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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Progress and outlook in structural biology of large viral RNAs

William A. Cantara 1, Erik D. Olson 1, Karin Musier-Forsyth *

Department of Chemistry and Biochemistry, Center for Retrovirus Research, Center for RNA Biology, The Ohio State University, Columbus, OH 43210, United States

ARTICLE INFO

Article history:
Available online 21 June 2014

Keywords:
Viral RNA
RNA structure
Viral translation
Genome packaging
tRNA-like structures

ABSTRACT

The field of viral molecular biology has reached a precipice for which pioneering studies on the structure of viral RNAs are beginning to bridge the gap. It has become clear that viral genomic RNAs are not simply carriers of hereditary information, but rather are active players in many critical stages during replication. Indeed, functions such as cap-independent translation initiation mechanisms are, in some cases, primarily driven by RNA structural determinants. Other stages including reverse transcription initiation in retroviruses, nuclear export and viral packaging are specifically dependent on the proper 3-dimensional folding of multiple RNA domains to recruit necessary viral and host factors required for activity. Furthermore, a large-scale conformational change within the 5′-untranslated region of HIV-1 has been proposed to regulate the temporal switch between viral protein synthesis and packaging. These RNA-dependent functions are necessary for replication of many human disease-causing viruses such as severe acute respiratory syndrome (SARS)-associated coronavirus, West Nile virus, and HIV-1. The potential for antiviral development is currently hindered by a poor understanding of RNA-driven molecular mechanisms, resulting from a lack of structural information on large RNAs and ribonucleoprotein complexes. Herein, we describe the recent progress that has been made on characterizing these large RNAs and provide brief descriptions of the techniques that will be at the forefront of future advances. Ongoing and future work will contribute to a more complete understanding of the lifecycles of retroviruses and RNA viruses and potentially lead to novel antiviral strategies.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

While significant progress has been made toward understanding the complex molecular choreography involved in viral replication, during recent years a bottleneck has emerged in the form of poorly understood mechanisms involving viral RNAs. The importance of understanding viral molecular biology goes beyond development of antiviral therapies. First, the dependence and co-evolution of viruses with their hosts along with host mechanisms inherent to viral infection have given unique insights into host cellular processes such as cap-independent translation (Filbin and Kieft, 2009; Miller et al., 2007; Nicholson and White, 2011) and frameshifting (Dinman, 2012; Giedroc and Cornish, 2009), as well as other basic molecular biology (Barozzi et al., 2007; Hershey and Chase, 1952; Roulston et al., 1999). Also, many technological developments have been driven by the need for novel methods of characterizing viral function (described herein). Finally, questions about immune system function and the evolutionary race between hosts and viruses with high mutation rates, particularly RNA viruses and retroviruses, are beginning to be deciphered (Cascalho and Platt, 2007; Ding and Voinnet, 2007; Iyer et al., 2006; Sanjuan et al., 2010). However, to date viral structural biology has mostly focused on the actions of viral proteins and their interactions with host factors.

Methods of characterizing large RNA structures are constantly improving, and there are large gaps in our understanding of viral replication that such structural data can help to fill. The small size of virus particles requires that, in many cases, a single RNA element can play a role in multiple distinct functions, thus optimizing virus replication. In the case of RNA viruses and retroviruses, the RNA genome acts as both the genetic material and as an active participant in many stages of the lifecycle. For example, specific functional domains of genomic and differentially spliced viral mRNAs have shown the ability to regulate cap-independent translation initiation mechanisms (Filbin and Kieft, 2009; Miller et al., 2007; Nicholson and White, 2011), transcription/translation (Dreher, 2009; Fechter et al., 2001), reverse transcription in
retroviruses (Lu et al., 2011b), cellular localization (Hammarskjold and Rekosh, 2011; Stake et al., 2013), programmed frameshifting (Dinman, 2012; Giedroc and Cornish, 2009) and viral packaging (Dreher, 2009; Fechter et al., 2001) (Fig. 1).

Viruses with RNA genomes constitute a number of human pathogens such as influenza virus, SARS-coronavirus, hepatitis C virus, West Nile virus, HIV-1 and HTLV-1. Indeed, there have been reports linking highly-structured viral RNA genomes to Persistence of infection in hosts (Davis et al., 2005). Viral RNA functions, although critical to viral lifecycle progression and infectivity, have been largely neglected in terms of antiviral therapy due in large part to our lack of high quality 3D structural data. We propose that improvements in technology and methods of RNA structure determination will revolutionize our understanding of viral replication and lead to novel antiviral strategies aimed at disrupting RNA-centric functions. Knowledge of viral RNA structure has already led to improvements in potential antiviral therapies. For example, information regarding the structured regions of the HIV-1 genome was used to help design more effective shRNA inhibitors of viral replication (Low et al., 2012), and a crystal structure of a hepatitis C virus inhibitor bound to a portion of the viral internal ribosome entry site (IRES) revealed the binding pocket and has led to subsequent improvements of the inhibitor (Dibrov et al., 2012; Zhou et al., 2013). Therefore, it is important to assess the current state of the field and address potential schemes for moving the field forward. In this review, we first describe recent progress that has been made toward solving large viral RNA structures, with a focus on non-enzymatic elements. As such, we will not be discussing viral ribozymes or other catalytic RNAs (reviews of this topic include Talini et al., 2009, 2011; Wu et al., 2009). Secondly, we highlight common methods of characterizing large viral RNA structure and provide advantages, disadvantages and recent advances for each.

2. Current progress in characterization of large viral RNA structure

2.1. Internal ribosome entry site (IRES) elements

IRES elements are found in viral RNAs that either require, or are capable of proceeding by cap-independent mechanisms for recruiting and engaging the ribosome. IRESes are employed across many viral families and genera. While all IRESes ultimately function as cis-acting enhancers of RNA translation, there is great variety in IRES structures and number of cofactors required. Presently, no consensus IRES sequence/structure has been observed, and there appears to be an inverse correlation between degree of IRES foldedness and number of cofactors required (reviewed in Filbin and Kieft, 2009). IRESes from the viral family Dicistroviridae are highly structured and able to recruit the 80S ribosome efficiently without the aid of any eukaryotic initiation factors (eIFs) or IRES trans-acting factors (ITAFs) (Wilson et al., 2000). The IRES structures of several Dicistroviridae family members have been characterized. For example, the Cricket paralysis virus (CrPV) IRES in complex with the 80S ribosome was solved to ~17 Å resolution using cryo-electron microscopy (cryo-EM) (Spahn et al., 2004), revealing tertiary structure characterized by an elongated, yet tightly defined shape that is able to manipulate the ribosome upon binding. A high-resolution crystal structure of the related Plautia staliinistitnine virus (PSIV) IRES alone was later solved, revealing that the RNA contained both rigid and flexible regions, explaining the observed order of ribosomal subunit recruitment (Pflügsten et al., 2006). A domain of the CrPV IRES containing a critical pseudoknot motif was also crystallized and revealed a striking RNA–mRNA structural mimicry, providing a structural explanation for its role in binding the P-site of the ribosome (Costantino et al., 2008). The importance of this structural mimicry was validated by additional crystal structures of CrPV...
and PSIV IRES domains bound to the 70S ribosome (Zhu et al., 2011). Recently, structures of the CrPV IRES bound to the ribosome were determined to atomic-level resolution using single-particle cryo-EM (Fernandez et al., 2014). The structures showed that IRES binding initially occurs at the ribosomal A site in a pretranslocation conformation, with translocation into the P site required to move the first codon of the IRES-associated message into the A site, allowing translation to initiate.

The successful characterization of the entire Dicistroviridae IRES structures is due, in part, to their relatively high degree of foldedness. IRESes from Hepatitis C virus (HCV) and related viruses are able to initiate translation internally, but are generally less structured and require additional cellular factors not needed by CrPV or PSIV-like IRESes (Otto and Puglisi, 2004; Pinede and Martinez-Salas, 2012). As such, high-resolution structural data only exists for subdomains of the HCV IRES. The focus of this review is large viral RNA structures, and so these works will not be extensively discussed (see also Filbin and Kieft, 2009; Lukavsky, 2009). Early probing and small-angle X-ray scattering (SAXS) characterization of the full-length HCV IRES confirmed that it folded into a defined tertiary structure of a non-globular nature (Kieft et al., 1999). Cryo-EM structures of the full-length HCV IRES bound to the 40S (Spahn et al., 2001) and 80S (Boehringer et al., 2005) ribosomes have been solved to ~20 Å and ~25 Å resolution, respectively. The 40S-IRES structure reveals that addition of the IRES is sufficient to alter the conformation of the 40S subunit, underscoring the active role that the RNA plays in initiating its own translation. The IRES-80S complex shows that upon 60S association, both the IRES and the ribosome structures are significantly altered. Siridechadilok and coworkers determined a cryo-EM structure of the HCV IRES in complex with elf3 (a required cofactor for 40S recruitment by this class of IRES), revealing that the IRES interacts with elf3 at the same position as the cap-dependent pathway cofactor elf4G (Siridechadilok et al., 2005). Furthermore, modeling this structure with the 40S suggests that the HCV IRES effectively mimics the positioning of elf4G, essentially acting as a functional replacement in translation initiation. Recently, the final subdomain of the HCV IRES was solved at high resolution by X-ray crystallography (XRC), and the structures were assembled into the available cryo-EM envelope of the full-length IRES revealing that they all fit well (Berry et al., 2011). This structure further supported the notion that the HCV IRES functions by specific positioning of the start codon in the ribosomal P-site. The 40S-elf3-CSFV IRES (Classical swine fever virus, a close HCV relative) ternary complex was solved to ~9 Å resolution by cryo-EM, showing that, in this system, the IRES binds the 40S subunit in the same position as elf3, relegateing elf3 to binding only with the IRES itself (Hashem et al., 2013). This structure helps rationalize how CSFV infection efficiently hijacks cellular translation in favor of viral translation; it not only competitively displaces a necessary cellular translation cofactor in favor of a viral IRES, it provides a novel site for the cofactor to dock and become sequestered. In all of the HCV IRES-complex structures, the RNA appears to adopt a single, elongated conformation. However, a recent study employing SAXS in conjunction with molecular dynamics suggests that the HCV apolIRES is composed of rigid parts that move independently of each other (Perard et al., 2013) (Fig. 1). Taken together, the HCV IRES demonstrates a complex and highly orchestrated mechanism of translation initiation in which structural characterization played a large role in our understanding.

