Coronavirus genomic RNA packaging

Paul S. Masters*

Wadsworth Center, New York State Department of Health, Albany, NY, 12201, United States

A R T I C L E I N F O

A B S T R A C T

RNA viruses carry out selective packaging of their genomes in a variety of ways, many involving a genomic packaging signal. The first coronavirus packaging signal was discovered nearly thirty years ago, but how it functions remains incompletely understood. This review addresses the current state of knowledge of coronavirus genome packaging, which has mainly been studied in two prototype species, mouse hepatitis virus and transmissible gastroenteritis virus. Despite the progress that has been made in the mapping and characterization of some packaging signals, there is conflicting evidence as to whether the viral nucleocapsid protein or the membrane protein plays the primary role in packaging signal recognition. The different models for the mechanism of genomic RNA packaging that have been prompted by these competing views are described. Also discussed is the recent exciting discovery that selective coronavirus genome packaging is critical for *in vivo* evasion of the host innate immune response.

1. Introduction: the genome packaging problem for RNA viruses

The concluding stages of viral replication typically involve assembly and egress of progeny virions from the infected cell. At this point viruses must address a critical issue consequential to their parasitic status – genome packaging. The most fundamental formulation of the packaging problem is: how do viruses ensure the specific incorporation of their genomes into assembled virions, faced with competition from numerous other nucleic acid species in the infected cell? For RNA viruses (most of which are cytoplasmic), the solutions that have evolved to overcome this problem vary considerably, according to the needs imposed by the wide range of lifestyles and replication strategies that RNA viruses can follow. A broad array of positive-strand RNA viruses (picornaviruses, flaviviruses, alphaviruses, coronaviruses, and retroviruses) and negative-strand RNA viruses (rhabdoviruses, paramyxoviruses, and orthomyxoviruses) are generally held to selectively package genome-length RNA of the correct polarity and to exclude other viral and host species. This process occurs with a high degree of efficiency, although noteworthy exceptions can be found. For example, retroviruses package abundant amounts of host 7SL RNA (Onafuwa-Nuga et al., 2006), and arenaviruses incorporate entire ribosomes (Sarute and Ross, 2017).

Specific packaging is often mediated by well-defined *cis*-acting RNA sequences or structures within the viral genome, designated packaging signals (PSs). One of the most intensively studied PS elements, named psi, is contained in the 5′ leader region of the HIV genome. The mechanism by which psi functions remains to be fully resolved, but recent studies have revealed that genome dimerization promotes the folding of psi and neighboring RNA elements into a conformer in which the genomic splice donor site is sequestered (Keane et al., 2015, 2016). These and other structural features suggest how the viral Gag protein selects unspliced genomic RNA and how packaging is coupled to dimerization of the pseudodiploid retroviral genome. Another well characterized PS occurs within the nsP1 coding region of the genome of Venezuelan equine encephalitis virus and is highly conserved in many other alphaviruses. This element consists of a cluster of short stem-loop structures, each containing a GGG loop motif, which appear to contribute additively to the efficiency of genome packaging (Kim et al., 2011).

An additional level of complexity is encountered in the packaging of segmented viral genomes. Although bi- and tri-segmented viruses may depend upon stochastic processes in order to accumulate virions with a complete genome, a recent analysis by single-molecule fluorescent *in situ* hybridization has shown that most individual influenza virions have a full complement of eight genome segments (Chou et al., 2012). Segment-specific PSs map near the termini of each segment, but how the network of segment-segment interactions is established is not yet precisely defined (Bolte et al., 2019). Influenza viruses, as well as rhabdoviruses and paramyxoviruses, also provide examples of an important distinction that can exist for some enveloped viruses – the difference between encapsidation and packaging. For these viruses, both genomic and antigenic RNA are encapsidated into helical ribonucleoproteins,

but only genomic RNA is packaged into virions.

A dedicated genomic structural element is not the only means by which an RNA virus can determine packaging specificity. Picornaviruses and flaviviruses, most of which lack a PS, tightly couple packaging to RNA replication (Nugent et al., 1999; Barrows et al., 2018). In particular, for poliovirus no clearly defined PS has ever been uncovered, despite extensive searches (Jiang et al., 2014). Moreover, exhaustive recoding experiments appear to rule out the existence of cryptic PSs in the poliovirus genome (Song et al., 2017). This has led to the hypothesis that polio capsid protein-replicase interactions are sufficient to ensure the specificity of genome packaging. An alternative view that is developing for many other icosahedral positive-strand RNA viruses is that there exist multiple dispersed (and in many cases, poorly defined) genomic PSs. These are postulated to serve as recognition sites for capsid subunits, which cause collapse of genomic RNA into a more condensed configuration through protein-protein interactions. The resulting compacted intermediate structure then recruits additional capsid subunits to complete assembly of the virion shell (Borodavka et al., 2012; Mendes and Kuhn, 2018).

2. Coronaviruses

Coronaviruses (CoVs) are a family of enveloped RNA viruses that infect mammals and birds, causing mainly respiratory and gastrointestinal diseases (Masters and Perlman, 2013). They have long been studied, in part because of their significant impact on livestock and companion animals. In humans, four coronaviruses, HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1, generally cause upper respiratory tract infections and are prevalent worldwide. Of paramount concern to human health, however, are two deadly coronaviruses that emerged just in this century (de Wit et al., 2016). The first of these, the causative agent of severe acute respiratory syndrome (SARS), originated in bats. SARS-CoV caused an outbreak of atypical pneumonia that began in China in late 2002, spread worldwide, and was extinguished by mid-2003 through the rigorous application of public health measures. All told, the SARS epidemic was characterized by high human-to-human transmission and a case-fatality rate of 10%. Although SARS-CoV was eliminated from the human population, very closely related coronaviruses persist in bat reservoirs, posing a constant threat for re-emergence. The second most noteworthy human coronavirus is the one responsible for Middle East respiratory syndrome (MERS), an outbreak of severe pneumonia that began in Saudi Arabia in 2012. At present, MERS is confined to the Arabian peninsula, from where cases have sporadically dispersed to other parts of the world. Like SARS-CoV, MERS-CoV has an ancestral origin in bats, but its proximal reservoir is dromedary camels, and transmission from this source in ongoing. In contrast to SARS-CoV, MERS-CoV does not spread as efficiently through human-to-human contact; nevertheless, it has an alarming case-fatality rate of 36%. The emergence of these two previously unknown coronaviruses has stimulated efforts to fully understand all aspects of the molecular biology of members of this family.

