Infection in a Mouse Model. *J. Virol.* **92**, (2018).
- 110. Pelegrin, M., Naranjo-Gomez, M. & Piechaczyk, M. Antiviral Monoclonal Antibodies: Can They Be More Than Simple Neutralizing Agents? *Trends Microbiol.* **23**, 653–665 (2015).
- 111. Caskey, M., Klein, F. & Nussenzweig, M. C. Broadly neutralizing anti-HIV-1 monoclonal antibodies in the clinic. *Nat. Med.* **25**, 547–553 (2019).
- 112. Samuel, C. E. Antiviral actions of interferons. Clin. Microbiol. Rev. 14, 778-809 (2001).
- 113. Park, M. S. *et al.* Towards the Application of Human Defensins as Antivirals. *Biomol. Ther. (Seoul).* **26**, 242–254 (2018).
- 114. Nicholson, B. L. & White, K. A. Exploring the architecture of viral RNA genomes. *Curr. Opin. Virol.* **12**, 66–74 (2015).
- 115. Li, J. & Liu, C. Coding or noncoding, the converging concepts of RNAs. *Front. Genet.* **10**, 1–10 (2019).

- 116. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **20**, 474–489 (2019).
- 117. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **20**, 474–489 (2019).
- 118. Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.* **83**, 441–466 (2014).
- 119. Mathews, D. H., Moss, W. N. & Turner, D. H. Folding and finding RNA secondary structure. *Cold Spring Harb. Perspect. Biol.* **2**, 1–15 (2010).
- 120. Bevilacqua, P. C., Ritchey, L. E., Su, Z. & Assmann, S. M. Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* **50**, 235–266 (2016).
- 121. Westhof, E. & Fritsch, V. RNA folding: beyond Watson–Crick pairs. *Structure* **8**, R55–R65 (2000).
- 122. Leontis, N. B. & Westhof, E. Analysis of RNA motifs. *Curr. Opin. Struct. Biol.* 13, 300–308 (2003).
- Hilbers, C. W., Heus, H. A., Van Dongen, M. J. P. & Wijmenga, S. S. The hairpin elements of nucleic acid structure: DNA and RNA folding. in *Nucleic acids and molecular biology* 56– 104 (Springer, 1994).
- 124. Butcher, S. E. & Pyle, A. M. The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. *Acc. Chem. Res.* 44, 1302–1311 (2011).
- 125. Halder, S. & Bhattacharyya, D. RNA structure and dynamics: a base pairing perspective. *Prog. Biophys. Mol. Biol.* **113**, 264–283 (2013).
- Staple, D. W. & Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* 3, e213–e213 (2005).
- 127. Draper, D. E. A guide to ions and RNA structure. *Rna* 10, 335–343 (2004).
- 128. Sun, L. Z., Zhang, D. & Chen, S. J. Theory and Modeling of RNA Structure and Interactions with Metal Ions and Small Molecules. *Annu. Rev. Biophys.* **46**, 227–246 (2017).
- 129. Jones, C. P. & Ferré-D'Amaré, A. R. RNA quaternary structure and global symmetry. *Trends Biochem. Sci.* **40**, 211–220 (2015).
- 130. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
- 131. Salim, N. *et al.* Thermodynamic and Kinetic Analysis of an RNA Kissing Interaction and Its Resolution into an Extended Duplex. *Biophys. J.* **102**, 1097–1107 (2012).
- 132. Seetin, M. G. & Mathews, D. H. RNA structure prediction: an overview of methods. *Methods Mol. Biol.* **905**, 99–122 (2012).
- Mailler, E., Paillart, J. C., Marquet, R., Smyth, R. P. & Vivet-Boudou, V. The evolution of RNA structural probing methods: From gels to next-generation sequencing. *Wiley Interdiscip. Rev. RNA* 10, 1–20 (2019).
- 134. Cruz, J. A. & Westhof, E. Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat. Methods* **8**, 513–521 (2011).
- 135. Markham, N. & Zuker, M. UNAFold: Software for Nucleic Acid Folding and Hybridization. *Methods Mol. Biol.* **453**, 3–31 (2008).
- 136. Lorenz, R. et al. ViennaRNA Package 2.0. Algorithms Mol. Biol. 6, 26 (2011).
- Xia, T. *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37, 14719–14735 (1998).
- Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288, 911–940 (1999).

- 139. Moss, W. N. & Steitz, J. A. In silico discovery and modeling of non-coding RNA structure in viruses. *Methods* **91**, 48–56 (2015).
- 140. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* **12**, 641–655 (2011).
- Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* 18, 321–329 (2008).
- 142. Mathews, D. Predicting RNA secondary structure by free energy minimization. *Theor. Chem. Acc.* **116**, 160–168 (2006).
- 143. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. RNA SHAPE Chemistry Reveals Nonhierarchical Interactions Dominate Equilibrium Structural Transitions in tRNAAsp Transcripts. J. Am. Chem. Soc. 127, 4659–4667 (2005).
- 144. Lorenz, R., Wolfinger, M. T., Tanzer, A. & Hofacker, I. L. Predicting RNA secondary structures from sequence and probing data. *Methods* **103**, 86–98 (2016).
- 145. Hammer, S., Günzel, C., Mörl, M. & Findeiß, S. Evolving methods for rational de novo design of functional RNA molecules. *Methods* 161, 54–63 (2019).
- 146. Miao, Z. & Westhof, E. RNA Structure: Advances and Assessment of 3D Structure Prediction. *Annu. Rev. Biophys.* **46**, 483–503 (2017).
- 147. Halweg-Edwards, A., Garst, A. & Batey, R. Determining Structures of RNA Aptamers and Riboswitches by X-Ray Crystallography. *Methods Mol. Biol.* **535**, 135–163 (2009).
- Kennedy, S. D. NMR Methods for Characterization of RNA Secondary Structure. *Methods Mol. Biol.* 1490, 253–264 (2016).
- Wilkinson, M. E., Lin, P.-C., Plaschka, C. & Nagai, K. Cryo-EM Studies of Pre-mRNA Splicing: From Sample Preparation to Model Visualization. *Annu. Rev. Biophys.* 47, 175–199 (2018).
- 150. Bermejo, G. A., Clore, G. M. & Schwieters, C. D. Improving NMR Structures of RNA. *Structure* 24, 806–815 (2016).
- 151. Keel, A. Y., Rambo, R. P., Batey, R. T. & Kieft, J. S. A general strategy to solve the phase problem in RNA crystallography. *Structure* **15**, 761–772 (2007).
- 152. Ziehler, W. A. & Engelke, D. R. Probing RNA structure with chemical reagents and enzymes. *Curr. Protoc. nucleic acid Chem.* 1–6 (2000).
- 153. Knapp, G. Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.* **180**, 192–212 (1989).
- 154. Madanecki, P., Nozell, S., Ochocka, R., Collawn, J. F. & Bartoszewski, R. RNAdigest: A web-based tool for the analysis and prediction of structure Specific RNAse digestion results. *PLoS One* **9**, 1–7 (2014).
- Desai, N. A. & Shankar, V. Single-strand-specific nucleases. *FEMS Microbiol. Rev.* 26, 457–491 (2003).
- 156. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107 (2010).
- 157. Fechter, P. et al. Traditional Chemical Mapping of RNA Structure In Vitro and In Vivo. *Methods Mol. Biol.* **1490**, 83–103 (2016).
- 158. Weeks, K. M. Advances in RNA structure analysis by chemical probing. *Curr. Opin. Struct. Biol.* **20**, 295–304 (2010).
- 159. Morgan, B. S., Forte, J. E. & Hargrove, A. E. Insights into the development of chemical probes for RNA. *Nucleic Acids Res.* **46**, 8025–8037 (2018).
- 160. Mitchell, D. 3rd, Assmann, S. M. & Bevilacqua, P. C. Probing RNA structure in vivo. *Curr. Opin. Struct. Biol.* **59**, 151–158 (2019).