**Human immunodeficiency virus type 1** (HIV-1) is capable of initiating translation via both cap-dependent and independent mechanisms (reviewed in Balvay et al., 2007). While an IRES element was initially proposed to be located within the gag ORF (Buck et al., 2001), subsequent studies have focused on the 5′UTR (Brasey et al., 2003; Plank et al., 2013; Vallejos et al., 2011); the results of these studies suggest that the gag ORF in fact inhibits IRES activity (Brasey et al., 2003; Plank et al., 2013). While the matter is still controversial, the possibility remains that both IRES locations exist, and that they each play distinct roles in the viral lifecycle (Yilmaz et al., 2006). It should be noted that the HIV-2 gag ORF has been unambiguously identified as an IRES, actively recruiting translation-initiation factors to the RNA (Herbreteau et al., 2003; Locker et al., 2011). Different splice variants of HIV-1 mRNAs containing the 5′ UTR have all been shown to possess the same IRES secondary structure, suggestive of a common mechanism for translational control (Charnay et al., 2009; Plank et al., 2014). In contrast to the HCV IRES, which specifically positions the start codon in the P-site of the ribosome, the HIV-1 IRES appears to function by first recruiting initiation factors to the RNA, after which they can scan until finding a start site (Plank et al., 2014). HIV-1 IRES activity is enhanced by still unidentified cellular cofactors, present only in the G2/M phase of the cell cycle, suggesting a specific timing when HIV-1 IRES activity is required (Vallejos et al., 2011). A clear picture has yet to emerge as to which mutations the HIV-1 IRES can tolerate while still maintaining functionality (Brasey et al., 2003; Plank et al., 2013, 2014; Vallejos et al., 2011), but overall it appears as though the RNA plays a more passive role than other viral IRESes. Consistent with this, HIV IRES activity requires more cellular factors than HCV, for example. Although the precise sequence that constitutes the HIV-1 IRES is not clearly defined, 3D structures of elements within the 5′ UTR have been elucidated (reviewed in Lu et al., 2011b). Recently, our lab used SAXS in conjunction with MD and simulated annealing to generate models of three ~100-nucleotide (nt) domains of the 5′ UTR, revealing that the primer–binding site (PBS) region mimics the 3D structure of a tRNA (Jones et al., 2014) (Fig. 1). It is possible that this tRNA-like element (TLE) interacts with the ribosome, although it has also been shown to bind lysyl-tRNA synthetase, which is co-packaged into viral particles along with the HIV-1 reverse transcription primer, tRNA^Lys (Jones et al., 2013). Further studies will be required to fully define the role(s) that the TLE plays in the viral lifecycle. Future efforts will likely be aimed at precisely determining the necessary cofactors for HIV-1 IRES activity, and further characterizing the whole 5′ UTR structure alone and in complexes. A number of studies have shown that translation initiation can also occur via ribosomal scanning in HIV-1 (Berkhout et al., 2011; Miele et al., 1996), and which mechanism is dominant is not entirely clear. It remains possible that HIV-1 uses both mechanisms, and additional conditions such as cell cycle progression or stress may factor into the choice. The Picornaviridae family comprises the final class of viral IRESes, of which Foot-and-mouth disease virus (FMDV) is an exemplar. This class is the least folded, requires the greatest number of cellular cofactors, and cannot even bind the 40S subunit on its own (reviewed in Filbin and Kieft, 2009). The tertiary structure of this class of IRESes alone or in relevant complexes is unknown; however, it has been suggested that tertiary contacts are indeed formed and are important for FMDV IRES activity (Fernandez-Miragall et al., 2006).

### 2.2. 3′ Cap-independent translation elements (CITEs)

The 3′ cap-independent translation element (CITE) is an RNA structure found within 3′ UTRs of plant viruses. These RNA motifs enhance translation of the viral genome, but are considered distinct from IRESes based on their location in the 3′ UTR. The mechanisms by which CITEs act to promote translation remain poorly understood, but interactions with 5′ UTRs, ribosomal subunits, and translation initiation factors have been described and are thought to be involved in enhancing translation (reviewed in Miller et al., 2007; Nicholson and White, 2011). To date, structural analysis of CITEs has largely been limited to examination of their secondary structures, with a few notable exceptions. The 3′ UTR of the Turnip
crinkle virus (TCV) contains a kissing-loop-T-shaped structure class of CITE that interacts with the 60S ribosomal subunit. Through a combination of SAXS and nuclear magnetic resonance (NMR) spectroscopy, a low-resolution structure was determined for a 102-nt RNA corresponding to the TCV CITE. This element was found to mimic the overall shape and certain nt motifs of tRNA (Zuo et al., 2010) (Fig. 1). This experimental structural study largely confirmed previous molecular modeling work that predicted that the tertiary structure would resemble tRNA (McCormack et al., 2008). These results also support data suggesting that the TCV CITE interacts with the 60S subunit via the P-site in a manner that mimics tRNA, and that this interaction is partly responsible for the observed translational enhancement (Stupina et al., 2008). In contrast, the Panicum mosaic virus-like (PMV) CITE class has been shown to interact with the initiation factor eIF4F via its eIF4E subunit, an interaction whose affinity correlates positively with the CITEs translational enhancement capabilities (Wang et al., 2009b). Furthermore, besides the TCV CITE, the PMV CITE is the only other CITE for which a 3D structure has also been elucidated (Wang et al., 2011). In this study, Wang and coworkers determined the secondary structures of CITEs from several viral genera using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) probing and RNA footprinting experiments, both in the presence and absence of eIF4E. The authors then built computational models of the apoCITE and docking models of CITE-eIF4E complexes using constraints derived from the probing experiments and previous literature. Interestingly, the PMV CITE interacts with eIF4E in a manner that resembles the interaction between eIF4E and the 5'-m'GpppN cap. Efforts are underway to examine the barley yellow dwarf virus CITE (representative of another class) by X-ray crystallography (Kraft et al., 2011). Future structural work in this field promises not only to improve our understanding of how this diverse set of RNA elements function, but also to reveal what makes them similar/different from the closely related IRES element (Nicholson and White, 2011).

2.3. Plant viral tRNA-like structures (TLSs)

The extreme 3' end of a number of (+) strand RNA plant viruses have been shown to mimic tRNA structures. These so-called tRNA-like structures (TLSs) are aminocoylated in vivo, and this activity has been linked to regulating translation vs. transcription of the viral genome, maintenance of its 3' end, and encapsidation of the RNA (reviewed in Dreher, 2009; Fechter et al., 2001). A majority of structural studies have focused on the Turnip yellow mosaic virus (TYMV) TLS, and through various probing studies combined with computational modeling found that it bares a striking resemblance to tRNA in terms of overall fold and tertiary interactions, despite differences in base pairing (de Smit et al., 2002; Dumas et al., 1987; Giegé et al., 1993). An NMR structure revealed that the 3' half of the TLS folds into a coaxially-stacked H-type pseudoknot that mimics the overall tRNA acceptor stem/T arm structure (Kolk et al., 1998). SAXS analysis confirmed that the RNA element mimics the overall shape of tRNA (Hammond et al., 2009). Further SAXS structures of the TYMV TLS showed that an upstream RNA element required for efficient aminocoylation folds as an independent domain, and suggested a role for the upstream domain in acting as a molecular switch, modulating TLS interactions with its protein partners (Hammond et al., 2010). A recent crystal structure confirmed many of the previous observations, and showed that the TYMV TLS strikingly resembles tRNA, despite being composed of distinct tertiary interactions (Colussi et al., 2014). Interestingly, the TLS contains one face that precisely mimics tRNA structure, while the other face is more divergent, providing a structural explanation for the ability of this RNA to serve as a substrate for aminocoylation, as well as in other non-tRNA roles.

The related Tobacco mosaic virus (TMV) and Brome mosaic virus (BMV) TLS tertiary structures have also been modeled. Early 3D models of the BMV TLS (Felden et al., 1994) (Fig. 1) and TMV TLS (Felden et al., 1996) generated based on RNA structure-probing suggested that both TLSs mimicked the canonical L shape of a tRNA and contained pseudoknot interactions reminiscent of the D–T interaction in tRNA. A more recent examination using SAXS and probing showed that, while both BMV and TMV TLSs retained some tRNA-like tertiary contacts, they did not fold into a static conformation like the TYMV TLS, but instead remained relatively dynamic in solution (Hammond et al., 2009). This result is interesting in light of the fact that all three TLSs are competent to be aminocoylated; further studies will be required to determine if this difference in RNA dynamics bears any significance on the viral lifecycle.