The virions of coronaviruses contain four conserved structural proteins (Fig. 1): the spike (S) protein, which governs binding to host cell receptors and virus entry into cells; the membrane (M) protein and the envelope (E) protein, which mediate virion budding; and the nucleocapsid (N) protein, which together with genomic RNA (gRNA) constitutes the nucleocapsid. S, E, and M are embedded in a membrane envelope derived from the site of budding, the Golgi-endoplasmic reticulum intermediate compartment. The nucleocapsid resides in the interior of the virion envelope. The genomes of coronaviruses are nonsegmented, positive-sense RNA molecules of exceptional length (25–32 kb), with 5′ caps and 3′ polyadenylate tails (Fig. 1). The downstream end of the genome contains the genes for the structural proteins in the invariant order S-E-M-N, as well as interspersed accessory genes. The 5′ two-thirds of the genome contains two open reading frames (rep1a and rep1b) encoding the viral replicase-transcriptase

polyprotein, which is expressed by a ribosomal frameshifting mechanism. The replicase is cotranslationally processed into 15 or 16 subunits designated nonstructural proteins (nsps) that carry out RNA-synthesis functions common to most RNA viruses, as well as a collection of enzymatic activities unique to coronaviruses and other members of the nidovirus order (Ziebuhr, 2005).

There are three aspects of coronavirus molecular biology most critical to a consideration of genome packaging. First, viral infection entails not only replication of gRNA but also production of a set of subgenomic (sg) RNAs that serve as mRNAs for the downstream genes. The sgRNAs, synthesized via negative-strand intermediates, form a nested set, each member of which contains a 5′ genomic leader RNA fused to a segment of the 3′ end of the genome (Fig. 1). Second, coronaviruses have helically symmetric nucleocapsids (Masters, 2006; Neuman et al., 2006; Gui et al., 2017), a feature that is unusual for positive-strand RNA viruses, which generally have icosahedral capsids. Third, the assembly of the nucleocapsid into the virion occurs through interactions between the carboxy termini of multiple monomers of N protein and of M protein (Escors et al., 2001; Hurst et al., 2005; Verma et al., 2006).

For coronaviruses, the packaging problem amounts to achieving highly specific recognition and selection of gRNA from among an abundance of viral sgRNAs, cellular mRNAs, and other RNA species in infected cells. Although there have been some reports suggesting packaging of sgRNA (Hofmann et al., 1990; Sethna et al., 1989; Zhao et al., 1993), it has generally been observed that stringently purified virions almost exclusively contain positive-sense gRNA (Makino et al., 1990; Escors et al., 2003; Kuo and Masters, 2013). In the relatively few cases where it has been examined in detail, this selectivity has been found to be accomplished by means of a genomic PS.

3. Coronavirus packaging signals

The PSs of coronaviruses are not broadly conserved across the family. Coronaviruses are taxonomically classified into four genera – alpha-, beta-, gamma-, and deltacoronaviruses – and these are further sorted into lineages or subgenera. In particular, the betacoronaviruses comprise four phylogenetic lineages. The first of these, the lineage A betacoronaviruses, provided the foundation for coronavirus packaging studies and, so far, have remained their principal focus.

3.1. Betacoronavirus lineage A packaging signals

The discovery and initial characterization of a coronavirus PS was carried out with the prototype lineage A betacoronavirus, mouse hepatitis virus (MHV), through experiments with defective-interfering (DI) RNAs. DI RNAs are extensively deleted gRNAs that retain the necessary *cis*-acting RNA elements that allow them to propagate by parasitizing the replicative machinery of the parental virus from which they originate. Comparisons of naturally arising and engineered DI RNAs that were either able or unable to be packaged into purified virions allowed the localization of the MHV PS to a portion of the replicase gene (Makino et al., 1990; van der Most et al., 1991). A more detailed study finely mapped the PS to a 190-nt segment of rep1b (Fosmire et al., 1992). This locus is some 20 kb downstream from the 5′ end of the genome, distant from *cis*-acting RNA elements required for viral RNA replication. As would be expected, it only appears in gRNA and not in the nested set of sgRNAs. Further mutational analysis identified a proposed 69-nt substructure within the 190-nt segment as the minimal functional PS (Fosmire et al., 1992). However, subsequent studies found that larger forms of the PS were more efficient for packaging of DI RNAs or heterologous RNAs into virions of helper virus (Cologna and Hogue, 2000) or into virus-like particles (VLPs) formed from expressed structural proteins (Bos et al., 1996; Narayanan and Makino, 2001). Moreover, it was shown that bovine coronavirus (BCoV), another lineage A betacoronavirus, recognizes the PS of MHV, and reciprocally, MHV recognizes the PS of BCoV (Cologna and Hogue,
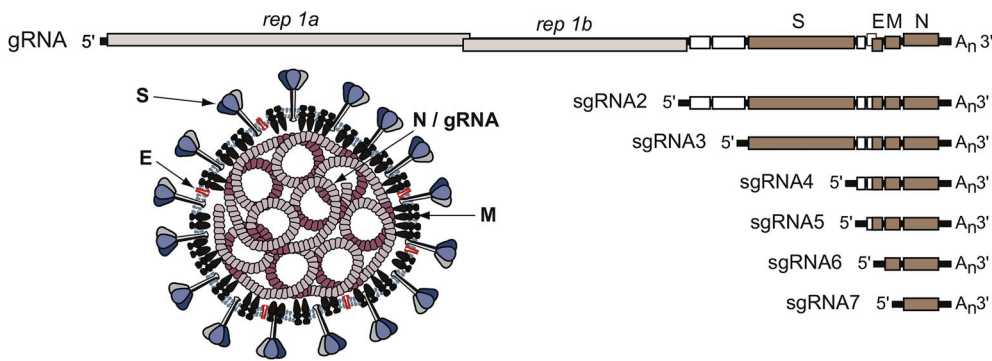
Fig. 1. Coronavirus RNA species and virions. *Top*, a schematic of the coronavirus genome (gRNA), typified by that of mouse hepatitis virus (MHV). The 5′ end contains the rep1a and rep1b genes, which encode the viral replicase-transcriptase, while the 3′ end contains the structural protein genes S, E, M, and N. Also at the 3′ end of the genome are accessory genes (unlabeled), the locations and numbers of which vary for different viruses; those shown are for MHV. Under the genome, to the *right*, are the nested set of subgenomic mRNAs (sgRNAs), of which MHV has six. To the *left* is a schematic of the virion, depicting the essential structural proteins: spike protein (S); membrane protein (M); envelope protein (E); and nucleocapsid protein (N).

2000).

A proposed model for the MHV PS was constructed through analysis of the folding of the homologous region from all lineage A betacoronavirus genomes (Chen et al., 2007). This structure, a 95-nt bulged stem-loop, contains four copies of an AGC/GUAAU motif, each displaying an AA (or GA bulge) on its 3′ side (Fig. 2A). These repeat units are spaced at regular intervals, with a two-fold quasi-symmetry centered around an internal loop. It was noted by Chen et al. (2007) that their model has a different RNA fold than the previously reported 69-nt substructure of Fosmire et al. (1992), and it is not in complete accord with that earlier study. The 95-nt bulged stem-loop structure was shown to be in good agreement with *in vitro* chemical and enzymatic probing experiments. It is also strongly supported by phylogenetic criteria (Fig. 2B). The terminal loop and the repeat units are absolutely conserved in every currently known lineage A betacoronavirus. All other stem segments are preserved by covariation of base pairs. In contrast, wider divergence is allowed in the central internal loop, where numerous nucleotide substitutions or small insertions occur. Notably, the PS of equine coronavirus (EqCoV), a close relative of BCoV, contains a deletion that precisely encompasses the central bulge and one of the repeat units (Fig. 2B).