- Busan, S., Weidmann, C. A., Sengupta, A. & Weeks, K. M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* 58, 2655– 2664 (2019).
- 162. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15**, 469–479 (2014).
- 163. Strobel, E. J., Yu, A. M. & Lucks, J. B. High-throughput determination of RNA structures. *Nat. Rev. Genet.* **19**, 615–634 (2018).
- 164. Bevilacqua, P. C. & Assmann, S. M. Technique development for probing RNA structure in vivo and genome-wide. *Cold Spring Harb. Perspect. Biol.* **10**, (2018).
- Kwok, C. K., Tang, Y., Assmann, S. M. & Bevilacqua, P. C. The RNA structurome: Transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem. Sci.* 40, 221–232 (2015).
- 166. Underwood, J. G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001 (2010).
- 167. Zheng, Q. *et al.* Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. *PLoS Genet.* **6**, e1001141 (2010).
- 168. Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M. & Bevilacqua, P. C. Determination of in vivo RNA structure in low-abundance transcripts. *Nat. Commun.* **4**, 2971 (2013).
- 169. Spitale, R. C. et al. RNA SHAPE analysis in living cells. Nat. Chem. Biol. 9, 18–20 (2013).
- 170. Leamy, K. A., Assmann, S. M., Mathews, D. H. & Bevilacqua, P. C. Bridging the gap between in vitro and in vivo RNA folding. *Q. Rev. Biophys.* **49**, e10–e10 (2016).
- Ding, Y., Kwok, C. K., Tang, Y., Bevilacqua, P. C. & Assmann, S. M. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* 10, 1050–1066 (2015).
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705 (2014).
- 173. Talkish, J., May, G., Lin, Y., Woolford, J. L. & McManus, C. J. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *Rna* **20**, 713–720 (2014).
- Incarnato, D., Neri, F., Anselmi, F. & Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* 15, 491 (2014).
- Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U. S. A.* 108, 11063–11068 (2011).
- 176. Watters, K. E., Yu, A. M., Strobel, E. J., Settle, A. H. & Lucks, J. B. Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods* **103**, 34–48 (2016).
- Poulsen, L. D., Kielpinski, L. J., Salama, S. R., Krogh, A. & Vinther, J. SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA* 21, 1042–1052 (2015).
- 178. Chan, D., Beasley, S., Zhen, Y. & Spitale, R. C. Facile synthesis and evaluation of a dualfunctioning furoyl probe for in-cell SHAPE. *Bioorg. Med. Chem. Lett.* **28**, 601–605 (2018).
- 179. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **10**, 1643–1669 (2015).
- Smola, M. J. & Weeks, K. M. In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.* 13, 1181–1195 (2018).

- Novoa, E. M., Beaudoin, J.-D., Giraldez, A. J., Mattick, J. S. & Kellis, M. Best practices for genome-wide RNA structure analysis: combination of mutational profiles and drop-off information. *bioRxiv* 176883 (2020) doi:10.1101/176883.
- 182. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11, 959–965 (2014).
- 183. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14**, 75–82 (2016).
- 184. Krokhotin, A., Mustoe, A. M., Weeks, K. M. & Dokholyan, N. V. Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* **23**, 6–13 (2017).
- 185. Wan, Y. *et al.* Genome-wide measurement of RNA folding energies. *Mol. Cell* **48**, 169–181 (2012).
- Chursov, A., Kopetzky, S. J., Bocharov, G., Frishman, D. & Shneider, A. RNAtips: Analysis of temperature-induced changes of RNA secondary structure. *Nucleic Acids Res.* 41, W486– W491 (2013).
- 187. Fischer, N. M., Polěto, M. D., Steuer, J. & Van Der Spoel, D. Influence of Na + and Mg 2+ ions on RNA structures studied with molecular dynamics simulations. *Nucleic Acids Res.* 46, 4872–4882 (2018).
- 188. Sun, L.-Z. & Chen, S.-J. A New Method to Predict Ion Effects in RNA Folding. *Methods Mol. Biol.* **1632**, 1–17 (2017).
- 189. Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).
- 190. Silverman, I. M. *et al.* RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol.* **15**, 1–16 (2014).
- 191. Schlick, T. & Pyle, A. M. Opportunities and Challenges in RNA Structural Modeling and Design. *Biophys. J.* **113**, 225–234 (2017).
- 192. Tian, S. & Das, R. RNA structure through multidimensional chemical mapping. *Q. Rev. Biophys.* **49**, e7 (2016).
- 193. Kladwang, W., VanLang, C. C., Cordero, P. & Das, R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.* **3**, 954–962 (2011).
- 194. Cheng, C. Y., Kladwang, W., Yesselman, J. D. & Das, R. RNA structure inference through chemical mapping after accidental or intentional mutations. *Proc. Natl. Acad. Sci.* 114, 9876 LP – 9881 (2017).
- 195. Cheng, C. Y. *et al.* Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife* **4**, e07600–e07600 (2015).
- 196. Tian, S., Cordero, P., Kladwang, W. & Das, R. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **20**, 1815–1826 (2014).
- 197. Ramani, V., Qiu, R. & Shendure, J. High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.* **33**, 980–984 (2015).
- 198. Ponce-Salvatierra, A. *et al.* Computational modeling of RNA 3D structure based on experimental data. *Biosci. Rep.* **39**, 1–25 (2019).
- Ouyang, Z., Snyder, M. P. & Chang, H. Y. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* 23, 377–387 (2013).
- 200. Bellaousov, S., Reuter, J. S., Seetin, M. G. & Mathews, D. H. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.* **41**, (2013).
- 201. Gruber, A. R., Bernhart, S. H. & Lorenz, R. The ViennaRNA web services. *Methods Mol. Biol.* **1269**, 307–326 (2015).