2.4. Nuclear export signals

A hallmark of retroviruses is reverse transcription of viral genomic RNA (vRNA) and integration into the host genome. Retroviruses hijack host transcriptional machinery to generate transcripts coding for viral proteins. Messages encoding longer retroviral proteins (such as Gag, Pol, and Env) must be exported from the nucleus containing some or all of their introns. A number of retroviruses bypass this hurdle by encoding cis-acting RNA elements that interact with nuclear export factors either directly, or indirectly via viral proteins (reviewed in Hammerskjöld and Rekosh, 2011; Stake et al., 2013). In the case of HIV-1, the virally encoded Rev protein oligomerizes on the Rev Response Element (RRE), subsequently engaging the Crm1 nuclear export pathway and localizing the unspliced or partially spliced viral RNAs to the cytoplasm (Stake et al., 2013). Efforts to characterize the HIV-1 RRE and Rev-RRE structures in their entirety have been hindered by the size of the RRE (~350 nt) and the propensity for Rev to aggregate. Early efforts were limited to characterization of short Rev-derived peptides bound to the high affinity RRE stem loop IIB solved by NMR (Battiste et al., 1996; Gossler et al., 2001; Peterson and Feigon, 1996; Ye et al., 1996) (Fig. 1), while the apo stem loop IIB structure was elucidated using X-ray crystallography (Ippolito and Steitz, 2000) and NMR (Peterson and Feigon, 1996). While there are minor discrepancies between the various studies, they all provide a clear structural rationale for Rev's critical interaction with a non-canonical G/G/C/A base pair motif in the RRE, in what is thought to be the nucleating event of Rev oligomerization. However, it is known that the entire RRE is required for functionality in vivo (Hammerskjöld and Rekosh, 2011). A more recent study using atomic force microscopy (AFM) characterized the full-length RRE by itself and bound to Rev protein (Paijlesen et al., 2009), showing that multiple copies of Rev coated the entire RRE. Protein engineering efforts by the Frankel laboratory yielded Rev mutants whose oligomeric state was controlled, leading to a SAXS structure of a full-length Rev dimer bound to RRE stem loop IIB (Daugherty et al., 2010a). In the same study, the authors examined the full-length RRE:Rev complex by electron microscopy (EM), showing that the RNP assembly was consistent with a single hexameric Rev formed on one RRE. The finding is significant as it suggests an active role of RRE in determining the extent of Rev oligomerization; indeed, in the absence of RRE, Rev forms heterogeneous oligomeric complexes in solution (Daugherty et al., 2010a). Recently, SAXS was used to characterize a ~230-nt RRE construct including stem loops IIB and IA, generating molecular envelopes and an ensemble of atomic models showing that the RRE fluctuates between “A” and “H” shaped conformations (Fang et al., 2013). Furthermore, the SAXS structure is consistent with the distance between the RNA binding domains of the Rev dimer (Daugherty et al., 2010b; DiMattia et al., 2010), suggesting a model where a Rev
dimer interacts with both helices of the RRE to nucleate a controlled oligomerization.

Much of the work toward determining the tertiary structure of viral nuclear export signals has focused on HIV-1, but a few other viruses have been studied. The only 3D structure of the HIV-2 RRE comes from a homology model based on the SHAPE-probed secondary structure (Lusvarghi et al., 2013a); the structure resembles the “A”- and “H”-shaped folds observed for HIV-1 RRE (Fang et al., 2013). This result is consistent with the observation that HIV-1 Rev is capable of exporting RNA containing either the HIV-1 or HIV-2 RRE (Dillon et al., 1990). Additional 3D characterization of nuclear export elements from other viruses will be needed to understand the structural basis for why certain viruses are able to interact directly with host export machinery, while others require viral specific adaptor molecules.

2.5. Frameshift signals

Programmed ribosomal frameshifting is a common mechanism employed by viruses to expand their typically limited genetic information. For example, many retroviruses encode their pol or gag ORFs without a start codon. These genes must be translated as the product of a highly controlled-1 ribosomal-frameshifting event that leads to a certain percentage of the gag gene stop codon being shifted out-of-frame, so that the ribosome may continue and produce gag-pol or gag-pro proteins. Furthermore, precise control of frameshift gene product levels has been linked to optimal viral function. These kinds of controlled-frameshifting events occur in a number of other viruses as well. Frameshift signals generally require two elements: a heptanucleotide “slippery sequence”, where the actual frameshifting occurs, and a structured RNA element 6–8 nt downstream that can stall the ribosome (the overall topic is reviewed in Dimman, 2012; Giedroc and Cornish, 2009). Here, we will focus on work characterizing the 3D structures of the RNA elements that induce frameshifting of the ribosome. Given their relatively small size (∼25–40 nt), there has been a great deal of success in high-resolution characterization of these elements.

Viral ribosome pausing elements generally fall into one of two structural categories: pseudoknot-containing elements and hyperstable stem loop structures. The Mouse mammary tumor virus (MMTV) was the first frameshift signal whose structure was determined at high resolution by NMR (Shen and Tinoco, 1995). This pseudoknot-containing frameshift signal was found to form two distinct coaxially-stacked helices punctuated by an intercalated adenosine that causes a distinct bend in the structure. Within the context of the MMTV system, this bend was shown to be critical for proper function, as mutations causing the pseudoknot to adopt a linear conformation (as confirmed by NMR) abrogated frameshifting activity (Chen et al., 1996). A crystal structure of the Beet western yellow virus (BYYV) frameshift pseudoknot solved to 1.6 Å resolution showed conservation of many of the structural elements observed in MMTV; the distinct bend and the characteristic triple-helix interactions were also present in the BYVV system (Egli et al., 2002; Su et al., 1999). A number of other viral frameshift signals have also been studied: NMR structures of the Pea enation mosaic virus (Nixon et al., 2002) and Sugarcane yellow leaf virus pseudoknots (Cornish et al., 2005), a crystal structure of the Potato leaf roll virus pseudoknot (Pallan et al., 2005), and a mass spectrometric three-dimensional (MS3D)-derived structure of the Feline immunodeficiency virus pseudoknot (Yu et al., 2005) have all been determined. While some differences exist in the nt composition and the specific interactions made, all of these models (determined by a variety of techniques) converge on a common overall fold; that is, one characterized by a distinctly bent conformation, and the presence of a triple helix formed by triple and quadruple base-pairing interactions. However, the bent pseudoknot motif is not universal, and several subsequent studies demonstrated that functional frameshift elements may adopt a linear conformation as well. NMR structures of the T2 bacteriophage pseudoknot (Du et al., 1996) and the Simian retrovirus-1 (SRV-1) frameshift signal (Du et al., 1997; Michiels et al., 2001) show this clearly. Nonetheless, these linear pseudoknots share many of the same features as their bent counterparts, such as a triple-helix motif.

The second class of frameshift elements is the hyperstable stem loop. HIV-1 is an example of this class and NMR structures of this frameshift signal have been solved (Gaudin et al., 2005; Staple and Butcher, 2005) (Fig. 1). While there was initially some controversy as to whether the HIV-1 frameshift element folded into a pseudoknot or discrete stem loop, these NMR structures unambiguously showed that only the stem loop conformation formed, at least in vitro in the absence of (viral) proteins. The stem loop folds into two distinct stems of roughly equal size that are bent by an internal bulge in a manner reminiscent of the bent-pseudoknot motif. An NMR structure of the Simian immunodeficiency virus (SIV) frameshift element was also reported, showing that while it too folds into a stem loop, it is not bent as is the case for HIV-1 (Marcheschi et al., 2007). Rather, the stem is coaxially stacked throughout and the apical-loop residues contain a number of H-bonding contacts, adding further stability.

Given the extremely diverse structural nature of frameshift elements, the question has naturally arisen as to what features govern the capability of these RNAs to cause the ribosome to stall and frameshifting to occur. The main hypotheses that have been favored are that structural stability of the frameshifting element, specific contacts with the ribosome by the frameshift element, or some combination of the two is critical for productive pausing of the ribosome. A consensus has not yet been reached (for more detailed discussion of this topic, see Firth and Brierley, 2012; Giedroc and Cornish, 2009). Clearly more structural data of frameshift elements bound to the ribosome would shed light on the debate. A low-resolution cryo-EM structure of an 80S ribosome paused at the Coronavirus frameshift pseudoknot contained bound elongation factor eEF2 in the A-site and an uncharacteristically bent tRNA at the P-site (Namy et al., 2006). This led the authors to propose a frameshifting mechanism based on the apparent strain that the pseudoknot was inducing within the ribosome. Higher-resolution structures of frameshift elements interacting with ribosomes are needed to establish whether there are any specific contacts made, or if the pausing is caused purely by intrinsic RNA stability.

2.6. Viral packaging signals

During the late stages of infection, viruses must assemble all of the required viral and host cell components and then egress from the cell. A variety of mechanisms exist to ensure that the correct proteins and nucleic acids are packaged into the nascent virions. Specific interactions occur between viral proteins and RNAs/DNAs in the assembly process, leading to packaging of the vRNA. In retroviruses, interactions occur between the Gag protein and a structured element termed the Psi (Ψ)-packaging signal within the vRNA that lead to selective incorporation of vRNA into virions. Retroviruses are exquisitely capable of packaging only a single unspliced and dimerized vRNA despite the presence of a vast excess of cellular RNAs and spliced vRNAs; understanding the nature of this recognition has been intensely investigated but is still incompletely understood (reviewed in D’Souza and Summers, 2005; Kuzembayeva et al., 2014). The Moloney murine leukemia virus (MLV) core-packaging signal has been defined, and has been shown to direct encapsidation of heterologous vRNAs into which it is transplanted (D’Souza and Summers, 2005). Employing a segmental isotopic-labeling strategy, NMR structures of the ~100-nt MLV Ψ-packaging signal in the apo (D’Souza et al., 2004) (Fig. 2)}
and NC-bound forms (D’Souza and Summers, 2004) have been determined. The apo structure revealed that two of the stem loops coaxially stacked with one another, forming an extended stem loop; the third stem loop formed independently and bifurcated the other two (D’Souza et al., 2004). The structure of the NC-RNA complex showed that a single NC molecule bound to a single-stranded region near the junction of the three stem loops that is only available if the packaging element is in the dimeric conformation, thus providing a structural explanation for the observation that only dimeric RNA is encapsidated (D’Souza and Summers, 2004). An NMR and cryo-ET characterization of dimeric MLV Ψ confirmed this result (Miyazaki et al., 2010). The minimal Rous sarcoma virus (RSV) Ψ element has also been identified, and an NMR structure of the 80-nt RNA bound to NC was reported (Zhou et al., 2007). Like MLV, the RSV-packaging element is composed of three stem loops, two of which are coaxially stacked, with the third forming independently. However, RSV NC was found to interact with two stem loops of the packaging signal via each of its zinc knuckles rather than at one site as in MLV.