The betacoronavirus lineage A PS falls in a region of rep1b that codes for the nsp15 subunit of the replicase-transcriptase (Deng and Baker, 2018). The PS RNA encodes a segment of nsp15 that is absent from the nsp15 sequences of betacoronavirus lineages B, C, and D, as well as those of alpha-, gamma-, and deltacoronaviruses (Fig. 2C). The encoded polypeptide corresponds to a surface loop of MHV nsp15 (Xu et al., 2006) that does not exist in the nsp15 structures of SARS-CoV (Joseph et al., 2007) or MERS-CoV (Zhang et al., 2018). Thus, there is no basis for the presumption that the same region of nsp15 contains the PS for either SARS-CoV (Hsieh et al., 2005) or MERS-CoV (Hsin et al., 2018). The ancillary positioning of the PS-encoded protein loop suggests that this PS was a late acquisition during the evolution of the coronavirus family. Alternatively, it remains possible that this PS was originally present in a common ancestor of all coronaviruses and was subsequently lost from most of them; however, the advantage provided *in vivo* by selective gRNA packaging (see below) makes such a scenario less likely.

Curiously, previously unnoticed PS elements appear elsewhere in the genomes of two lineage A betacoronaviruses. Three of the four isolates of rabbit coronavirus (RbCoV-HKU-14; Lau et al., 2012) contain a 116-nt insertion that, it turns out, harbors a duplication of the upper half of the native PS (Fig. 2D). Remarkably, the remainder of the insertion carries an additional AGC/GUAAU repeat unit. Similarly, a duplication of the terminal segment of the PS can be found in a North American isolate of EqCoV (Zhang et al., 2007), but not in two Japanese isolates (Nemoto et al., 2015). Both the RbCoV and the EqCoV partial PS duplications occur at exactly the same site in a highly variable

region of the replicase nsp3 subunit, situated between two domains designated NAB and G2M (Neuman et al., 2008). The functional significance of these duplications, if any, remains to be determined.

3.2. Packaging signals of other coronaviruses

As discussed above, other lineages and genera of coronaviruses clearly do not possess a counterpart of the lineage A betacoronavirus PS. Among these, the only species for which packaging has been examined in depth is the alphacoronavirus transmissible gastroenteritis virus (TGEV). TGEV PS studies began with the characterization of three naturally arising DI RNAs that were found to be efficiently packaged (Méndez et al., 1996). Notably, DI RNA-containing particles could be separated from those of virions by density gradient centrifugation, indicating that they were packaged independently of full-length gRNA. However, it is not clear if this is a general property of other packaged non-genomic RNAs, since even the smallest of the DI RNAs analyzed was quite large (9.7 kb).

Dissection of smaller constructed DI RNAs, accompanied by stringent virion purification, led to the mapping of the TGEV PS to a segment spanning nucleotides 100–649 from the 5′ end of the viral genome (Escors et al., 2003). This same segment was shown to contain the 5′ *cis*-acting elements necessary for RNA replication. A subsequent study, which assayed packaging by duplication of the packaging region within a sgRNA of an engineered virus, reduced the downstream boundary to nt 598 (Fig. 3A) (Morales et al., 2013). Further limitation of the RNA structures required for virion incorporation could not be attained through deletion analysis with this system, leading the authors to conclude that the TGEV PS encompasses the entire 5′-most 598 nt and possibly other parts of the genome. If correct, this would indicate that the TGEV PS is much more highly complex than that of MHV.

A different, but more speculative, perspective on the PSs of TGEV and other alpha- and betacoronaviruses came from an analysis of the predicted RNA structures of the genomic 5′ ends of all coronaviruses (Chen and Olsthoorn, 2010). Among the features that are conserved across all genera is an articulated bulged stem-loop, denoted SL5, which straddles the 5′ UTR-nsp1 boundary. In the alphacoronaviruses, as well as in all betacoronavirus lineages except lineage A, a set of substructures emerge from the apex of SL5. For TGEV, these SL5 insertions consist of three smaller stem-loops displaying repeats of the motif UUCCG(U/C) (Fig. 3B); three highly similar repeat units appear in all other alphacoronaviruses. Moreover, for the lineage B, C, and D betacoronaviruses, including SARS-CoV and MERS-CoV, SL5 contains two or three repeating substructures, each with the loop motif UUUCGU. The repetitive nature of these elements, akin to that of the lineage A betacoronavirus PS (above) and to those of the PSs of alphaviruses (Kim et al., 2011), led to the proposal that these are the corresponding PSs of the alpha- and betacoronaviruses in which they occur (Chen and
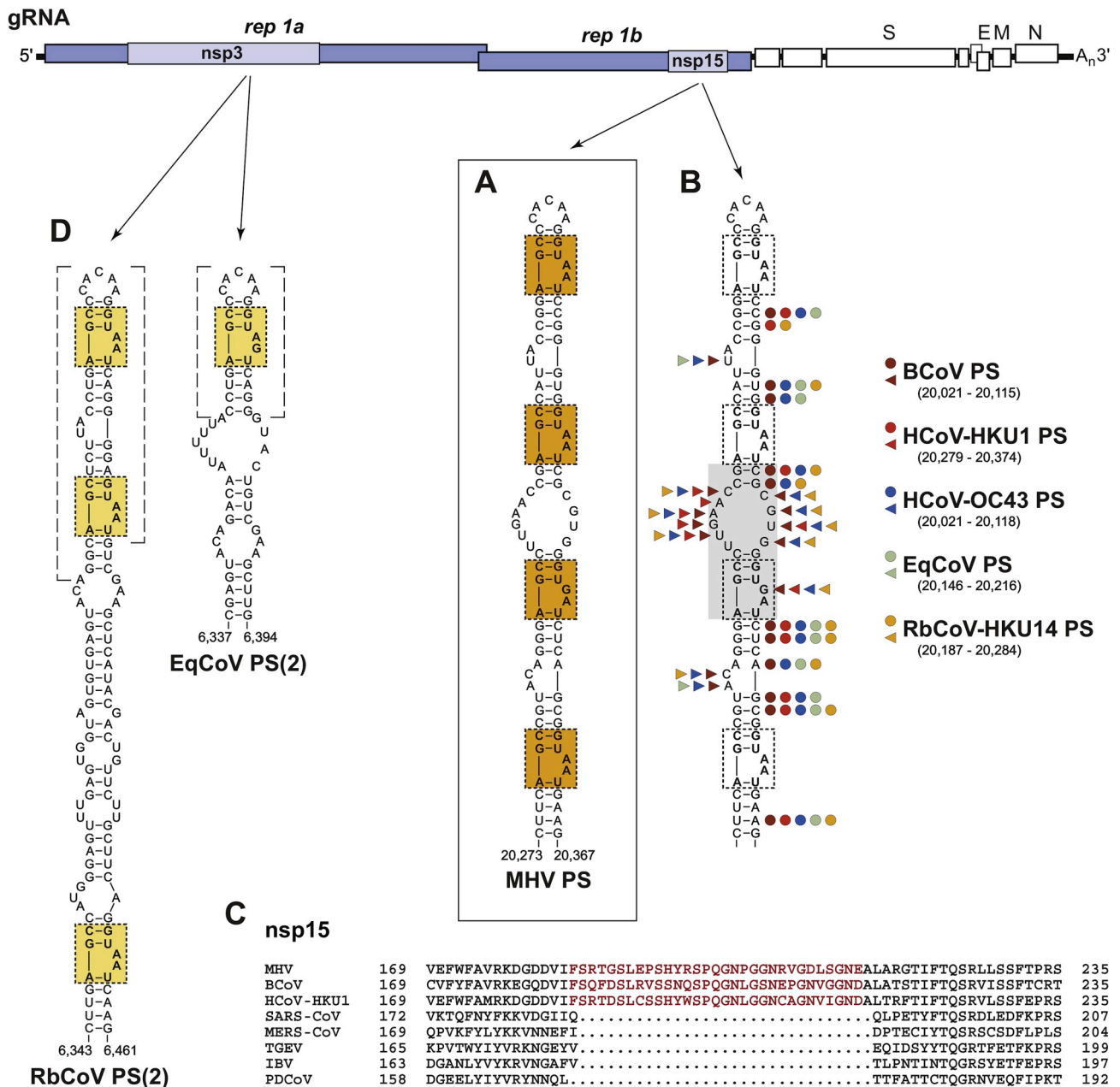
Fig. 2. Betacoronavirus lineage A packaging signals. *Top*, a map of the lineage A betacoronavirus genome, showing the loci of PS and PS-like elements. (A) Structure of the MHV PS, modeled by Chen et al. (2007). Repeat units are highlighted in orange. (B) Phylogenetic conservation of the PS among lineage A betacoronaviruses. Circles denote basepairs in double-stranded sections that co-vary between MHV and one of the other viruses; additional base-pairing covariation is found among different strains of MHV and of HCoV-HKU1 (not shown). Triangles denote bases in single-stranded sections that differ from MHV; the shaded rectangle indicates the part of the structure that is deleted in EqCoV. (C) Amino-acid alignment of part of the nsp15 subunit of the replicase-transcriptase. Highlighted in red is the segment encoded by the PS of lineage A betacoronaviruses. This segment is absent from all other lineages and genera, here represented by SARS-CoV (betacoronavirus, lineage B), MERS-CoV (betacoronavirus, lineage C), TGEV (alphacoronavirus), IBV (gammacoronavirus), and PDCoV (deltacoronavirus). (D) Upstream PS-like structures, denoted PS(2), in RbCoV and EqCoV. In each, broken-line brackets show the region that is a duplication of PS sequence in nsp15; repeat units are highlighted in yellow. GenBank accession numbers for the sequences shown: MHV, AY700211; BCoV, U00735; HCoV-HKU1, AY597011; HCoV-OC43, AY903460; EqCoV, EF446615; RbCoV-HKU14, JN874559; SARS-CoV, AY278741; MERS-CoV, JX869059; TGEV, AJ271965; IBV, AJ311317; PDCoV, JQ065042.