- 202. Boerneke, M. A., Ehrhardt, J. E. & Weeks, K. M. Physical and Functional Analysis of Viral RNA Genomes by SHAPE. *Annu. Rev. Virol.* **6**, 93–117 (2019).
- 203. Sükösd, Z. *et al.* Characterising RNA secondary structure space using information entropy. *BMC Bioinformatics* **14 Suppl 2**, 1–9 (2013).
- 204. Schroeder, S. J. Challenges and approaches to predicting RNA with multiple functional structures. *Rna* 24, 1615–1624 (2018).
- 205. Watters, K. E. *et al.* Probing of RNA structures in a positive sense RNA virus reveals selection pressures for structural elements. *Nucleic Acids Res.* **46**, 2573–2584 (2018).
- 206. Sükösd, Z. *et al.* Full-length RNA structure prediction of the HIV-1 genome reveals a conserved core domain. *Nucleic Acids Res.* **43**, 10168–10179 (2015).
- 207. Villordo, S. M., Carballeda, J. M., Filomatori, C. V. & Gamarnik, A. V. RNA Structure Duplications and Flavivirus Host Adaptation. *Trends Microbiol.* **24**, 270–283 (2016).
- 208. Hesketh, E. L. *et al.* Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nat. Commun.* **6**, 1–10 (2015).
- Larman, B. C., Dethoff, E. A. & Weeks, K. M. Packaged and Free Satellite Tobacco Mosaic Virus (STMV) RNA Genomes Adopt Distinct Conformational States. *Biochemistry* 56, 2175–2183 (2017).
- 210. Li, P. *et al.* Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity. *Cell Host Microbe* **24**, 875-886.e5 (2018).
- Sanjuán, R. Mutational fitness effects in RNA and single-stranded DNA viruses: Common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 1975–1982 (2010).
- 212. Hermann, T. Small molecules targeting viral RNA. *Wiley Interdiscip. Rev. RNA* 7, 726–743 (2016).
- Crary, S. M., Towner, J. S., Honig, J. E., Shoemaker, T. R. & Nichol, S. T. Analysis of the role of predicted RNA secondary structures in Ebola virus replication. *Virology* 306, 210–218 (2003).
- 214. Weik, M., Enterlein, S., Schlenz, K. & Mühlberger, E. The Ebola virus genomic replication promoter is bipartite and follows the rule of six. *J. Virol.* **79**, 10660–10671 (2005).
- 215. Gebhard, L. G., Filomatori, C. V & Gamarnik, A. V. Functional RNA elements in the dengue virus genome. *Viruses* **3**, 1739–1756 (2011).
- 216. Silva, P. A. G. C., Pereira, C. F., Dalebout, T. J., Spaan, W. J. M. & Bredenbeek, P. J. An RNA Pseudoknot Is Required for Production of Yellow Fever Virus Subgenomic RNA by the Host Nuclease XRN1. J. Virol. 84, 11395 LP – 11406 (2010).
- 217. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
- 218. Jayaraman, D. & Kenyon, J. C. New windows into retroviral RNA structures. *Retrovirology* **15**, 11 (2018).
- Rausch, J. W., Sztuba-Solinska, J. & Le Grice, S. F. J. Probing the structures of viral RNA regulatory elements with SHAPE and related methodologies. *Front. Microbiol.* 8, 1–15 (2018).
- 220. Lu, K., Heng, X. & Summers, M. F. Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410**, 609–633 (2011).
- 221. Keane, S. C. *et al.* Structure of the HIV-1 RNA packaging signal. *Science (80-. ).* **348**, 917–921 (2015).
- 222. Comas-Garcia, M. *et al.* Dissection of specific binding of HIV-1 Gag to the 'packaging signal' in viral RNA. *Elife* **6**, (2017).

- 223. Girard, F. *et al.* Dimer initiation sequence of HIV-1Lai genomic RNA: NMR solution structure of the extended duplex. *J. Biomol. Struct. Dyn.* **16**, 1145–1157 (1999).
- 224. Dubois, N., Marquet, R., Paillart, J.-C. & Bernacchi, S. Retroviral RNA Dimerization: From Structure to Functions. *Front. Microbiol.* **9**, 527 (2018).
- 225. Bannwarth, S. & Gatignol, A. HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. *Curr. HIV Res.* **3**, 61–71 (2005).
- Schulze-Gahmen, U. & Hurley, J. H. Structural mechanism for HIV-1 TAR loop recognition by Tat and the super elongation complex. *Proc. Natl. Acad. Sci.* 115, 12973 LP – 12978 (2018).
- 227. Harrich, D., Mavankal, G., Mette-Snider, A. & Gaynor, R. B. Human immunodeficiency virus type 1 TAR element revertant viruses define RNA structures required for efficient viral gene expression and replication. *J. Virol.* **69**, 4906–4913 (1995).
- 228. Rausch, J. W. & Le Grice, S. F. J. HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective. *Viruses* 7, 3053–3075 (2015).
- 229. Suhasini, M. & Reddy, T. R. Cellular proteins and HIV-1 Rev function. *Curr. HIV Res.* 7, 91–100 (2009).
- 230. Fernandes, J., Jayaraman, B. & Frankel, A. The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA Biol.* **9**, 6–11 (2012).
- 231. Adams, R. L., Pirakitikulr, N. & Pyle, A. M. Functional RNA structures throughout the Hepatitis C Virus genome. *Curr. Opin. Virol.* 24, 79–86 (2017).
- 232. You, S. & Rice, C. M. 3' RNA Elements in Hepatitis C Virus Replication: Kissing Partners and Long Poly(U). J. Virol. 82, 184 LP 195 (2008).
- 233. Reigadas, S. *et al.* HCV RNA-dependent RNA polymerase replicates in vitro the 3' terminal region of the minus-strand viral RNA more efficiently than the 3' terminal region of the plus RNA. *Eur. J. Biochem.* **268**, 5857–5867 (2001).
- Pirakitikulr, N., Kohlway, A., Lindenbach, B. D. & Pyle, A. M. The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures. *Mol. Cell* 62, 111–120 (2016).
- 235. Mauger, D. M. *et al.* Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3692–3697 (2015).
- 236. Lukavsky, P. J. Structure and function of HCV IRES domains. *Virus Res.* 139, 166–171 (2009).
- 237. Vopálenský, V. *et al.* Characterization of Hepatitis C Virus IRES Quasispecies From the Individual to the Pool. *Frontiers in Microbiology* vol. 9 731 (2018).
- 238. Yamamoto, H. *et al.* Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. *EMBO J.* **34**, 3042–3058 (2015).
- Honda, M., Brown, E. A. & Lemon, S. M. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* 2, 955–968 (1996).
- 240. Noble, E. *et al.* Biophysical analysis of influenza A virus RNA promoter at physiological temperatures. *J. Biol. Chem.* **286**, 22965–22970 (2011).
- 241. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360 (2014).
- 242. Bae, S.-H. *et al.* Structural features of an influenza virus promoter and their implications for viral RNA synthesis. *Proc. Natl. Acad. Sci.* **98**, 10602 LP 10607 (2001).
- 243. Dadonaite, B. *et al.* The structure of the influenza A virus genome. *Nat. Microbiol.* **4**, 1781–1789 (2019).