In contrast to both MLV and RSV, the minimal packaging signal of HIV-1 has been more difficult to identify. Conflicting reports exist as to whether the entire 5’UTR plus part of the gag ORF is required to encapsidate a heterologous RNA, or if a smaller region is sufficient (reviewed in D’Souza and Summers, 2005; Kuzembayeva et al., 2014). The general consensus seems to be that the entire 5’UTR region is necessary, but efforts to determine the minimal signal are still underway (Heng et al., 2012). Within the 5’UTR, four stem loop (SL) elements (SL1–SL4) are generally believed to be necessary for HIV-1 vRNA packaging. The structure of the individual stem loops of HIV-1 Ψ (SL1–SL3) in both the apo and NC-bound states have been extensively characterized by NMR (as reviewed in Lu et al., 2011b). A variety of studies using different techniques have attempted to elucidate the structure of the entire Ψ domain. An MS3D-based model of Ψ showed that the RNA adopts a compact conformation with tertiary contacts occurring between the apical loop of SL4 and an internal bulge within SL1 and the loops of SL2 and SL3 (Yu et al., 2008a). However, subsequent studies using Förster resonance energy transfer (FRET)-derived distance restraints and molecular dynamics (MD) (Stephenson et al., 2013), as well as SANS in conjunction with MD and simulated annealing (Jones et al., 2014), suggested that Ψ lacked tertiary contacts, and that SL1 and SL3 coaxially stack on each other with SL2 bifurcating. Interestingly, there is a high degree of similarity between the overall structures of MLV and HIV-1 Ψ (Fig. 2), suggesting that packaging elements from a diverse set of retroviruses may share a conserved 3D shape. A structure of NC bound to the larger HIV-1 Ψ element has yet to be determined, so it is possible that the apORNA fluctuates dynamically between a compact and open conformation, and that NC binding stabilizes one or the other.

Similar to the molecular switch observed in MLV Ψ, it has been proposed that the larger HIV-1 5’UTR alternates between a packaging-competent and non-packaging-competent conformation (Abbink and Berkhout, 2003; Abbink et al., 2005; Berkhout et al., 2002; Huthoff and Berkhout, 2001; Kenyon et al., 2013; Lu et al., 2011a; Ooms et al., 2004a,b; Seif et al., 2013). In the long-distance interaction (LDI) conformation, SL1 of Ψ interacts with an upstream region within the 5’UTR, preventing dimerization, while in the branched multiple-helix (BMH) conformation, SL1 is exposed. Efforts are underway to determine the structure of the 5’UTR in both conformations, and these results should further our understanding of why only dimeric vRNAs are packaged in HIV-1. AFM images of dimeric 5’UTRs suggest the formation of two intermolecular contacts, presumably between the Ψ SL1 dimerization initiation site (DIS) loops, and between the trans-activation region (TAR) loops (Andersen et al., 2004; Pallesen, 2011). A SAXS-derived 3D structure of the TAR and polyA hairpins (Fig. 1) is consistent with this possibility, as the stem loops coaxially stack and the TAR loop is available for intermolecular interactions (Jones et al., 2014). The same study showed that the SL1 hairpin loop of an HIV-1 Ψ variant that was mutated to prevent dimerization, is solvent exposed and available for intermolecular contacts. However, higher-resolution structures of the dimeric RNA will be required to confirm the presence and exact nature of these interactions.

3. Techniques for structural characterization of large functional viral RNAs and RNP

Our understanding of the molecular mechanisms underlying viral replication is currently incomplete due to limited knowledge of the tertiary structure of genomic RNA elements and their complexes with host and viral proteins. The large size and inherent conformational flexibility of these functional RNAs and ribonucleoprotein complexes (RNP) significantly reduces the effectiveness of conventional methods such as NMR and XRC. However, recent advances in methods and technology have made structural characterization of large RNAs and RNPs feasible. Comprehensive descriptions of the technologies have been reviewed extensively and will be referred to in the text. Here, we will focus on applications to large RNA structure determination. Each method is characterized by inherent size and resolution limitations (Fig. 3), as well as specific advantages, disadvantages, challenges and recent advances that will also be discussed.

All of the methods reviewed here suffer from a few common challenges that must be overcome. First, sufficient quantities of the RNA of interest must be obtained in a pure form. In vitro transcription using T7 RNA polymerase generally provides efficient synthesis of RNAs up to ~400–500 nt; however, since the polynucleotide commonly adds one (n + 1) or two (n + 2) extra nt to the 3’ end, yields can be significantly reduced by virtue of the requirement for RNAs with homogeneous ends. Denaturing polyacrylamide gel electrophoresis generally affords single-nt resolution of RNAs up to ~50 nt (Lu et al., 2010; Reyes et al., 2009). Alternative tactics must be used for RNAs of larger size. Such approaches include the use of ribozyme cassettes flanking the sequence of interest.
Fig. 3. General size and resolution limitations of common approaches to RNA structure determination. NMR spectroscopy can yield atomic-resolution (∼1–5 Å) structures, but has the most restrictive size limitations for RNA (∼3–35 kDa). XRC is also capable of atomic-resolution (∼1–5 Å) and can be readily applied to RNP structures up to 10MDa with the caveat that they must be capable of forming large, well-ordered crystals. Cryo-EM has not been used specifically for RNA structure determination but has demonstrated the ability to achieve sub-nanometer resolution (up to 3 Å) in certain circumstances, however structures must be large enough to provide ample phase contrast (∼200 kDa to 20,000 MDa). SAXS bridges the gap between NMR and cryo-EM in size limitations (∼10 kDa to 3500 MDa) at the cost of resolution (currently ∼10–50 Å).

(Batey and Kieft, 2007; Di Tomasso et al., 2012, 2011; Ferrer-D’Amare and Doudna, 1996; Salvail-Lacoste et al., 2013; Walker et al., 2003, trans-acting ribozymes (Ferrer-D’Amare and Doudna, 1996; Perrotta and Been, 1992), RNase H (Inoue et al., 1987; Lapham and Crothers, 1996), RNase P (Ziehler and Engelke, 1996) or trans-acting DNazymes (Santoro and Joyce, 1997). Additionally, it has recently been shown that Marine cyanophage Syn5 RNA polymerase exhibits significantly reduced number of n + 1/n + 2 products during run-off transcription (Zhu et al., 2014). Second, proper folding of the RNA into a functional structure must be verified and alternative conformations rigorously removed from the sample. RNAs have a propensity to become trapped in local energy minima during folding, resulting in multiple misfolded forms, which has been referred to as “alternative conformer hell” (Fedor and Westhof, 2002; Uhlenbeck, 1995). In some cases, primarily with large RNAs, thermal denaturation/renaturation strategies may prove ineffective, requiring use of native purification (Batey and Kieft, 2007; Lukavsky and Puglisi, 2004; Toor et al., 2008) or even recombinant methods of overexpressing and purifying RNA from Escherichia coli (Nelissen et al., 2012). Mutations may be introduced to aid in sample preparation, but these changes should be vetted for function. Proper function of viral RNAs often hinges on their ability to bind viral and host proteins. Therefore, binding or other activity assays should be performed with all constructs prior to structural characterization. Also, when possible, mutations should be incorporated into viral infectivity assays. In some cases, nature has already performed this experiment through genetic diversity, highlighting the importance of comprehensive sequence analysis of viral isolates prior to structural experimentation.

3.1. Chemical modification, probing and comparative sequence analysis

While a detailed description of chemical-probing techniques is outside the scope of this review (for comprehensive reviews, see Weeks, 2010 and Lu et al., 2011b), such studies contribute important details needed for efficient tertiary structure determination. Chemical probing has been used to identify backbone flexibility and base-pairing interactions at single-nt resolution (Merino et al., 2005), leading to determination of secondary structure. However, although standardized protocols have been published over the past three decades (Low and Weeks, 2010; Lusvarghi et al., 2013; McGinnis et al., 2009; Mortimer and Weeks, 2009; Steen et al., 2011; Wilkinson et al., 2006), unambiguous secondary structure assignment remains elusive for many RNAs largely due to RNA conformational heterogeneity and dynamics, with the 5′UTR of HIV-1 as a prime example (Lu et al., 2011b).

Recent successes in deriving the secondary structures and conformational changes of viral RNA genomes illustrates the usefulness of SHAPE probing (Archer et al., 2013; Grohman et al., 2013; Watts et al., 2009; Wilkinson et al., 2006). An initial estimate suggested that the addition of SHAPE-derived pseudo-free energy parameters to nearest neighbor free energy terms improved secondary structure prediction to 96–100% accuracy (Deigan et al., 2009). However, more recent rigorous benchmarking has revealed more modest improvements (false negative rates near 17% and false discovery rate near 20%) (Kladwang et al., 2011). Despite inconsistencies between probing data and high-resolution structures (Miyazaki et al., 2010), methods to improve accuracy are ongoing. These include the use of differential SHAPE methods, where reactivity differences between two reagents are used to correct for inherent biases of each reagent (Rice et al., 2014). Another approach has been to use dilution as a means of normalizing different data sets based on an internal standard, such as the number of full-length primer extension products (Kladwang et al., 2014). As with many other structural techniques, even low levels of sample heterogeneity can severely affect the accuracy of SHAPE-derived secondary structure models. To mitigate this effect and offer the ability to examine different conformers within a heterogeneous population, in-gel SHAPE has been developed (Kenyon et al., 2013). Briefly, a polydisperse sample is separated by native polyacrylamide electrophoresis followed by excision of each conformer, which can then be probed within the gel matrix.

Despite certain instances of uncertainty, chemical probing offers critical information toward efficient tertiary structure determination. First, knowledge of secondary structure allows for construction of RNAs that either retain or disrupt functional base-pairing networks. In instances of structural heterogeneity, mutations can also be introduced to stabilize desired structures or destabilize alternative, nonfunctional conformations. Also, regions of highly flexible residues correlate with difficulty in determining tertiary structure. These residues can be stabilized or eliminated from constructs to improve the quality of structural data. Finally, probing techniques can be used to generate data for direct use in molecular modeling of tertiary structures. All current methods of 3D structure determination require some degree of molecular modeling. For instance, XRC requires building a model into an electron density map that is further refined. Structure calculation from NMR involves generation of an energy-minimized model using NMR-derived distance and angle restraints. Chemical-probing data can similarly be used to constrain molecular models. For example, quantitative measures of solvent accessibility can be obtained from hydroxyl-radical probing (Pastor et al., 2000; Tuillius and Greenbaum, 2005), greatly constraining tertiary structure models. Direct distance restraints can also be probed using bifunctional crosslinking analyzed by mass spectrometry, also known as MS3D (Fabris and Yu, 2010; Yu et al., 2008a,b, 2005). Caution must be taken in analyzing and interpreting crosslinking data, as transient conformations may inadvertently be trapped, obscuring the structure of the RNA.