Olsthoorn, 2010). A straightforward test of this hypothesis by reverse genetics has yet to be carried out.

Gamma- and deltacoronaviruses, on the other hand, do not contain a version of the SL5 loop insertions. For the prototype gammacoronavirus, infectious bronchitis virus (IBV), an early study was able to show packaging of a constructed 5.9-kb DI RNA composed of a mosaic of genome segments, although further localization of a PS was not pursued (Dalton et al., 2001). For the deltacoronaviruses, genome packaging studies may be delayed by the fact that currently only one

species from this genus, porcine deltacoronavirus (PDCoV), can be propagated *in vitro* (Hu et al., 2015). However, the hallmarks of a coronavirus PS appear to be its insertion into a highly conserved region of the replicase gene and its display of repeated units of RNA sequence and structure. These criteria draw attention to a 104-nt RNA segment of the rep1b gene of some deltacoronaviruses that contains six small stem-loops, five of which present a repeated GUAC loop motif (Fig. 3C). This segment falls within a 40-amino-acid insertion at the nsp13-nsp14 junction that is present in roughly half of the known deltacoronavirus
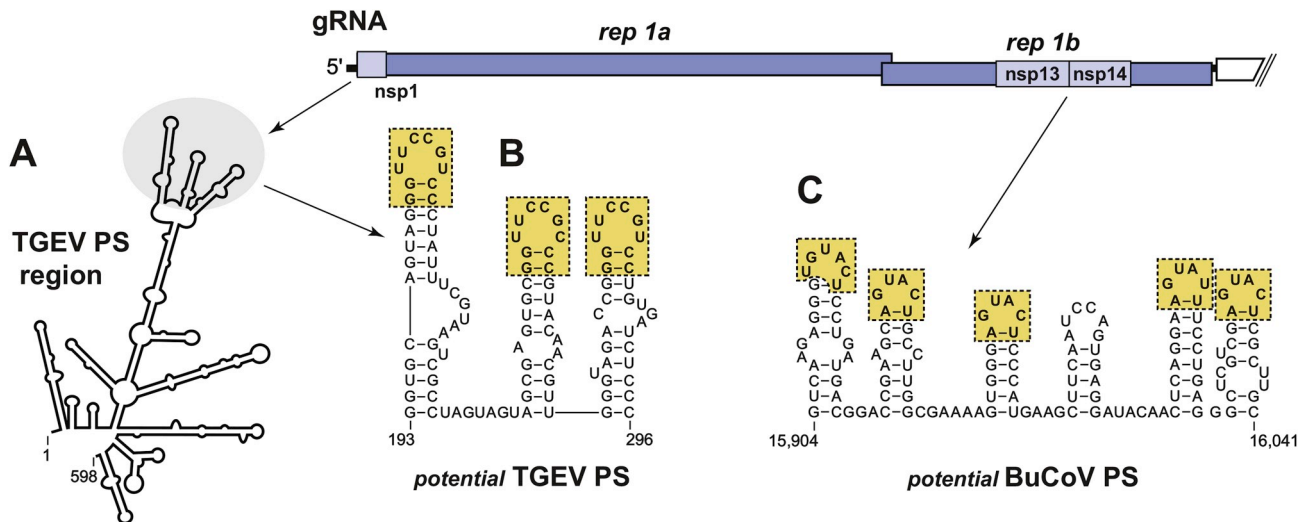
Fig. 3. Other coronavirus PSs. (A) The genomic segment to which the PS for the alphacoronavirus TGEV has been localized (Morales et al., 2013); nucleotides 1–598 include the 5′ UTR and almost all of the nsp1 gene. (B) The TGEV PS predicted by Chen and Olsthoorn (2010), corresponding to the shaded area in (A). (C) A potential PS, containing multiple repeats of a loop motif, found at the nsp13-nsp14 junction of a subset of deltacoronaviruses, including BuCoV. In (B) and (C), repeat units are highlighted in yellow. GenBank accession numbers for the sequences shown: TGEV, AJ271965; BuCoV-HKU11, FJ376620.

species, including bulbul coronavirus (BuCoV), but is absent from other members of the genus, including PDCoV.