- 244. Simon, L. M. *et al.* In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs. *Nucleic Acids Res.* **47**, 7003–7017 (2019).
- 245. Huber, R. G. *et al.* Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.* **10**, (2019).
- 246. Chapman, E. G. *et al.* The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* **344**, 307–310 (2014).
- 247. Akiyama, B. M. *et al.* Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science (80-. ).* **354**, 1148–1152 (2016).
- 248. Slonchak, A. & Khromykh, A. A. Subgenomic flaviviral RNAs: What do we know after the first decade of research. *Antiviral Res.* **159**, 13–25 (2018).
- 249. Manfredonia, I. *et al.* Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* 1–17 (2020) doi:10.1093/nar/gkaa1053.
- 250. Ziv, O. *et al.* The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol. Cell* 1–11 (2020) doi:10.1016/j.molcel.2020.11.004.
- 251. Lauber, C. *et al.* The Footprint of Genome Architecture in the Largest Genome Expansion in RNA Viruses. *PLoS Pathog.* **9**, (2013).
- 252. Manfredonia, I. & Incarnato, D. Structure and regulation of coronavirus genomes: state-of-the-art and novel insights from SARS-CoV-2 studies. *Biochem. Soc. Trans.* **0**, 1–12 (2020).
- 253. Sanders, W. *et al.* Comparative analysis of coronavirus genomic RNA structure reveals conservation in SARS-like coronaviruses. *bioRxiv* 2020.06.15.153197 (2020) doi: 10.1101/2020.06.15.153197.
- 254. Wacker, A. *et al.* Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* **48**, 12415–12435 (2020).
- 255. Li, L. *et al.* Structural Lability in Stem-Loop 1 Drives a 5' UTR-3' UTR Interaction in Coronavirus Replication. *J. Mol. Biol.* **377**, 790–803 (2008).
- 256. Madhugiri, R. *et al.* Structural and functional conservation of cis-acting RNA elements in coronavirus 5'-terminal genome regions. *Virology* **517**, 44–55 (2018).
- 257. Liu, P. *et al.* A U-turn motif-containing stem-loop in the coronavirus 5' untranslated region plays a functional role in replication. *RNA* **13**, 763–780 (2007).
- 258. Yang, D., Liu, P., Giedroc, D. P. & Leibowitz, J. Mouse Hepatitis Virus Stem-Loop 4 Functions as a Spacer Element Required To Drive Subgenomic RNA Synthesis. J. Virol. 85, 9199 LP – 9209 (2011).
- Hsue, B. & Masters, P. S. A bulged stem-loop structure in the 3' untranslated region of the genome of the coronavirus mouse hepatitis virus is essential for replication. *J. Virol.* 71, 7567–7578 (1997).
- Hsue, B., Hartshorne, T. & Masters, P. S. Characterization of an Essential RNA Secondary Structure in the 3' Untranslated Region of the Murine Coronavirus Genome. J. Virol. 74, 6911 LP – 6921 (2000).
- 261. Yang, D. & Leibowitz, J. L. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **206**, 120–133 (2015).
- 262. Goebel, S. J., Hsue, B., Dombrowski, T. F. & Masters, P. S. Characterization of the RNA Components of a Putative Molecular Switch in the 3' Untranslated Region of the Murine Coronavirus Genome. J. Virol. 78, 669 LP – 682 (2004).
- 263. Huston, N. C., Wan, H., Araujo Tavares, R. de C., Wilen, C. & Pyle, A. M. Comprehensive in-vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *bioRxiv* 2020.07.10.197079 (2020) doi:10.1101/2020.07.10.197079.
- 264. Robertson, M. P. *et al.* The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* **3**, (2005).

- 265. Rangan, R. *et al.* RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA* **26**, 937–959 (2020).
- 266. Kelly, J. A. *et al.* Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS-CoV-2. *bioRxiv Prepr. Serv. Biol.* 2020.03.13.991083 (2020) doi: 10.1101/2020.03.13.991083.
- 267. Plant, E. Ribosomal Frameshift Signals in Viral Genomes. in (2012). doi:10.5772/26550.
- 268. Plant, E. P. *et al.* A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol.* **3**, e172 (2005).
- 269. Brierley, I., Digard, P. & Inglis, S. C. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537–547 (1989).
- 270. Zhang, K. *et al.* Cryo-electron Microscopy and Exploratory Antisense Targeting of the 28kDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome. *bioRxiv* 2020.07.18.209270 (2020) doi:10.1101/2020.07.18.209270.
- Plant, E. P., Sims, A. C., Baric, R. S., Dinman, J. D. & Taylor, D. R. Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses* 5, 279–294 (2013).
- 272. Dinman, J. D. Programmed –1 Ribosomal Frameshifting in SARS Coronavirus. *Mol. Biol. SARS-Coronavirus* 63–72 (2009) doi:10.1007/978-3-642-03683-5\_5.
- 273. Schlick, T., Zhu, Q., Jain, S. & Yan, S. Structure-Altering Mutations of the SARS-CoV-2 Frameshifting RNA Element. *Biophys. J.* 1–14 (2020) doi:10.1016/j.bpj.2020.10.012.
- 274. Tavares, R. de C. A., Mahadeshwar, G., Wan, H., Huston, N. C. & Pyle, A. M. The global and local distribution of RNA structure throughout the SARS-CoV-2 genome. *J. Virol.* JVI. 02190-20 (2020) doi:10.1128/JVI.02190-20.
- 275. Huston, N. C. *et al.* Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell* (2021) doi:https://doi.org/ 10.1016/j.molcel.2020.12.041.
- Ziv, O. *et al.* The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol. Cell* 80, 1067-1077.e5 (2020).
- Cui, H. & Zhang, L. G-Quadruplexes Are Present in Human Coronaviruses Including SARS-CoV-2. *Front. Microbiol.* 11, 1–8 (2020).
- Zhang, R., Xiao, K., Gu, Y., Liu, H. & Sun, X. Whole Genome Identification of Potential G-Quadruplexes and Analysis of the G-Quadruplex Binding Domain for SARS-CoV-2. *Front. Genet.* 11, 587829 (2020).
- 279. Rizvi, N. F. & Smith, G. F. RNA as a small molecule druggable target. *Bioorg. Med. Chem. Lett.* 27, 5083–5088 (2017).
- 280. Rizvi, N. F. & Smith, G. F. RNA as a small molecule druggable target. *Bioorganic Med. Chem. Lett.* 27, 5083–5088 (2017).
- 281. Donlic, A. & Hargrove, A. E. Targeting RNA in mammalian systems with small molecules. *Wiley Interdiscip. Rev. RNA* **9**, 1–21 (2018).
- 282. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* 23, 995–1011 (2017).
- 283. Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* **12**, 35–48 (2014).
- 284. Hashemian, S. M. R., Farhadi, T. & Ganjparvar, M. Linezolid: A review of its properties, function, and use in critical care. *Drug Des. Devel. Ther.* **12**, 1759–1767 (2018).
- 285. Aboul-ela, F. Strategies for the design of RNA-binding small molecules. *Future Med. Chem.* 2, 93–119 (2010).