In addition to chemical-probing methods, nucleic acid secondary structures can also be assessed through phylogenetic and thermodynamic approaches (reviewed in Seetin and Mathews, 2012). Indeed, one of the most successful methods of secondary-structure prediction is comparative sequence analysis (CSA). CSA is predicated on the notion that structural conservation
supersedes that of primary sequence conservation. Here, polymorphisms within a large number of homologous sequences can be used to differentiate between paired or unpaired residues using covariation analysis (Eddy and Durbin, 1994). Indeed, application of CSA to ribosomal RNAs was strikingly accurate (~97%) (Gottel et al., 2002). A caveat of this technique is the need for large datasets of homologous sequences. Such large databases exist for many viral genomic RNAs such as HIV (Foley et al., 2013) and HCV (Kuiken et al., 2008). In cases where many sequence homologs are unknown or as a stand-alone method, secondary structure can be predicted based on energy minimization using the nearest neighbor model. Several tools such as Mfold (Zuker, 2003) and RNA structure (Bellaousov et al., 2013; Reuter and Mathews, 2010) are freely available for automation of such calculations. As with all models, predicted secondary-structure models must be experimentally scrutinized for accuracy. For example, site-directed mutagenesis and compensatory mutation can be used to further validate predictions. Additionally, in vivo SELEX (systematic evolution of ligands by exponential enrichment), whereby specific nt in a proviral plasmid are randomized and transfected into cells followed by multiple rounds of viral replication, can also be used to identify the importance of different structural elements in viral infectivity (reviewed in Berkhout and Das, 2009). Similarly, forced-evolution experiments in which mutated viruses with replication defects are used as parents for identification of spontaneous reversions that result in partial or full rescue of replication efficiency (Abbink and Berkhout, 2008; Berkhout and Das, 2009; Berkhout et al., 1997; Klaver and Berkhout, 1994), can reveal important insights into viral RNA structure.

While it is important to consider RNA structure in order to understand how various functions are carried out, most viral RNA functions require proteins. Therefore, deciphering the mechanisms of RNA action in viral biology requires a structural rationale for how viral RNA and proteins interact to perform a specific function. While many of the methods discussed below can be applied to RNA:protein systems, additional approaches can be used to further characterize the RNPs. Here restraints based on probing of specific amino acids in the protein followed by identification using mass spectrometry can be helpful. This technique has been used in a viral context previously to identify surface lysines on HIV-1 reverse transcriptase that are protected from covalent modification upon binding to the tRNA-primed viral genome (Kvaratskhelia et al., 2002). In addition many other reagents have been identified that can be used to specifically probe carboxylic acids, cysteine, histidine, tryptophan and tyrosine residues (Mendoza and Vachet, 2009). Due to the stringency of some of these reagents and the conditions in which they are active, care should be taken to ensure that neither the modifying reagents nor the conditions appreciably affect the properties of the RNP (Kvaratskhelia et al., 2002; Shell et al., 2005; Shkriabai et al., 2006).

### 3.2. X-ray crystallography

In honor of the centenary anniversary of the derivation of Bragg’s Law (Bragg, 1913), detailed reviews discussing current and proposed technological advances in XRC are abundant (Garman, 2014; Howard and Probert, 2014; Miller, 2014; Zhang and Ferre-D’Amare, 2014). For large RNA-only structures, XRC has been a workhorse, accounting for 138 of the 142 entries in the protein data bank (PDB) larger than 100 nt (Fig. 3). In fact, XRC is the only atomic-resolution method that has been used independently to solve an RNA-only structure larger than 101 nt. However, due to complications in crystallizing RNA such as high intrinsic flexibility and poor packing tendencies arising from the repulsive nature of polyanionic molecules (Holbrook and Kim, 1997), many RNAs are difficult to crystallize and resolution tends to wane as a function of RNA size (Fig. 4A). Obtaining well-ordered crystals for large RNAs can be exceedingly difficult; therefore, construct design is a critical first step. Careful inspection of phylogenetic variation aids in the design of constructs that effectively sample sequence space and sometimes brute-force trial-and-error of numerous variants is required. Such analyses were important factors that led to determination of RNase P RNA (Kazantsev et al., 2005), the glmS ribozyme (Klein and Ferre-D’Amare, 2006), and a number of riboswitches (Garst et al., 2008; Johnson et al., 2012b; Liberman et al., 2013; Montange and Batey, 2006; Spitale et al., 2009). Other considerations for construct design include defining domain boundaries (Correll et al., 1997), incorporating stable tetraloops (Pley et al., 1994; Scott et al., 1995) and addition of crystallization modules designed to make intermolecular contacts (Ferre-D’Amare et al., 1998b). Finally, favorable crystal contacts can be introduced by engineering hairpins into RNA constructs that are specifically recognized by small crystallization chaperone proteins such as U1A (Ferre-D’Amare and Doudna, 2000; Ferre-D’Amare et al., 1998a) and monoclonal antibody fragments (Koldobskaya et al., 2011; Piccirilli and Koldobskaya, 2011).

Advances at synchrotron facilities are continually improving both the quality of data and the ability to solve structures of samples once thought to be intractable to XRC. An emphasis on automation, from sample preparation and data collection to image processing and model refinement, is creating a more user-friendly experience. In addition to hardware improvements, software development has led to programs (such as HKL-3000 (Minor et al., 2006), PHENIX (Adams et al., 2004, 2010, 2011; Afonine et al., 2012) and the Auto-Rickshaw web server (Panjikar et al., 2005, 2009)) with the ability to perform data reduction, phasing, generation of electron density maps and model building in a very short period of time and with minimal user intervention. On the hardware side, as improvements in technology using traditional storage ring sources continue, the advent of X-ray free-electron lasers (XFELs) promises to revolutionize the field by enabling structure determination from macromolecules unable to form large, highly-ordered crystals and monitoring dynamic molecular motions at femtosecond timescale resolution (Aquila et al., 2012; Neutze and Moffat, 2012). XFELs can produce femtosecond-duration pulses with ~10^{12} photons at high energy (Emma et al., 2010). The feasibility of these types of studies has been demonstrated for known structures (Boutet et al., 2012; Chapman et al., 2011; Demirci et al., 2013; Kern et al., 2012, 2013; Redecke et al., 2013) and for de novo structure determination (Barends et al., 2014). The advantages of using XFELs parallel the needs of RNA XRC. For example, time-resolved diffraction is advantageous for analysis of dynamic molecules. Also, XFELs are ideally suited to analysis of micro/nanocrystals of RNA constructs for which large crystals cannot be grown.

#### 3.3. Nuclear magnetic resonance spectroscopy

NMR-restrained modeling has produced 442 RNA-only atomic-resolution structures in the PDB ranging from 8 to 101 nt in length. However, despite this large range of sizes, the median structure is only 24 nt (Figs. 3 and 4B). Even for conformationally homogeneous RNAs, NMR characterization of large RNAs suffers many drawbacks (Tolbert et al., 2010). (1) Large RNAs consisting of multiple secondary-structure elements require long-range restraints on relative interhelical orientations, but interproton NOEs are only useful over very short distances (<5 Å). (2) Only four different residue types are present in RNA leading to poor chemical shift dispersion; indeed, the ribose moiety of all four residues is chemically identical. Therefore, as the size of the RNA increases, spectral overlap eventually renders unambiguous resonance assignments impossible. (3) Compared to protein, RNA is decidedly less proton rich. (4) RNAs tend to be less globular than proteins, resulting in extended structures with slower tumbling times. This further exacerbates
the signal loss due to proportionally shorter transverse relaxation times as molecular mass increases.

To alleviate many of these shortcomings, new experimental strategies have been developed. A detailed description of these techniques is outside the scope of this review, but have been reviewed (Fernandez and Wider, 2003; Gao, 2013; Lukavsky and Puglisi, 2005; Tzakos et al., 2006). Briefly, isotope-labeling strategies can be used to partially alleviate spectral overlap. This can be achieved either through the combination of labeled and unlabeled NTPs during in vitro transcription reactions, or through segmental labeling and splint-mediated ligation (Lu et al., 2010). For both of these methods, commercial sources are available for uniformly-labeled NTPs. Also, selectively deuterated and 13C-enriched NTPs can be chemically synthesized (Davis et al., 2005; Scott et al., 2000). In addition to mitigating spectral overlap, residual dipolar couplings (RDCs) can be measured on selectively-labeled RNAs, yielding relative interdomain angular restraints (Molova et al., 2000). Finally, transverse relaxation-optimized spectroscopy (TROSY) significantly reduces the line broadening resulting from short transverse relaxation times in slowly tumbling RNAs (Pervushin et al., 1997). Overall, although many large RNAs remain intractable for atomic-resolution NMR characterization, improvements are continuing to push past size barriers. For sequences up to ~100 nt in length, NMR remains a viable option for solution-based structure determination of many RNAs.

3.4. Cryo-electron microscopy

Due to inherent size limitations, single-particle cryo-electron microscopic techniques have been applied only sparingly to RNAs. Therefore the current advantages and limitations of this technique will only be briefly described in the context of how it may be used in future studies. For more detailed reviews, see (Baker et al., 2010; Zhou, 2011). For cryo-EM, the signal-to-noise ratio increases with particle size, making it ideal for structure determination of large macromolecular complexes (Allegretti et al., 2014; Amunts et al., 2014; Fernandez et al., 2014; Liao et al., 2013). Currently, the limit of detection for sub-nanomolar resolution structure determination is ~200 kDa (roughly equivalent to 550–600 nt), significantly larger than most of the viral RNAs described in this review (Fig. 3). However, technology is currently being pursued to improve phase contrast for smaller molecules to reduce the size barrier to 100 kDa and below (Wu et al., 2013).