4. Role of the packaging signal in the viral genome

The actual *in situ* role of the PS in the coronavirus genome, rather than in DI RNAs, was explored through the construction of MHV mutants in which this element was altered, deleted, or relocated (Kuo and Masters, 2013). A mutant, named silPS, was created through the introduction of 20 coding-silent mutations into the PS (Fig. 4A). This allowed the complete disruption of PS RNA primary sequence and secondary structure in a manner that did not modify the product of the nsp15 gene in which it resides. Analysis of extensively purified virions revealed that the silPS mutant, in contrast to the wild type, packaged large amounts of sgRNAs in addition to gRNA (Fig. 4A). The same loss of RNA packaging specificity was observed in another mutant in which the PS was entirely deleted. Moreover, it was shown that the PS was fully functional if deleted at its native site and transposed to another locus downstream of rep1b (but still unique to gRNA). It should be noted that in infected cells, all of these mutants synthesized the same complement of gRNA and sgRNAs, and in the same amounts, as the wild type.

Contrary to initial expectations, the silPS or PS-deletion mutations had little to no effect on viral plaque size or growth kinetics in tissue culture. Virions of silPS did not have a measurably different particle-to-PFU ratio compared to that of the wild type. Nevertheless, growth competition experiments demonstrated that viruses containing the PS had a fitness advantage over mutants lacking this element (Kuo and Masters, 2013). Similarly, in a separate study, a mutant was constructed in which most of the PS was replaced by a segment encoding an epitope tag and its complement (Athmer et al., 2017). This mutant also packaged greatly elevated amounts of sgRNA into virions, and it was demonstrated to be out-competed by wild-type MHV during serial passaging in tissue culture.

Thus, the different PS mutants showed that in coronavirus replication the function of the PS is to ensure the selective incorporation of gRNA into virions, as opposed to being strictly required for gRNA packaging. The minimally defective growth phenotype of PS mutants was at first surprising, because the prior studies with DI RNAs had fixed the notion that the PS is necessary for packaging of an RNA species into virions. However, the latter may only be true for DI RNAs, because in

order to be packaged they must compete with the (PS-containing) genomes of helper viruses. In the intact viral genome the function of the coronavirus PS resembles that of the alphavirus PS, which confers selective packaging of its gRNA over that of a single, more numerous sgRNA. PS-negative alphavirus mutants, like their coronavirus counterparts, package both gRNA and sgRNA (Kim et al., 2011). In contrast, PS-negative alphavirus mutants have much more severely growth-defective phenotypes, likely reflecting a limitation on the amount of RNA that can be accommodated by an icosahedral capsid. The helical coronavirus nucleocapsid, surrounded by a pleomorphic membrane envelope, is apparently considerably more tolerant to the inclusion of extraneous RNA.

5. *In vivo* role of gRNA packaging selectivity

The minimal phenotype of MHV packaging mutants in tissue culture raised the question of whether packaging selectivity has a significant function. To address this issue, Athmer et al. (2018) examined viral replication in the mouse and discovered that packaging plays a decisive role in evasion of host innate immunity. This was most dramatically demonstrated by constructing the set of 20 silPS mutations in the highly neurovirulent JHM strain of MHV. Mice infected with the MHV-JHM-silPS mutant had decreased weight loss and greatly enhanced survival, compared to the substantial weight loss and nearly 100% mortality of those infected with wild-type MHV-JHM. Additionally, the mutant produced lower viral titers in the brains of infected mice and was cleared more rapidly than the wild type. This attenuation was shown to be strongly dependent upon type-I interferon (IFN) signaling, as virulence of the MHV-JHM-silPS mutant was almost completely restored in IFN alpha/beta receptor-knockout (IFNAR[-/-]) mice. However, the mutant remained attenuated in either mitochondrial antiviral signaling protein-knockout (MAVS[-/-]) mice or in toll-like receptor 7-knockout (TLR7[-/-]) mice. These results indicated that neither MAVS nor TLR7 signaling alone can account for attenuation, suggesting that the two pathways are redundant or else that other circuitry is involved in initiation of the IFN response to the packaging mutant virus.

One interpretation of these findings would be that packaged sgRNA exhibits a pathogen-associated molecular pattern detectable by an innate immune sensor that is somehow blind to gRNA (Athmer et al., 2018). This might suggest that MHV genome replication introduces base modifications to gRNA that are absent from, or inefficiently
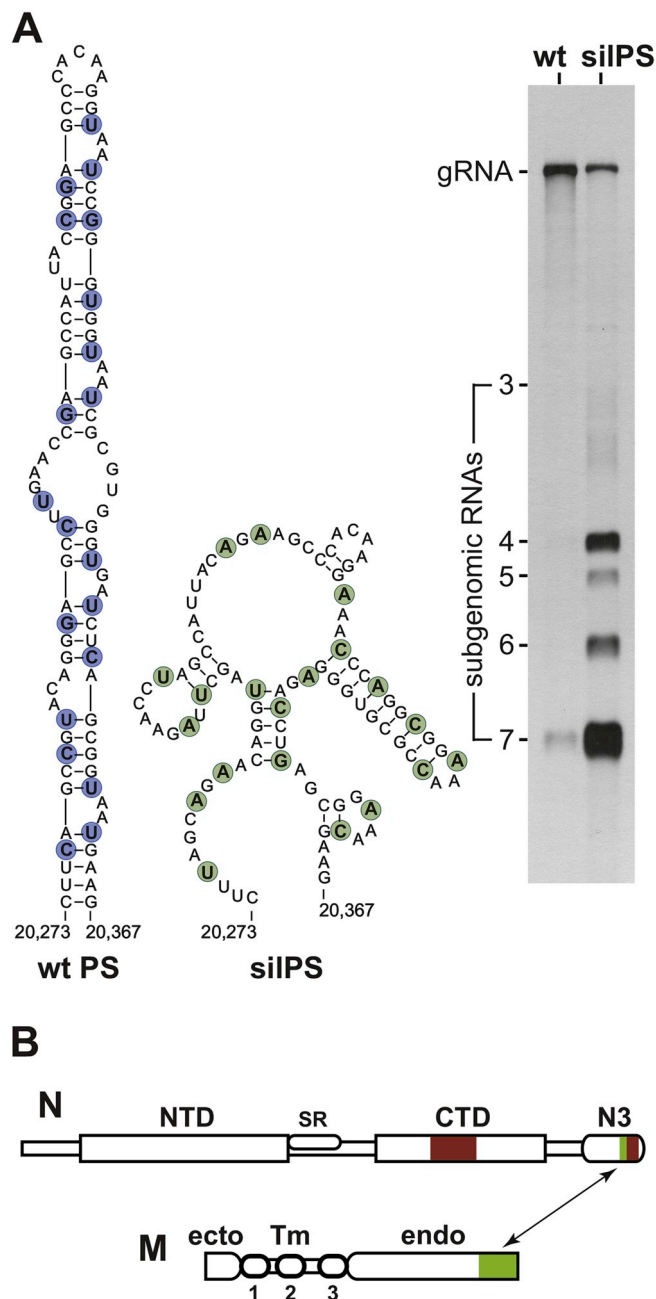
Fig. 4. Function and recognition of the PS. (A) Role of the PS in the viral genome. *Left*, the primary sequence and RNA secondary structure of the wild-type (wt) MHV PS was disrupted with 20 coding-silent mutations to create a mutant designated silPS (Kuo and Masters, 2013). Altered nucleotides are those highlighted in blue in the wt PS structure and in green in the silPS structure. *Right*, a Northern blot of RNA from highly purified virions of wt MHV and the silPS mutant detected with a probe specific for the 3′ end of the genome, which is common to gRNA and all sgRNAs. (B) Linear schematics of the MHV N and M proteins. In the N protein, there are two RNA-binding structural domains (NTD and CTD) and the carboxy-terminal domain (N3). At the amino terminus, and connecting the three domains, are unstructured segments; the central spacer contains a serine- and arginine-rich region (SR). In the M protein, the amino-terminal ectodomain (ecto) is connected to the carboxy-terminal endodomain (endo) by three transmembrane segments (Tm). *Red*, loci of substitutions or point mutations of N protein, in either the CTD (Kuo et al., 2014) or in N3 (Kuo et al., 2016b), which separately abolish packaging selectivity in viruses that contain a wt PS. *Green*, loci of critical residues in N protein (Hurst et al., 2005; Verma et al., 2006) and M protein (Escors et al., 2001; Kuo and Masters, 2002; Verma et al., 2007) essential for N-M virion assembly interactions (arrow).