- 286. Morgan, B. S., Forte, J. E., Culver, R. N., Zhang, Y. & Hargrove, A. E. Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angew. Chem. Int. Ed. Engl.* 56, 13498–13502 (2017).
- Chittapragada, M., Roberts, S. & Ham, Y. W. Aminoglycosides: Molecular insights on the recognition of RNA and aminoglycoside mimics. *Perspect. Medicin. Chem.* 2009, 21–37 (2009).
- 288. Kirsebom, L. A., Virtanen, A. & Mikkelsen, N.-E. Aminoglycoside interactions with RNAs and nucleases. in *RNA Towards Medicine* 73–96 (Springer, 2006).
- Guan, L. & Disney, M. D. Recent Advances in Developing Small Molecules Targeting RNA. ACS Chem. Biol. 7, 73–86 (2012).
- 290. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).
- 291. Hewitt, W. M., Calabrese, D. R. & Schneekloth, J. S. Evidence for ligandable sites in structured RNA throughout the Protein Data Bank. *Bioorganic Med. Chem.* 27, 2253–2260 (2019).
- 292. Anderson-Lee, J. *et al.* Principles for predicting RNA secondary structure design difficulty. *J. Mol. Biol.* **428**, 748–757 (2016).
- 293. Schlick, T. & Pyle, A. M. Opportunities and Challenges in RNA Structural Modeling and Design. *Biophysical journal* vol. 113 225–234 (2017).
- 294. Gooding, K. B. *et al.* High throughput screening of library compounds against an oligonucleotide substructure of an RNA target. *J. Am. Soc. Mass Spectrom.* **15**, 884–892 (2004).
- 295. Ursu, A. *et al.* Design of small molecules targeting RNA structure from sequence. *Chem. Soc. Rev.* **49**, 7252–7270 (2020).
- 296. Zhu, S., Rooney, S. & Michlewski, G. RNA-targeted therapies and high-throughput screening methods. *Int. J. Mol. Sci.* **21**, 1–29 (2020).
- 297. Connelly, C. M., Moon, M. H. & Schneekloth, J. S. J. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem. Biol.* **23**, 1077–1090 (2016).
- 298. Haniff, H. S., Knerr, L., Chen, J. L., Disney, M. D. & Lightfoot, H. L. Target-Directed Approaches for Screening Small Molecules against RNA Targets. *SLAS Discov.* **25**, 869–894 (2020).
- 299. Disney, M. D., Yildirim, I. & Childs-Disney, J. L. Methods to enable the design of bioactive small molecules targeting RNA. *Org. Biomol. Chem.* **12**, 1029–1039 (2014).
- 300. Hanspach, G., Trucks, S. & Hengesbach, M. Strategic labelling approaches for RNA singlemolecule spectroscopy. *RNA Biol.* **16**, 1119–1132 (2019).
- Llano-Sotelo, B. & Chow, C. S. RNA-aminoglycoside antibiotic interactions: fluorescence detection of binding and conformational change. *Bioorg. Med. Chem. Lett.* 9, 213–216 (1999).
- 302. Parsons, J. & Hermann, T. Conformational flexibility of ribosomal decoding-site RNA monitored by fluorescent pteridine base analogues. *Tetrahedron* **63**, 3548–3552 (2007).
- Benz, A., Singh, V., Mayer, T. U. & Hartig, J. S. Identification of Novel Quadruplex Ligands from Small Molecule Libraries by FRET-Based High-Throughput Screening. *ChemBioChem* 12, 1422–1426 (2011).
- Menendez-Gil, P., Caballero, C. J., Solano, C. & Toledo-Arana, A. Fluorescent Molecular Beacons Mimicking RNA Secondary Structures to Study RNA Chaperone Activity. in *RNA Chaperones* 41–58 (Springer, 2020).

- 305. Sommer, G. *et al.* Applying a high-throughput fluorescence polarization assay for the discovery of chemical probes blocking La:RNA interactions in vitro and in cells. *PLoS One* **12**, e0173246–e0173246 (2017).
- 306. Moon, M. H., Hilimire, T. A., Sanders, A. M. & Schneekloth, J. S. J. Measuring RNA-Ligand Interactions with Microscale Thermophoresis. *Biochemistry* **57**, 4638–4643 (2018).
- 307. Silvers, R., Keller, H., Schwalbe, H. & Hengesbach, M. Differential scanning fluorimetry for monitoring RNA stability. *Chembiochem* **16**, 1109–1114 (2015).
- 308. Connelly, C. M., Abulwerdi, F. A. & Schneekloth Jr, J. S. Discovery of RNA Binding Small Molecules Using Small Molecule Microarrays. *Methods Mol. Biol.* **1518**, 157–175 (2017).
- 309. Abulwerdi, F. A. & Schneekloth, J. S. J. Microarray-based technologies for the discovery of selective, RNA-binding molecules. *Methods* **103**, 188–195 (2016).
- 310. Childs-Disney, J. L. *et al.* A Massively Parallel Selection of Small Molecule-RNA Motif Binding Partners Informs Design of an Antiviral from Sequence. *Chem* **4**, 2384–2404 (2018).
- 311. Disney, M. D. *et al.* Informa 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem. Biol.* **11**, 1720–1728 (2016).
- 312. Hoa, X. D., Kirk, A. G. & Tabrizian, M. Towards integrated and sensitive surface plasmon resonance biosensors: a review of recent progress. *Biosens. Bioelectron.* 23, 151–160 (2007).
- Hendrix, M., Priestley, E. S., Joyce, G. F. & Wong, C. H. Direct observation of aminoglycoside-RNA interactions by surface plasmon resonance. J. Am. Chem. Soc. 119, 3641–3648 (1997).
- 314. Win, M. N., Klein, J. S. & Smolke, C. D. Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucleic Acids Res.* **34**, 5670–5682 (2006).
- Greig, M. J. & Robinson, J. M. Detection of oligonucleotide-ligand complexes by ESI-MS (DOLCE-MS) as a component of high throughput screening. *J. Biomol. Screen.* 5, 441–454 (2000).
- 316. Rizvi, N. F. *et al.* Discovery of Selective RNA-Binding Small Molecules by Affinity-Selection Mass Spectrometry. *ACS Chem. Biol.* **13**, 820–831 (2018).
- Rizvi, N. F. & Nickbarg, E. B. RNA-ALIS: Methodology for screening soluble RNAs as small molecule targets using ALIS affinity-selection mass spectrometry. *Methods* 167, 28–38 (2019).
- 318. Rizvi, N. F. *et al.* Discovery of Selective RNA-Binding Small Molecules by Affinity-Selection Mass Spectrometry. *ACS Chem. Biol.* **13**, 820–831 (2018).
- Rizvi, N. F. *et al.* Targeting RNA with Small Molecules: Identification of Selective, RNA-Binding Small Molecules Occupying Drug-Like Chemical Space. *SLAS Discov.* 25, 384–396 (2020).
- 320. Thompson, R. D., Baisden, J. T. & Zhang, Q. NMR characterization of RNA small molecule interactions. *Methods* **167**, 66–77 (2019).
- 321. Calabrese, D. R., Connelly, C. M. & Schneekloth, J. S. Chapter Seven Ligand-observed NMR techniques to probe RNA-small molecule interactions. in *RNA Recognition* (ed. Hargrove, A. E. B. T.-M. in E.) vol. 623 131–149 (Academic Press, 2019).
- 322. Edwards, A. L., Garst, A. D. & Batey, R. T. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol. Biol.* **535**, 135–163 (2009).
- 323. Jones, S. Protein-RNA interactions: structural biology and computational modeling techniques. *Biophys. Rev.* **8**, 359–367 (2016).
- 324. Daldrop, P. & Brenk, R. Structure-based virtual screening for the identification of RNAbinding ligands. *Methods Mol. Biol.* **1103**, 127–139 (2014).