Despite the size limitations, many RNAs of this size would be beneficial to study. Indeed, with recent advances in cryo-EM technology, such as software for automatic post picking (Nicholson and Gaeser, 2001; Norouzi et al., 2013), correction methods for beam-induced particle movement (Bai et al., 2013; Shigematsu and Sigworth, 2013), introduction of improved phase plates (Wu et al., 2013) and development of direct electron detectors (Grigorieff, 2013; McMullan et al., 2009a,b,c), near-atomic resolution can be achieved for macromolecules and complexes >200 kDa. For example, genome dimerization mediated by sequences and structures within the 5′UTR of many retroviruses are a key checkpoint regulating genomic RNA packaging; however, little is known of the structural mechanisms of dimerization. Dimers of retroviral 5′UTR constructs could, in principle, be designed to be large enough for cryo-EM analysis. Additionally, protein components known to regulate dimerization, such as Gag, could also be characterized in an RNA-bound complex.

Although cryo-EM has the capacity to achieve sub-nanomolar resolution for large macromolecules, cryo-EM and cryo-electron tomography (cryo-ET) techniques can also be used in combination with NMR spectroscopy to characterize the structures of smaller (<200 kDa) molecules. Briefly, as described in Section 3.3, a key strength of NMR spectroscopy is the ability to identify short-range distances within macromolecules; however, detection of long-range interhelical restraints such as RDCs is technically difficult and requires conformational homogeneity. Cryo-EM/ET techniques can overcome the lack of long-range distance restraints inherent to NMR. The validity of this hybrid NMR/cryo-EM/ET approach has been demonstrated for the structure of the Moloney murine leukemia virus packaging element (Miyazaki et al., 2010). For this system, alternative techniques such as measuring RDCs or SAXS were unsuccessful due to sample heterogeneity. Single-particle cryo-EM/ET methods have the powerful advantage of removing conformational subpopulations from analysis.

3.5. Small-angle X-ray scattering

As described in a recent review, the popularity of SAXS applied to biomolecules has been steadily increasing over the past decade (Graewert and Svergun, 2013). The upsurge in SAXS use can be partially attributed to key advantages of SAXS over other techniques. For instance, SAXS is a true solution technique that can be performed under biologically relevant conditions. Additionally,
SAXS does not require generation of ordered crystals or incorporation of labels and does not have intrinsic size limitations (Fig. 3). In the current state of the field, however, structural information for large RNAs is limited to low resolution (~10–50 Å), yielding only general architectural features such as orientations of helices or large conformational changes.

Also contributing to the increased popularity of SAXS (and further bolstering its usage) has been the development of improved hardware and software systems that facilitate automated, high-throughput data collection (Franke et al., 2012; Hura et al., 2009; Martel et al., 2012) and comprehensive data analysis (Blanchet et al., 2012; Forster et al., 2010; Grant et al., 2011; Konarev et al., 2006). As for XRC, the push for speed and quality in SAXS data collection has been further aided by the improved performance of single-photon pixel detectors such as the PILATUS line (Kraft et al., 2009). The recently developed successor to the PILATUS detectors, EIGER, promises to reduce pixel size and readout dead time while enhancing frame rates (Johnson et al., 2012a). Similarly, software improvements have begun to tackle key issues. A powerful application of SAXS data is in the evaluation of structural models or the validation of crystal structures in a more biologically relevant context (in solution without crystal packing influences); however, complications in evaluating the contribution of solvent has reduced the validity of theoretical scattering profiles, especially at wider angles (WAXS). This becomes important to improving the resolution limits of SAXS/WAXS experiments because incorpora-

tion of high quality wide angle data has the potential to yield sub-nanomolar resolution structural information (Kofinger and Hummer, 2013). Attempts to alleviate this issue have come in the form of three new algorithms that use all-atom explicit solvent MD simulations (Virtanen et al., 2010, 2011), Zernike expansion (Liu et al., 2012) and Poisson–Boltzmann–Langvin formalism (Poitven et al., 2011) to improve the quality of theoretical scattering curves by incorporating solvent effects more rigorously.

The low-resolution nature of current SAXS technology and the requirement for highly homogenous samples limits its potential as a stand-alone technique. However, in combination with complementary methods, SAXS data can help to refine structural models for RNAs that are poorly suited to XRC or NMR. In a divide-and-conquer strategy, where NMR or XRC is used to determine atomic-resolution structures of short segments composing a larger RNA, SAXS can provide important long-range and global structure information for rigid–body docking to determine how the segments fit together in the context of the full macromolecule. While this has been used scarcely for large RNA complexes (Zuo et al., 2008), the feasibility of this strategy has been demonstrated for large, multidomain proteins such as CaMKII (Rosenberg et al., 2005), the hepatocyte growth factor/scattering factor (Gherardi et al., 2006), MLV Gag (Datta et al., 2011), HIV-1 gag (Datta et al., 2007) and the multi-aminoacyl-tRNA synthetase complex (Dias et al., 2013). While rigid–body docking is applicable in certain systems, SAXS-derived de novo structure characterization requires more complex modeling techniques. Molecular modeling techniques using only SAXS data currently require the ability to produce starting structures with matching overall architectural features. Over the past decade, significant progress has been made in the field of RNA 3D structure prediction leading to freely-available software and web servers (Das and Baker, 2007; Flores et al., 2011; Jonakas et al., 2009; Parisien and Major, 2008; Popenda et al., 2012; Sharma et al., 2008). De novo models can then be evaluated by calculating theoretical scattering profiles to find the best starting structure for modeling. Depending on the amount of desired conforma-
tional space sampling, different methods of MD simulation (replica exchange, accelerated, course-grain, simulated annealing, Monte Carlo, etc.) can be used to refine the starting model by calculating theoretical SAXS data from each sampled conformation (Jones et al., 2014). Also, simulated annealing protocols can be supplemented with SAXS-derived energy restraints (Fang et al., 2013; Jones et al., 2014). Indeed, other restraints, such as secondary structure and RDCs from NMR experiments can also be used (Bourke et al., 2012; Fang et al., 2013; Grishaev et al., 2008; Henning and Sattler, 2014; Jones et al., 2014; Lipfert et al., 2008; Wang et al., 2005a, 2010; Zuo et al., 2010). As SAXS/WAXS technology progresses, software improvements should be directed toward more seamless incorporation of experimental restraints such as FRET distances, chemical/ enzymatic probing results and NOE and RDC data from NMR experiments, thus providing a more high-throughput and user-friendly method.

4. Concluding remarks

Our understanding of viral replication has benefited from recent methodological and technological advances aimed at visualizing RNA structure. Many viral RNAs of functional importance such as IRESes, CITEs, TLSs and nuclear export, frameshifting and packaging signals have been studied; however, the field of viral RNA structural biology is still at a stage of development and our most successful high-resolution structures to date have been limited to small RNAs. Despite the limited number of RNAs studied, high-resolution structures and structural models have improved our understanding of viral replication and molecular biology. Continuing progress will hinge on expanding our abilities to solve large, functional RNA domains. In this review, we have demonstrated that such studies are possible and highlighted the techniques that are at the forefront of current capabilities. For example, studies that combine different techniques such as divide-and-conquer approaches and conflation of SAXS and NMR methods will be successful for many RNAs. Technological advances in SAXS and cryo-EM hold promise for improving resolution of large RNA structures. Careful application of new technologies, with an eye toward their limitations will provide a structural rationale for the many open-ended questions that continue to pervade viral molecular biology.

Acknowledgments

This work was supported by National Institutes of Health (NIH) Grant R01 AI077387 (to K.M.-F. and Dr. L. Kleiman). E.D.O. is supported by The Ohio State University Chemistry-Biology Interface Program training grant and W.A.C. by the Pelotonia Postdoctoral Fellowship Program.

References

Abbkinn, T.E., Berkhourth, B., 2003. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. J. Biol. Chem. 278 (13), 11601–11611.

Abbkinn, T.E., Berkhourth, B., 2008. RNA structure modules splicing efficiency at the human immunodeficiency virus type 1 major splice donor. J. Virol. 82 (6), 3090–3098.

Abbkinn, T.E., Ooms, M., Haasnoot, P.C., Berkhourth, B., 2005. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. Biochemistry 44 (25), 9058–9066.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D: Biol. Crystallogr. 66 (Pt 2), 213–221.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2011. The Phenix software for automated determination of macromolecular structures. Methods 55 (1), 94–106.

Adams, P.D., Czakó, K., Grosse-Kunstleve, R.W., Hung, L.W., Iorger, T.R., McCoy, A.J., Moriarty, N.W., Pai, R.K., Read, R.J., Romo, T.D., Sacchettini, J.C., Sauter, N.K., Storoni, L.C., Terwilliger, T.C., 2004. Recent developments in the Phenix software for automated crystallographic structure determination. J. Synchrotron Radiat. 11 (Pt 1), 53–55.
Dias, J., Renault, L., Perez, J., Miranda, M., 2013. Small-angle X-ray solution scattering analyses of a multi-aminocycl-RRNA synthesis complex reveals an elongated and multi-armed particle. J. Biol. Chem. 288 (33), 32970–32989.

Dibrov, S.M., Ding, K., Brunn, N.D., Parker, M.A., Bergdahl, B.M., Wyles, D.L., Hermann, T., 2012. Structure of a hepatitis C virus RNA domain in complex with a trans-activating RNA element binding a motif reminiscent of riboswitches. Proc. Natl. Acad. Sci. U.S.A. 109 (14), 5223–5228.

Dillon, P.J., Nelbock, P., Perkins, A., Rosen, C.A., 1990. Function of the human immunodeficiency virus type 1 and 2 Rev proteins is dependent on their ability to interact with a structured region present in env gene mRNA. J. Virol. 64 (9), 4428–4437.