applied to, sgRNAs during transcription. A different, but less likely, possibility is that the PS itself antagonizes an innate immune sensor. Irrespective of mechanism, the critical *in vivo* impairment of MHV packaging mutants argues that gRNA packaging selectivity must be an essential property of all coronaviruses. Alternatively, coronaviruses that fail to maintain gRNA packaging selectivity would need to have evolved activities that suppress the particular components of innate immunity that become activated by packaging-negative virions.

6. Recognition of the packaging signal

The two candidates for a viral molecular partner that recognizes the PS are the N protein and the M protein. The coronavirus N protein (ca. 50 kDa) comprises two highly basic, independently-folded domains, designated NTD and CTD, and a mostly acidic carboxy-terminal domain, N3 (Fig. 4B). The NTD and CTD are each RNA-binding modules, and they exhibit high structural homology among different species across multiple genera (Chang et al., 2014); additionally, the CTD serves as the dimerization domain for N molecules. Spacer segments appear at the amino terminus and linking the three domains of N; the central spacer contains a serine- and arginine-rich region (SR). In contrast to the NTD and CTD, N3 and the spacer segments are thought to be intrinsically disordered polypeptides.

The M protein (ca. 30 kDa) consists of a small ectodomain, three transmembrane domains, and a large endodomain (Fig. 4B). The endodomain has a compact globular structure, only the final 20 residues of which constitute an accessible carboxy-terminal tail. The functional units of M are also dimers, which are connected into larger oligomeric arrays in the membrane (Neuman et al., 2011).

N protein wraps the viral genome into a helical nucleocapsid that is bound to the network of M endodomains at the site of virion budding, which is the endoplasmic reticulum-Golgi intermediate compartment. N and M molecules associate with each other via their respective carboxy termini (Fig. 4B). Genetic and biochemical analyses have localized critical residues in both N (Hurst et al., 2005; Verma et al., 2006) and M (Escors et al., 2001; Kuo and Masters, 2002; Verma et al., 2007) that are essential for these assembly interactions. In virion ultrastructural studies, the N-M interaction has been visualized as what has been described as thread-like connections between the two proteins (Bárcena et al., 2009; Neuman et al., 2011). M is the most abundant virion structural protein, present in a roughly 1.4:1 molar ratio to N protein (recomputed from Sturman et al., 1980). Thus, in principle, there are sufficient M endodomain tails on the internal surface of the virion membrane to accommodate an interaction with every domain N3 in the nucleocapsid. However, there is structural evidence that virion M endodomains actually exist in two conformations, compact and elongated, with only the latter making contact with the nucleocapsid (Neuman et al., 2011).

6.1. Earlier studies of PS binding by N protein

Although N protein must possess broadly nonspecific RNA-binding activity in order to coat the entire viral genome, efforts have been made to identify sequence-specific RNA substrates that may be relevant to packaging or other functions. Molenkamp and Spaan (1997) were able to demonstrate specific binding to the MHV PS by N protein in lysates from MHV-infected cells, using either electrophoretic mobility shift assays or UV crosslinking followed by immunoprecipitation. Specificity for the PS RNA was also observed by the latter method with N purified from virions or N expressed in the absence of other viral proteins.

A contrasting view of this interaction came from a study by Cologna et al. (2000), which made use of a filter-binding assay to gauge the affinity of RNA binding by N protein in MHV-infected cell lysates. This was applied to RNA substrates spanning the entire 5.5 kb of one of the smaller of the packaged MHV DI RNAs (van der Most et al., 1991). From this work it was found that, although the PS was bound more efficiently

than non-viral RNA, the regions of the DI RNA having the highest efficiency of binding by N protein were located outside of the PS. This result did not appear to support a role in which the PS serves simply as the primary nucleation site for gRNA encapsidation.

6.2. Studies indicating M protein is the principal, and possibly sole, PS recognition element

The determination that there are multiple genomic regions of preferred binding by N protein was consistent with the observation by multiple groups that anti-N antibodies coimmunoprecipitated all positive-sense viral RNAs, both gRNA and all sgRNAs, from MHV- or BCoV-infected cell lysates (Baric et al., 1988; Cologna et al., 2000; Narayanan et al., 2000). Further pursuit of this observation by Narayanan et al. (2000) led to unexpected results. These investigators characterized the N-M interaction in infected cells, showing that anti-N antibody could coimmunoprecipitate M protein, and conversely, anti-M antibody could coimmunoprecipitate N protein. Moreover, pulse-labeling experiments demonstrated that M protein brought down by anti-N antibody was mostly unglycosylated; the O-glycosylation of MHV M protein occurs in the Golgi. Therefore, M and N associated in a pre-Golgi compartment assumed to be the virion budding site. Most significantly, in contrast to the precipitation of the full spectrum of viral RNA species by anti-N antibody, anti-M antibody coimmunoprecipitated only gRNA. Finally, it was found that the N-M interaction could not be initiated merely by co-expression of N and M, suggesting a requirement for some other viral component.

In a follow-up report, the PS was identified as that other required viral component (Narayanan and Makino, 2001). Analysis of a large set of constructed DI RNAs demonstrated that all of them were intracellularly associated with N protein, irrespective of the presence or absence of the PS. However, intracellular association of M protein with a given DI RNA correlated completely with the packaging ability of that DI RNA, which, in turn, was absolutely dependent upon the presence of the PS. Furthermore, it was shown that in MHV-infected cells N protein became bound to an expressed (nonreplicating) nonviral RNA (CAT gene RNA), whether or not it harbored the PS. Only the version of CAT gene RNA containing the PS was bound by M protein and was packaged into virions.