- 325. Wehler, T. & Brenk, R. Structure-based discovery of small molecules binding to RNA. in *RNA Therapeutics* 47–77 (Springer, 2017).
- 326. Guilbert, C. & James, T. L. Docking to RNA via Root-Mean-Square-Deviation-Driven Energy Minimization with Flexible Ligands and Flexible Targets. J. Chem. Inf. Model. 48, 1257–1268 (2008).
- 327. Lang, P. T. *et al.* DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **15**, 1219–1230 (2009).
- 328. Pfeffer, P. & Gohlke, H. DrugScoreRNA--knowledge-based scoring function to predict RNA-ligand interactions. *J. Chem. Inf. Model.* **47**, 1868–1876 (2007).
- 329. Philips, A., Milanowska, K., Lach, G. & Bujnicki, J. M. LigandRNA: computational predictor of RNA-ligand interactions. *RNA* **19**, 1605–1616 (2013).
- 330. Ruiz-Carmona, S. *et al.* rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.* **10**, e1003571–e1003571 (2014).
- 331. Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nature reviews. Drug discovery* vol. 10 188–195 (2011).
- 332. Li, Q. Application of Fragment-Based Drug Discovery to Versatile Targets. *Front. Mol. Biosci.* 7, 180 (2020).
- 333. Suresh, B. M. et al. A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. Proc. Natl. Acad. Sci. 202012217 (2020) doi:10.1073/pnas. 2012217117.
- 334. Otto, S., Furlan, R. L. E. & Sanders, J. K. M. Dynamic combinatorial chemistry. *Drug Discov. Today* 7, 117–125 (2002).
- McAnany, J. D. & Miller, B. L. Dynamic combinatorial chemistry as a rapid method for discovering sequence-selective RNA-binding compounds. *Methods Enzymol.* 623, 67–84 (2019).
- 336. Wang, J., Ge, Y. & Xie, X.-Q. Development and Testing of Druglike Screening Libraries. J. *Chem. Inf. Model.* **59**, 53–65 (2019).
- 337. Mehta, A. *et al.* SMMRNA: A database of small molecule modulators of RNA. *Nucleic Acids Res.* **42**, 132–141 (2014).
- 338. Morgan, B. S. *et al.* R-BIND: An Interactive Database for Exploring and Developing RNA-Targeted Chemical Probes. *ACS Chem. Biol.* **14**, 2691–2700 (2019).
- 339. Li, Q. *et al.* G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **41**, D1115-23 (2013).
- 340. Kumar Mishra, S. & Kumar, A. NALDB: nucleic acid ligand database for small molecules targeting nucleic acid. *Database (Oxford)*. **2016**, (2016).
- 341. Hermann, T. Viral RNA targets and their small molecule ligands. in *RNA Therapeutics* 111–134 (Springer, 2017).
- 342. Mei, H. Y. *et al.* Discovery of selective, small-molecule inhibitors of RNA complexes--I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. *Bioorg. Med. Chem.* **5**, 1173–1184 (1997).
- 343. Mei, H.-Y. *et al.* Inhibition of an HIV-1 Tat-derived peptide binding to TAR RNA by aminoglycoside antibiotics. *Bioorg. Med. Chem. Lett.* **5**, 2755–2760 (1995).
- 344. Yang, M. Discoveries of Tat-TAR interaction inhibitors for HIV-1. *Curr. Drug Targets. Infect. Disord.* **5**, 433–444 (2005).
- 345. Abulwerdi, F. A. & Le Grice, S. F. J. Recent Advances in Targeting the HIV-1 Tat/TAR Complex. *Curr. Pharm. Des.* 23, 4112–4121 (2017).
- 346. Mousseau, G., Mediouni, S. & Valente, S. T. Targeting HIV transcription: the quest for a functional cure. *Curr. Top. Microbiol. Immunol.* **389**, 121–145 (2015).