DiMattia, M.A., Watts, N.R., Stahl, S.J., Rader, C., Wingfield, P.T., Stuart, D.J., Steven, A.C., Grimes, J.M., 2004. Implications of the HIV-1 Rev dimer structure at 3.2A resolution for non-ionic binding to the Rev response element. Proc. Natl. Acad. Sci. U.S.A. 101 (13), 5810–5814.

Ding, S.W., Voinnet, O., 2007. Antiviral immunity directed by small RNAs. Cell 130 (3), 413–426.

Dinnan, J.D., 2012. Mechanisms and implications of programmed translational frame-shifting. Wiley Interdiscip. Rev. RNA 3 (5), 661–673.

Drewe, T.W., 2009. Role of RNA-like structures in controlling plant virus replication. Virus Res. 139 (2), 217–229.

Du, Z., Giedroc, D.P., Hoffman, D.W., 1996. Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots. Biochemistry 35 (11), 4187–4199.

Du, Z., Holland, J.A., Hansen, M.R., Giedroc, D.P., Hoffman, D.W., 1997. Base pairings within the RNA pseudoknot associated with the simian retrovirus-1-gag-pro sequence. J. Mol. Biol. 270 (3), 464–470.

Dumas, P., Moras, D., Florentz, C., Giege, R., Verlaan, P., Van Kentum, A., Pleij, C.W., 1987. 3-D graphics modelling of the RNA-like 3′-end of turnip yellow mosaic virus RNA: structural and functional implications. J. Biomol. Struct. Dyn. 4 (5), 925–941.

Egli, T., Tsou, D., Rich, A., 2002. Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4302–4307.

Emma, P., Akre, J.A., Bionta, R., Bostedt, C., Bozek, J., Brachmann, A., Bucksbaum, P.F., Colby, C., Decker, F.J., Ding, Y., Dowell, D., Edstrom, S., Fisher, A., Frisch, J., Gieiche, L., Jastings, H., Hays, G., Hering, P., Huang, Z., Iversson, R., Loos, H., Messerschmidt, M., Nahmiah, A., Moeller, S., Nuhn, H.D., Pile, G., Ratter, D., Rezapiela, J., Schultz, D., Smith, T., Stefani, P., Tompkins, H., Turner, J., Vickers, J., White, W., Wu, J., Yockey, C., Galadzio, J., 2010. First lasing and operation of an Angstrom-wavelength free-electron laser. Nat. Photonics 4, 641–647.

Fabris, D., Yu, E.T., 2010. Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS/MS. J. Mass Spectrom. 45 (8), 841–860.

Fang, X., Wang, J., O’Carroll, I.P., Mitchell, M., Zuo, X., Wang, Y., Yu, P., Liu, Y., Rausch, J.W., Dyba, M.A., Kjems, J., Schiotters, C.D., Seiler, S., Winans, R.E., Watts, N.R., Stahl, S.J., Wingfield, P.T., Byrd, R.A., Le Grice, S.F., Reim, A., Wang, Y.X., 2013. An unusual topological structure of the HIV-1 Rev response element. Cell 155 (3), 594–605.

Feicht, P., Rudinger-Thirion, J., Florentz, C., Giege, R., 2001. Novel features in the RNA-like world of plant viral RNAs. Cell. Mol. Life Sci. 58 (11), 1547–1561.

Fedor, M.J., Westhof, E., 2002. Ribosomes: the first 20 years. Mol. Cell 10 (4), 703–704.

Felden, B., Florentz, C., Giege, R., Westhof, E., 1994. Solution structure of the 3′-end of tobacco mosaic RNA genomic RNAs. Conformational mimicry with canonical tRNAs. J. Mol. Biol. 235 (2), 508–531.

Felden, B., Florentz, C., Giege, R., Westhof, E., 1996. A central pseudoknotted three-way junction imposes RNA-like mimicry and the orientation of three 5′ upstream pseudoknots in the 3′ terminus of tobacco mosaic virus RNA. RNA 2 (3), 201–212.

Fernandez-Miragall, O., Ramos, R., Ramajo, J., Martinez-Salas, E., 2006. Evidence of reciprocal tertiary interactions between conserved motifs involved in gaining RNA structure essential for internal initiation of translation. RNA 12 (2), 223–234.

Fernandez, C., Wider, G., 2003. TROSY in NMR studies of the structure and function of eukaryotic RNA and RNA-like macromolecules. Curr. Opin. Struct. Biol. 13 (5), 570–580.

Fernandez, I.S., Bai, X.C., Murshudov, G., Sherer, S.H., Ramakrishnan, V., 2014. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. Cell 157 (4), 823–831.

Ferre-D’Amare, A.R., Doudna, J.A., 1996. Use of cis- and trans-ribosymes to remove 5′ and 3′ heterogeneities from milligrams of in vitro transcribed RNA. Nucl. Acids Res. 24 (5), 997–1003.

Ferre-D’Amare, A.R., Doudna, J.A., 2000. Crystalization and structure determination of a hepatitis delta virus ribosome: use of the RNA-binding protein U1A as a crystallization module. J. Mol. Biol. 295 (3), 541–556.

Ferre-D’Amare, A.R., Doudna, J.A., 1999a. Crystal structure of a hepatitis delta virus ribosome. Nature 395 (6702), 567–574.

Ferre-D’Amare, A.R., Zhou, K., Doudna, J.A., 1999b. A general module for RNA crystallization. J. Mol. Biol. 279 (3), 621–631.

Finley, M.F., Kieft, J.S., 2009. Toward a structural understanding of IRES RNA function. Curr. Opin. Struct. Biol. 19 (3), 267–276.

Firth, A.E., Brierey, L., 2012. Non-canonical translation in RNA viruses. J. Gen. Virol. 93 (P1), 1385–1409.
Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K., Lukasik, P., Bartol, N., Blachowska, A., 2012. Automated 3D structure construction for large RNAs. Nucl. Acids Res. 40 (14), e112.

Redecke, L., Noss, K., DePonte, D.P., White, T.A., Reinders, D., Barty, A., Stellato, F., Liang, M., Barends, T.R., Boutet, S., Williams, G.J., Messerschmidt, M., Seibert, M., Blank, T.R., 2011. Natively induced 3D structure determination of a twin-arginine translocase pore protein. PLoS Biol. 9 (11), e1001221.

Reeve, K., Lian, W., Royall, S., Mares, C.M., 2007. The RNA recognition motif: a family of protein-nucleic acid interfaces. RNA 13, 1182–1190.

Remus, T., Lamikanra, O., van den Heuvel, B., 2012. Auto-picking of cryo-EM images of homogeneous recombinant RNA — towards large-scale production of RNA. Nucl. Acids Res. 40 (13), e102.

Reuze, J., Mathews, D.H., 2010. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinform. 11, 129.

Reyes, F.E., Garst, A.D., Batey, R.T., 2005. Strategies in RNA crystallography. Methods Enzymol. 408, 39–53.

Rice, G.M., Leonard, C.W., Weeks, K.M., 2014. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. RNA, http://dx.doi.org/10.1261/rna.043323.113.

Rosenberg, O.S., DeInds, S., Sung, R.J., Nairn, A.C., Kuriyan, J., 2005. Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. Cell 123 (5), 849–860.

Roulston, A., Alcalde, R.C., Branton, P.E., 1999. Viruses and apoptosis. Annu. Rev. Microbiol. 53, 571–578.

Salvai-Lacoste, A., Di Tomaso, G., Piette, B.L., Legault, P., 2013. Affinity purification of T7 RNA transcripts with homogeneous ends using Aribio and CRISP tags. RNA 19 (11), 1930–1941.

Sanjuan, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belsham, R., 2010. Viral mutation rates. J. Virol. 84 (19), 9733–9748.

Santoro, S.W., Joyce, G.F., 1997. A general purpose RNA-cleaving enzyme. Proc. Natl. Acad. Sci. USA 94 (9), 4262–4267.

Scott, W.G., Finch, J.T., Klug, A., 1995. The crystal structure of an all-RNA hammerhead ribozyme. Nucl. Acids Symp. Ser. 34, 214–216.

Scott, L.G., Tolbert, T.J., Williamson, J.R., 2000. Preparation of specifically 2′- and 3′-O-(2-thiodiacyclopropyl)ribonucleosides. Methods Enzymol. 317, 18–38.

Seetin, M.G., Mathews, D.H., 2012. RNA structure prediction: an overview of methods. Methods Mol. Biol. 905, 99–122.

Sel, E., Niu, M., Kleiman, L., 2013. Linking to sequences within the primer binding site loop promotes an HIV-1 RNA conformation favoring RNA dimerization and packaging. RNA 19 (10), 1384–1393.

Sharma, S., Ding, F., Dokholyan, N.V., 2008. ifoldRNA: three-dimensional RNA structure prediction with the MultiFold algorithm. Nucleic Acids Res. 36, 2241–2252.

Shell, S.M., Hess, S., Kvaratskhelia, M., Zou, Y., 2005. Mass spectrometric identification of lssines involved in the interaction of human replication protein a with a single-stranded RNA. Biochemistry 44 (3), 971–978.

Shen, L.X., Tinoco, I., 1995. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. J. Mol. Biol. 247 (5), 963–978.

Shigeoka, H., Sigworth, F.J., 2013. Noise models and cryo-EM drift correction with a direct-electron camera. Ultramicroscopy 131, 61–69.

Shkiriaib, N., Datta, S.A., Zhao, Z., Hess, S., Rein, A., Kvaratskhelia, M., 2006. Interactions of HIV-1 Gag with assembly cofactors. Biochemistry 45 (15), 4077–4083.

Sidique, A., Frasch, C.E., Kiefer, T., 2009. Riboswitches: a foxRNA: flexiblity in the recognition of complex nanocolloids: RNA riboswitches for the recognition and folding. Bioinformatics 24 (17), 1951–1952.

Shell, S.M., Hess, S., Kvaratskhelia, M., Zou, Y., 2005. Mass spectrometric identification of lssines involved in the interaction of human replication protein a with a single-stranded RNA. Biochemistry 44 (3), 971–978.