This work was brought to a surprising conclusion in a subsequent study in which the basic coimmunoprecipitation experiments were recapitulated using only MHV components that were expressed from Sindbis virus vectors, in addition to RNA substrates that were produced from transfected plasmids with vaccinia virus-expressed T7 RNA polymerase (Narayanan et al., 2003). Here it was shown that anti-M antibody precipitated CAT gene (PS+) RNA, but not CAT gene (PS-) RNA, from cells that were expressing both M protein and N protein. Remarkably, the same result was obtained if only M protein, but not N protein, was expressed. Finally, it was demonstrated that CAT gene (PS +) RNA was packaged into VLPs produced solely by expression of the MHV M and E proteins, in the absence of N protein. These results strongly implied that M protein (possibly with the assistance of E) is the viral component that recognizes and binds to the PS, and that formation of a nucleocapsid with N protein is not required for packaging specificity.

6.3. Studies implicating roles for both the CTD and domain N3 of N protein in PS recognition

The presence in the coronavirus N protein of two distinct RNA-binding domains, the NTD and the CTD, is an unusual characteristic for an RNA virus nucleocapsid protein (Chang et al., 2014). To probe the functional significance of this arrangement, a genetic study was carried out in which each of these domains in the MHV N protein was substituted by its counterpart from the SARS-CoV N protein (Kuo et al., 2014). The NTDs and CTDs of MHV and SARS-CoV have 44% and 35%

amino-acid identity, respectively. This degree of homology was expected to be sufficiently close to allow such substitutions to be viable, while also being sufficiently divergent to uncover sequence-specific interactions between N protein domains and other viral components. Indeed, both substitutions were well tolerated by MHV, and neither affected the N protein-M protein interaction or the fidelity of the leader-body sgRNA junctions formed during transcription. However, substitution of the CTD profoundly affected packaging. Notably, the SARS-CoV CTD substitution mutant had the same packaging-defective phenotype as that of the silPS mutant (above, Fig. 4A), despite having a completely wild-type copy of the PS. By contrast, the NTD substitution mutant had a wild-type packaging phenotype. Further construction of partial SARS-CoV chimeras of the MHV N CTD showed that substitution by just a 30-amino-acid central region of the SARS-CoV CTD was sufficient to abolish packaging (Fig. 4B). As discussed above, the MHV PS is unique to lineage A betacoronaviruses, and the genome of SARS-CoV (a lineage B betacoronavirus) does not contain a homolog of this RNA structure (Chen and Olsthoorn, 2010). Thus, the SARS-CoV N protein likely would not have evolved to recognize the MHV PS. This outcome was reinforced by the subsequent demonstration that a mutant with a substitution of the even more divergent N protein CTD of the alpha-coronavirus TGEV also exhibited a packaging-defective phenotype (Kuo et al., 2016b).

These results provided strong support for a role for the CTD of the N protein in recognition of the PS. This finding (distinguishing the CTD, not the NTD) was somewhat counterintuitive, based on the many unliganded structures that have been solved for coronavirus N protein NTDs and CTDs (Chang et al., 2014). The NTD displays a beta-platform presenting a potential RNA-binding groove rich in basic and aromatic amino-acid residues, the latter of which are speculated to be able to form sequence-specific contacts with RNA bases. The CTD dimer, on the other hand, is a rectangular slab with a putative RNA-binding groove on one face that is lined with basic residues thought to be best suited for nonspecific interactions with the phosphodiester backbone. Nevertheless, the exact character of RNA binding by either the NTD or CTD awaits determination of structures containing bound RNA.

A separate region of the N molecule was later shown to also be involved in PS recognition (Kuo et al., 2016b). This finding resulted from examination of the effect of complete or partial replacement of the MHV M protein by the M protein of SARS-CoV. Such substitutions abolished packaging, but it could not be concluded that this defect was solely due to M. The construction of SARS-CoV M chimeric mutants was possible only because they also incorporated substitution of the SARS-CoV domain N3, the carboxy-terminal tail of N protein that is necessary and sufficient for virion assembly interactions between N and M. To clarify this outcome, multiple previously isolated MHV domain N3 mutants (all containing wild-type MHV M protein) were screened for their packaging ability, and some were found to have lost selective packaging of gRNA. As with the CTD mutants, packaging-defective domain N3 mutants had the same phenotype as that of the silPS mutant but had a completely intact wild-type PS. Comparison of packaging-competent and packaging-defective viruses allowed mapping of the packaging defect to a 9-amino-acid segment of domain N3 (Fig. 4B) that was adjacent to, but distinct from, key residues that interact with M protein in virion assembly (Hurst et al., 2005).

Together, these studies implicated both the N protein CTD and domain N3 as major determinants of PS recognition, a conclusion markedly in contrast with that of prior studies that suggested the primary role for M protein. To date, genetic evidence has not been obtained that could either rule in or rule out participation of the M protein in PS recognition, in part because the M protein endodomain is much less tolerant to mutational alteration than is the N protein (Kuo et al., 2016a).
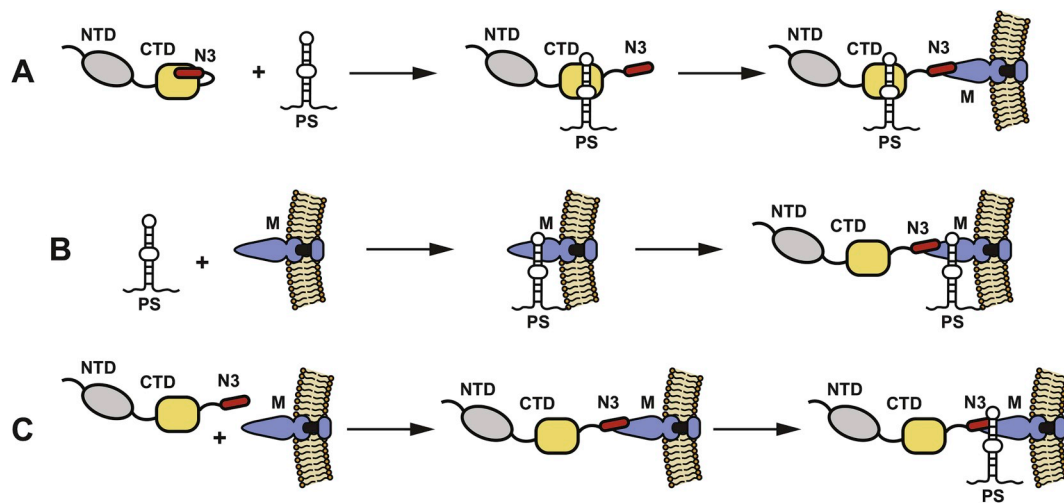
Fig. 5. Models for the mechanism of coronavirus gRNA packaging. (A) The N protein CTD binds to the PS, concomitantly releasing domain N3; N3 then binds to the M protein endodomain. (B) The M protein binds to the PS, initiating encapsidation of the genome by N protein. (C) Complex formation by the N and M proteins precedes PS recognition. See text for details. For simplicity, in all three models subsequent cooperative N-RNA, N-N, N-M, and M-M interactions have been omitted.