- 347. Abulwerdi, F. A. *et al.* Development of Small Molecules with a Noncanonical Binding Mode to HIV-1 Trans Activation Response (TAR) RNA. *J. Med. Chem.* **59**, 11148–11160 (2016).
- 348. Maity, D., Kumar, S., Curreli, F., Debnath, A. K. & Hamilton, A. D. α-Helix-Mimetic Foldamers for Targeting HIV-1 TAR RNA. *Chem. A Eur. J.* **25**, 7265–7269 (2019).
- 349. Chavali, S. S., Bonn-Breach, R. & Wedekind, J. E. Face-time with TAR: Portraits of an HIV-1 RNA with diverse modes of effector recognition relevant for drug discovery. *J. Biol. Chem.* **294**, 9326–9341 (2019).
- 350. Davis, B. *et al.* Rational Design of Inhibitors of HIV-1 TAR RNA through the Stabilisation of Electrostatic 'Hot Spots'. *J. Mol. Biol.* **336**, 343–356 (2004).
- 351. Sztuba-Solinska, J. *et al.* Identification of biologically active, HIV TAR RNA-binding small molecules using small molecule microarrays. *J. Am. Chem. Soc.* **136**, 8402–8410 (2014).
- 352. Zeiger, M. et al. Fragment based search for small molecule inhibitors of HIV-1 Tat-TAR. Bioorg. Med. Chem. Lett. 24, 5576–5580 (2014).
- 353. Joly, J.-P. *et al.* Artificial nucleobase-amino acid conjugates: a new class of TAR RNA binding agents. *Chemistry* **20**, 2071–2079 (2014).
- 354. Patwardhan, N. N. *et al.* Amiloride as a new RNA-binding scaffold with activity against HIV-1 TAR. *Medchemcomm* **8**, 1022–1036 (2017).
- 355. Luedtke, N. W. & Tor, Y. Fluorescence-based methods for evaluating the RNA affinity and specificity of HIV-1 Rev-RRE inhibitors. *Biopolymers* **70**, 103–119 (2003).
- 356. Cao, Y. & Liu, X. HIV-1 Rev and related inhibitors. Yao Xue Xue Bao 42, 347-351 (2007).
- 357. Shuck-Lee, D. *et al.* Heterocyclic compounds that inhibit Rev-RRE function and human immunodeficiency virus type 1 replication. *Antimicrob. Agents Chemother.* **52**, 3169–3179 (2008).
- 358. Wynn, J. E. & Santos, W. L. HIV-1 drug discovery: targeting folded RNA structures with branched peptides. *Org. Biomol. Chem.* **13**, 5848–5858 (2015).
- 359. Prado, S. *et al.* Bioavailable inhibitors of HIV-1 RNA biogenesis identified through a Revbased screen. *Biochem. Pharmacol.* **107**, 14–28 (2016).
- 360. Warui, D. M. & Baranger, A. M. Identification of specific small molecule ligands for stem loop 3 ribonucleic acid of the packaging signal Ψ of human immunodeficiency virus-1. J. Med. Chem. 52, 5462–5473 (2009).
- Ingemarsdotter, C. K., Zeng, J., Long, Z., Lever, A. M. L. & Kenyon, J. C. An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *Retrovirology* 15, 1–14 (2018).
- 362. Tam, V. K., Kwong, D. & Tor, Y. Fluorescent HIV-1 dimerization initiation site: Design, properties, and use for ligand discovery. *J. Am. Chem. Soc.* **129**, 3257–3266 (2007).
- 363. Ennifar, E. *et al.* Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res.* **34**, 2328–2339 (2006).
- 364. Lozano, G., Fernandez, N. & Martinez-Salas, E. Modeling three-dimensional structural motifs of viral IRES. *J. Mol. Biol.* **428**, 767–776 (2016).
- Berry, K. E., Waghray, S., Mortimer, S. A., Bai, Y. & Doudna, J. A. Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. *Structure* 19, 1456– 1466 (2011).
- 366. Moon, J. S. *et al.* Inhibition of hepatitis C virus in mice by a small interfering RNA targeting a highly conserved sequence in viral IRES pseudoknot. *PLoS One* **11**, 1–21 (2016).
- 367. Bhat, P. *et al.* Targeting ribosome assembly on the HCV RNA using a small RNA molecule. *RNA Biol.* **9**, 1110–1119 (2012).
- 368. Paulsen, R. B. *et al.* Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc. Natl. Acad. Sci.* **107**, 7263 LP 7268 (2010).

- 369. 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).
- 370. Dibrov, S. M. *et al.* Hepatitis C Virus Translation Inhibitors Targeting the Internal Ribosomal Entry Site. *J. Med. Chem.* **57**, 1694–1707 (2014).
- 371. Wang, W. *et al.* Hepatitis C viral IRES inhibition by phenazine and phenazine-like molecules. *Bioorg. Med. Chem. Lett.* **10**, 1151–1154 (2000).
- 372. Jefferson, E. A. *et al.* Biaryl guanidine inhibitors of in vitro HCV-IRES activity. *Bioorganic & amp; Med. Chem. Lett.* **14**, 5139—5143 (2004).
- 373. 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).
- 374. Parsons, J. *et al.* Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol.* **5**, 823–825 (2009).
- 375. Dibrov, S. M., Johnston-Cox, H., Weng, Y.-H. & Hermann, T. Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew. Chem. Int. Ed. Engl.* **46**, 226–229 (2007).
- 376. 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.* 109, 5223 LP 5228 (2012).
- 377. Zhou, S., Rynearson, K. D., Ding, K., Brunn, N. D. & Hermann, T. Screening for inhibitors of the hepatitis C virus internal ribosome entry site RNA. *Bioorg. Med. Chem.* 21, 6139– 6144 (2013).
- Carnevali, M., Parsons, J., Wyles, D. L. & Hermann, T. A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. *Chembiochem* 11, 1364– 1367 (2010).
- 379. Ding, K., Wang, A., Boerneke, M. A., Dibrov, S. M. & Hermann, T. Aryl-substituted aminobenzimidazoles targeting the hepatitis C virus internal ribosome entry site. *Bioorg. Med. Chem. Lett.* 24, 3113–3117 (2014).
- 380. Rynearson, K. D. *et al.* 2-Aminobenzoxazole ligands of the hepatitis C virus internal ribosome entry site. *Bioorg. Med. Chem. Lett.* **24**, 3521–3525 (2014).
- 381. Bradford, S. & Cowan, J. A. Catalytic metallodrugs targeting HCV IRES RNA. *Chem. Commun. (Camb).* **48**, 3118–3120 (2012).
- 382. Bradford, S. S., Ross, M. J., Fidai, I. & Cowan, J. A. Insight into the recognition, binding, and reactivity of catalytic metallodrugs targeting stem loop IIb of hepatitis C IRES RNA. *ChemMedChem* **9**, 1275–1285 (2014).
- 383. Ross, M. J., Bradford, S. S. & Cowan, J. A. Catalytic metallodrugs based on the LaR2C peptide target HCV SLIV IRES RNA. *Dalton Trans.* 44, 20972–20982 (2015).
- 384. Kim, H. *et al.* Aminoglycoside antibiotics bind to the influenza A virus RNA promoter. *Mol. Biosyst.* **8**, 2857–2859 (2012).
- 385. Lee, M.-K. *et al.* A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun. (Camb).* **50**, 368–370 (2014).
- 386. Bottini, A. *et al.* Targeting Influenza A Virus RNA Promoter. *Chem. Biol. Drug Des.* **86**, 663–673 (2015).
- 387. 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).

- 388. Ritchie, D. B., Soong, J., Sikkema, W. K. A. & Woodside, M. T. Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. J. Am. Chem. Soc. 136, 2196–2199 (2014).
- 389. Lan, T. C. T. *et al.* Structure of the full SARS-CoV-2 RNA genome in infected cells. *bioRxiv* 2020.06.29.178343 (2020) doi:10.1101/2020.06.29.178343.
- 390. 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).
- 391. Haniff, H. S. *et al.* Targeting the SARS-COV-2 RNA genome with small molecule binders and ribonuclease targeting chimera (RiboTAC) degraders. *ACS Cent. Sci.* **6**, 1713–1721 (2020).
- 392. Costales, M. G., Matsumoto, Y., Velagapudi, S. P. & Disney, M. D. Small Molecule Targeted Recruitment of a Nuclease to RNA. *J. Am. Chem. Soc.* **140**, 6741–6744 (2018).
- 393. Di Giorgio, A. & Duca, M. New chemical modalities enabling specific RNA targeting and degradation: Application to SARS-CoV-2 RNA. *ACS Cent. Sci.* **6**, 1647–1650 (2020).
- 394. Manfredonia, I. *et al.* Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **48**, 12436–12452 (2020).
- 395. Zafferani, M. *et al.* Amilorides inhibit SARS-CoV-2 replication in vitro by targeting RNA structures. *bioRxiv* 2020.12.05.409821 (2020).
- 396. Zhao, C. *et al.* Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angew. Chemie* **133**, 436–442 (2021).