Shen, L.X., Tinoco, I., 1995. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. J. Mol. Biol. 247 (5), 963–978.

Shigeoka, H., Sigworth, F.J., 2013. Noise models and cryo-EM drift correction with a direct-electron camera. Ultramicroscopy 131, 61–69.

Shkiriaib, N., Datta, S.A., Zhao, Z., Hess, S., Rein, A., Kvaratskhelia, M., 2006. Interactions of HIV-1 Gag with assembly cofactors. Biochemistry 45 (15), 4077–4083.

Sidique, A., Frasch, C.E., Kiefer, T., 2009. Riboswitches: a foxRNA: flexiblity in the recognition of complex nanocolloids: RNA riboswitches for the recognition and folding. Bioinformatics 24 (17), 1951–1952.

Shell, S.M., Hess, S., Kvaratskhelia, M., Zou, Y., 2005. Mass spectrometric identification of lssines involved in the interaction of human replication protein a with a single-stranded RNA. Biochemistry 44 (3), 971–978.

Shen, L.X., Tinoco, I., 1995. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. J. Mol. Biol. 247 (5), 963–978.

Shigeoka, H., Sigworth, F.J., 2013. Noise models and cryo-EM drift correction with a direct-electron camera. Ultramicroscopy 131, 61–69.

Shkiriaib, N., Datta, S.A., Zhao, Z., Hess, S., Rein, A., Kvaratskhelia, M., 2006. Interactions of HIV-1 Gag with assembly cofactors. Biochemistry 45 (15), 4077–4083.

Sidique, A., Frasch, C.E., Kiefer, T., 2009. Riboswitches: a foxRNA: flexiblity in the recognition of complex nanocolloids: RNA riboswitches for the recognition and folding. Bioinformatics 24 (17), 1951–1952.

Shell, S.M., Hess, S., Kvaratskhelia, M., Zou, Y., 2005. Mass spectrometric identification of lssines involved in the interaction of human replication protein a with a single-stranded RNA. Biochemistry 44 (3), 971–978.

Shen, L.X., Tinoco, I., 1995. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. J. Mol. Biol. 247 (5), 963–978.

Shigeoka, H., Sigworth, F.J., 2013. Noise models and cryo-EM drift correction with a direct-electron camera. Ultramicroscopy 131, 61–69.
Virtanen, G., Gallori, E., Maurel, M.C., 2009. Natural and unnatural ribozymes: back to the pYondRNA. RNA World Res. Microbiol. 52 (3), 457–465.

Tolbert, B.S., Miyazaki, Y., Barton, S., Kinde, B., Starck, P., Singh, R., Bax, A., Case, D.A., Summers, M.F., 2010. Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. J. Biomol. NMR 47 (3), 205–219.

Toor, N., Keating, K.S., Taylor, S.D., Pyle, A.M., 2008. Crystal structure of a self-spliced group II intron. Science 320 (5872), 77–82.

Tullius, T.D., Greenbaum, J.A., 2005. Mapping nucleic acid structure by hydroxyl radical cleavage. Curr. Opin. Biotechnol. 9 (2), 127–134.

Tzakos, A.G., Grace, C.R., Lukavsky, P.J., Riek, R., 2006. NMR techniques for very large proteins and RNAs in solution. Annu. Rev. Biophys. Biomol. Struct. 35, 319–342.

Uhlenbeck, O.C., 1995. Keeping RNA happy. RNA 1 (1), 4–6.

Vallejos, M., Deforges, J., Plank, T.D., Letelier, A., Ramdohr, P., Abraham, C.G., Valiente-Echeverria, F., Kieft, J.S., Sargueil, B., Lopez-Lastra, M., 2011. Activity of the human immunodeficiency virus type 1 cell cycle-dependent internal ribosomal entry site is modulated by IRES trans-acting factors. Nucl. Acids Res. 39 (14), 6186–6200.

Virtanen, J.J., Makowski, L., Sosnick, T.R., Freed, K.F., 2010. Modeling the hydration layer around proteins: HyPred. Biophys. J. 99 (5), 1611–1619.

Virtanen, J.J., Makowski, L., Sosnick, T.R., Freed, K.F., 2011. Modeling the hydration layer around proteins: applications to small- and wide-angle X-ray scattering. Biophys. J. 101 (8), 2061–2069.

Walker, S.C., Avir, J.M., Conn, G.L., 2003. General plasmids for producing RNA in vitro transcripts with homogeneous ends. Nucl. Acids Res. 31 (15), e82.

Wang, J., Zuo, X., Yu, P., Xu, H., Starich, M.R., Tiede, D.M., Shapiro, B.A., Schwieters, C.D., Wang, Y.X., 2009a. A method for helical RNA global structure determination in solution using small-angle X-ray scattering and NMR measurements. J. Mol. Biol. 393 (3), 717–734.

Wang, Y.X., Zuo, X., Wang, J., Yu, P., Butcher, S.E., 2010. Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. Methods 52 (2), 180–191.

Wang, Z., Parisien, M., Scheets, K., Miller, W.A., 2011. The cap-binding translation initiation factor, elf4E, binds a pseudoknot in a viral cap-independent translation element. Structure 19 (6), 868–880.

Wang, Z., Treder, K., Miller, W.A., 2009b. Structure of a viral cap-independent translation element that functions via high affinity binding to the elf4E subunit of elf4F. J. Biol. Chem. 284 (21), 14189–14202.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess Jr., J.W., Svanstrom, R., Burch, C.L., Weeks, K.M., 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460 (7256), 711–716.

Weeks, K.M., 2010. Advances in RNA structure analysis by chemical probing. Curr. Opin. Struct. Biol. 20 (3), 295–304.

Wilkinson, K.A., Merino, E.J., Weeks, K.M., 2006. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. 1 (3), 1610–1616.

Wilson, J.E., Pestova, T.V., Hellen, C.U., Sarnow, P., 2000. Initiation of protein synthesis from the A site of the ribosome. Cell 102 (4), 511–520.

Wu, Q., Huang, L., Zhang, Y., 2009. The structure and function of catalytic RNAs. Sci. China C Life Sci. 52 (3), 394–405.

Wu, Y., Wang, C., Chang, J., Chen, Y., Miyazaki, N., Murata, K., Nagayama, K., Chang, W., 2013. Zernike phase contrast cryo-electron microscopy reveals 100 kDa component in a protein complex. J. Phys. D: Appl. Phys. 46, 494008–494017.

X., G., Gorin, A., Ellington, A.D., Patel, D.J., 1996. Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 REV peptide-RNA aptamer complex. Nat. Struct. Biol. 3 (12), 1026–1033.

Yilmaz, A., Bolinger, A., Boris-Lawrie, K., 2006. Retrovirus translation initiation: issues and hypotheses derived from study of HIV-1. Curr. HIV Res. 4 (2), 131–139.

Yu, E.T., Hawkins, A., Eaton, J., Fabris, D., 2008a. MS3D structural elucidation of the HIV-1 packaging signal. Proc. Natl. Acad. Sci. U.S.A. 105 (34), 12248–12253.

Yu, E.T., Hawkins, A., Kuntz, I.D., Rahn, L.A., Rothfuss, A., Sale, K., Young, M.M., Yang, C.L., Pancereilla, C.M., Fabris, D., 2008b. The collaboratory for MS3D: a new cybersubstrate for the structural elucidation of biological macromolecules and their assemblies using mass spectrometry-based approaches. J. Proteome Res. 7 (11), 4848–4857.

Yu, E.T., Zhang, Q., Fabris, D., 2005. Untying the IVV frameshifting pseudoknot structure by MS3D. J. Mol. Biol. 345 (1), 69–80.

Zhang, J., Ferre-D’Amare, A.R., 2014. New molecular engineering approaches for crystallographic studies of large RNAs. Curr. Opin. Struct. Biol. 26, 9–15.

Zhou, J., Bean, R.L., Vogt, V.M., Summers, M., 2007. Solution structure of the Rous sarcoma virus nucleocapsid protein: mufSI RNA packaging signal complex. J. Mol. Biol. 365 (2), 453–467.

Zhou, S., Rynearson, K.D., Ding, K., Brunn, N.D., Hermann, T., 2013. Screening for inhibitors of the hepatitis C virus internal ribosome entry site RNA. Bioorg. Med. Chem. 21 (20), 6139–6144.

Zhou, Z.H., 2011. Atomic resolution cryo electron microscopy of macromolecular complexes. Adv. Protein Chem. Struct. Biol. 82, 1–35.

Zhu, B., Tabor, S., Richardson, C.C., 2014. Syn5 RNA polymerase synthesizes precise run-off RNA products. Nucl. Acids Res. 42 (5), e33.

Zhu, J., Korostelev, A., Costantino, D.A., Donohue, J.P., Noller, H.F., Kieft, J.S., 2011. Crystal structures of complexes containing domains from two viral internal ribosome entry site (IRES) RNAs bound to the 70S ribosome. Proc. Natl. Acad. Sci. U.S.A. 108 (5), 1839–1844.

Ziehler, W.A., Engelke, D.R., 1996. Synthesis of small RNA transcripts with discrete 5′ and 3′ ends: Biotechniques 20 (4), 622–624.

Zuker, M., 2003. Mfold web server for nucleic acid folding and prediction. Nucl. Acids Res. 31 (13), 3406–3415.

Zuo, X., Wang, J., Foster, T.R., Schwieters, C.D., Tiede, D.M., Butcher, S.E., Wang, Y.X., 2008. Global molecular structure and interfaces: refining an RNA/RNA complex structure using solution X-ray scattering data. J. Am. Chem. Soc. 130 (11), 3292–3293.

Zuo, X., Wang, J., Yu, P., Eyler, D., Xu, H., Starich, M.R., Tiede, D.M., Simon, A.E., Kaspzik, W., Schwieters, C.D., Shapiro, B.A., Wang, Y.X., 2010. Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3′ UTR of turnip crinkle virus. Proc. Natl. Acad. Sci. U.S.A. 107 (4), 1385–1390.