# FIGURES AND TABLES SOURCES

Figure 1: <u>http://what-when-how.com/molecular-biology/rna-structure</u>

**Figure 2:** Mailler, Elodie, et al. "The evolution of RNA structural probing methods: From gels to next-generation sequencing." *Wiley Interdisciplinary Reviews: RNA* 10.2 (2019): e1518.

**Figure 3:** Strobel, Eric J., et al. "RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs." *Current opinion in biotechnology* 39 (2016): 182-191.

**Figure 4:** Boerneke, Mark A., Jeffrey E. Ehrhardt, and Kevin M. Weeks. "Physical and functional analysis of viral RNA genomes by SHAPE." *Annual review of virology* 6 (2019): 93-117. <u>Modified</u>

**Figure 5a:** Keane, Sarah C., et al. "Structure of the HIV-1 RNA packaging signal." *Science* 348.6237 (2015): 917-921.

**Figure 5b:** Ennifar, Eric, Philippe Walter, and Philippe Dumas. "Cation-dependent cleavage of the duplex form of the subtype-B HIV-1 RNA dimerization initiation site." *Nucleic acids research* 38.17 (2010): 5807-5816.

**Figure 6a:** Hermann, Thomas. "Small molecules targeting viral RNA." *Wiley Interdisciplinary Reviews: RNA* 7.6 (2016): 726-743.

**Figure 6b/6c**: Fernandes, Jason, Bhargavi Jayaraman, and Alan Frankel. "The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex." *RNA biology* 9.1 (2012): 6-11.

**Figure 7:** Adams, Rebecca L., Nathan Pirakitikulr, and Anna Marie Pyle. "Functional RNA structures throughout the hepatitis C virus genome." *Current opinion in virology* 24 (2017): 79-86.

**Figure 8:** Adams, Rebecca L., Nathan Pirakitikulr, and Anna Marie Pyle. "Functional RNA structures throughout the hepatitis C virus genome." *Current opinion in virology* 24 (2017): 79-86.

**Figure 9a:** Bae, Sung-Hun, et al. "Structural features of an influenza virus promoter and their implications for viral RNA synthesis." *Proceedings of the National Academy of Sciences* 98.19 (2001): 10602-10607.

**Figure 9b**: Ferhadian, Damien, et al. "Structural and functional motifs in influenza virus RNAs." *Frontiers in microbiology* 9 (2018): 559.

**Figure 10:** Simon, Lisa Marie, et al. "In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs." *Nucleic acids research* 47.13 (2019): 7003-7017.

Figure 11a: Li, Pan, et al. "Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity." *Cell host & microbe* 24.6 (2018): 875-886.

**Figure 11b:** Adams, Rebecca L., Nathan Pirakitikulr, and Anna Marie Pyle. "Functional RNA structures throughout the hepatitis C virus genome." *Current opinion in virology* 24 (2017): 79-86.

**Figure 12:** Wacker, Anna, et al. "Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy." *Nucleic acids research* 48.22 (2020): 12415-12435.

**Figure 13:** Schlick, Tamar, et al. "Structure-Altering Mutations of the SARS-CoV-2 Frameshifting RNA Element." *Biophysical journal* (2020).

**Figure 14:** Haniff, Hafeez S., et al. "Target-Directed Approaches for Screening Small Molecules against RNA Targets." *SLAS DISCOVERY: Advancing the Science of Drug Discovery* (2020): 2472555220922802.

**Figure 15a/15b:** Perez-Gonzalez, Cibran, Daniel A. Lafontaine, and J. Carlos Penedo. "Fluorescence-based strategies to investigate the structure and dynamics of aptamer-ligand complexes." *Frontiers in chemistry* 4 (2016): 33.

**Figure 15c/15d:** Haniff, Hafeez S., et al. "Target-Directed Approaches for Screening Small Molecules against RNA Targets." *SLAS DISCOVERY: Advancing the Science of Drug Discovery* (2020): 2472555220922802.

**Figure 16:** Haniff, Hafeez S., et al. "Target-Directed Approaches for Screening Small Molecules against RNA Targets." *SLAS DISCOVERY: Advancing the Science of Drug Discovery* (2020): 2472555220922802.

**Figure 17:** Childs-Disney, Jessica L., et al. "A massively parallel selection of small molecule-RNA motif binding partners informs design of an antiviral from sequence." *Chem* 4.10 (2018): 2384-2404.

**Figure 18:** Haniff, Hafeez S., et al. "Target-Directed Approaches for Screening Small Molecules against RNA Targets." *SLAS DISCOVERY: Advancing the Science of Drug Discovery* (2020): 2472555220922802.

**Figure 19a:** Hermann, Thomas. "Viral RNA targets and their small molecule ligands." *RNA Therapeutics*. Springer, Cham, 2017. 111-134.

**Figure 19b:** Sztuba-Solinska, Joanna, et al. "Identification of biologically active, HIV TAR RNAbinding small molecules using small molecule microarrays." *Journal of the American Chemical Society* 136.23 (2014): 8402-8410.

**Figure 20:** Prado, Silvia, et al. "Bioavailable inhibitors of HIV-1 RNA biogenesis identified through a Rev-based screen." *Biochemical Pharmacology* 107 (2016): 14-28.

**Figure 21:** Hermann, Thomas. "Viral RNA targets and their small molecule ligands." *RNA Therapeutics*. Springer, Cham, 2017. 111-134.

**Figure 22a:** Hermann, Thomas. "Viral RNA targets and their small molecule ligands." *RNA Therapeutics*. Springer, Cham, 2017. 111-134.

**Figure 22b/22c:** Hermann, Thomas. "Small molecules targeting viral RNA." *Wiley Interdisciplinary Reviews: RNA* 7.6 (2016): 726-743.

**Figure 23a:** Haniff, Hafeez S., et al. "Target-Directed Approaches for Screening Small Molecules against RNA Targets." *SLAS DISCOVERY: Advancing the Science of Drug Discovery* (2020): 2472555220922802.

**Figure 23b/23c:** Hermann, Thomas. "Small molecules targeting viral RNA." *Wiley Interdisciplinary Reviews: RNA* 7.6 (2016): 726-743.

**Figure 24:** Park, So-Jung, Yang-Gyun Kim, and Hyun-Ju Park. "Identification of RNA pseudoknotbinding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening." *Journal of the American Chemical Society* 133.26 (2011): 10094-10100.

**Figure 25:** Haniff, Hafeez S., et al. "Targeting the SARS-CoV-2 RNA genome with small molecule binders and ribonuclease targeting chimera (RIBOTAC) degraders." *ACS central science* 6.10 (2020): 1713-1721.

**Table 2:** Mailler, Elodie, et al. "The evolution of RNA structural probing methods: From gels to next-generation sequencing." *Wiley Interdisciplinary Reviews: RNA* 10.2 (2019): e1518. <u>Modified</u>