7. Models for the mechanism of packaging

MHV packaging studies based on genetic manipulations and studies based on analyses of DI RNAs and VLPs thus seem at odds as to whether M protein or N protein is the dominant player in choosing gRNA for virion assembly. However, these two sets of results are not necessarily mutually exclusive, and they possibly will be brought into alignment by further work. The apparent discrepancy between the outcomes of the two approaches may be analogous to the difference, discussed above, between the observed role of the MHV PS in DI RNA systems as opposed to its role in the intact viral genome. At this stage, the currently available data suggest three types of models for how coronavirus packaging selectivity is achieved (Fig. 5).

7.1. Primary PS recognition by N protein

One model, which is derived from the genetic studies (Kuo et al., 2014 and Kuo et al., 2016b), builds upon the known roles of the CTD as an RNA-binding module (Chang et al., 2014) and of domain N3 as the sole region of N that interacts with M protein (Hurst et al., 2005; Verma et al., 2006; Kuo et al., 2016a). In this proposed mechanism, the largely acidic domain N3 is originally sequestered by the CTD and then becomes dislodged, directly or indirectly, only as a result of the CTD binding to the PS (Fig. 5A). This may come about because the PS competes for the same site on the CTD that is occupied by the N3 peptide, or alternatively, because binding of the PS induces a conformational change in the CTD that frees N3 from a separate site. The resulting CTD-PS interaction would nucleate cooperative binding of N monomers along the whole length of the gRNA, with each monomer releasing its N3 domain to bind to an M monomer endodomain. The key feature of this model is that domain N3 is not available to interact with M protein until the CTD has bound to the PS. Thereby, RNA molecules that do not contain the PS are excluded from packaging, even though they may be nonspecifically bound by other N monomers.

This model invokes a mode of RNA-binding specificity similar to that observed for the spliceosomal protein U1A and its target substrate, U1 hairpin II RNA (Law et al., 2013). In free U1A, a short carboxy-terminal helix occludes the beta-sheet platform of the RNA-binding domain of the molecule; only the sequence-specific U1 hairpin II can initiate events that displace the helix to form part of the RNA-binding surface. The role of domain N3 in the model is also akin to the chaperone-like function of the amino terminus of the paramyxovirus phosphoprotein, which maintains its nucleocapsid protein in an open

state to prevent nonspecific RNA binding (Yabukarski et al., 2014).

7.2. Primary PS recognition by M protein

A second model, which directly follows from the results of Narayanan et al. (2003), proposes that the M protein endodomain, which has oligomerized on the intracellular membrane of the assembly site, specifically recognizes the PS. This M-PS binding event then serves as the nucleation point for the condensation of all other M-N and N-N interactions that drive budding (Fig. 5B). Although N protein is most likely the first virion structural protein to encounter the genome during, or shortly following, gRNA synthesis in the replication compartment (Bost et al., 2001; Stertz et al., 2007), its initial mode of binding to gRNA is proposed to be the same as that to all sgRNAs. The specific binding of M to the PS is thus envisioned to trigger formation of the helical ribonucleoprotein by N and concomitant incorporation of the nucleocapsid into nascent virions (Narayanan and Makino, 2001; Narayanan et al., 2003). This would suggest that in each assembled virion, there exists a unique point of protein-gRNA contact, between M and the PS, which cannot be discerned at the current level of resolution of ultrastructural work (Bárcena et al., 2009; Neuman et al., 2011).

At the time this model was proposed there was no precedent for specific RNA binding by a transmembrane protein (Narayanan et al., 2003). Subsequently, however, there have emerged multiple examples of RNA recognition by membrane-bound cellular or viral proteins (Einav et al., 2008; Hsu et al., 2018). In particular, components of the essential Sec61 protein translocation pore complex form structurally well-defined bridges with certain rRNA helices of the large ribosomal subunit (Becker et al., 2009).

7.3. Complex formation between N and M proteins, followed by recognition of the PS

A third possible model is that neither the N protein nor the M protein by itself can bind to the PS, but that prior association of domain N3 and the M endodomain creates a surface that recognizes the PS (Fig. 5C). This model would seem to be ruled out by the existence of N3 mutants that have fully retained normal N-M assembly but have lost selective gRNA packaging (Kuo et al., 2016b). Nevertheless, it has been shown that not only the two structural domains, the NTD and CTD, but also unstructured segments of N, including N3, contribute to the affinity of RNA binding (Chang et al., 2009). Thus, it is conceivable that a unique conformational state of N induced by association with M could

generate or stabilize a specific PS-binding site.

8. Future perspectives

Considerable further interrogation of the above models remains to be done to obtain a more complete understanding of coronavirus genome packaging. Genetic methods have yielded insights into the participation of N protein, but attempts to construct MHV mutants with partially chimeric M protein endodomains have been largely unsuccessful (Kuo et al., 2016b). A more productive approach might be to target basic amino-acid residues in the M endodomain that are common to lineage A betacoronaviruses but are not conserved in the other lineages. For N protein, a similar strategy applied to clusters of charged surface residues of the CTD could pinpoint interactions with domain N3 or the PS. Complementary biochemical investigations need to be undertaken to examine the RNA-binding properties of full-length N protein, as well constructs of the CTD and CTD-N3. The methods used in many recent studies to examine N binding to nonspecific RNA and DNA (Chang et al., 2009, 2014) ought to now be applied to PS RNA. *In vitro* analysis of RNA binding by M protein would also be important but necessarily more difficult owing to the need to maintain M in a oligomeric membrane-bound state.

Perhaps the largest contribution to be made to our knowledge of packaging mechanisms will come from structural biology. Understanding virion assembly cannot be separated from the need to more generally comprehend the molecular details of coronavirus N-N and N-RNA interactions. Although structures are currently available for 6 NTDs and 4 CTDs from different viruses, there is as yet no structure for an RNA-bound complex of either of these domains. Likewise, no structure has been solved for an entire N protein, with or without an RNA ligand, but a recent low-resolution cryo-EM reconstruction of MHV N has taken the first step in that direction (Gui et al., 2017). At the next level, we lack a clear picture of the coronavirus nucleocapsid. More basically, we lack even fundamental parameters, such as the stoichiometry of N to RNA. One speculative model of the nucleocapsid has 7 nucleotides of RNA being bound by one monomer of N (Chang et al., 2014); however, another estimate places this ratio at 14 to 40 nucleotides per N monomer (Neuman et al., 2011). A high-resolution nucleocapsid structure would provide details of how each NTD and CTD encapsidates gRNA and why the helical coronavirus nucleocapsid is much more flexible than the helical nucleocapsids of rhabdo- and paramyxoviruses. Similarly, a more detailed model of M protein structure, oligomerization, and M-N interactions in the virion would be highly valuable, although this will be a still more technically daunting achievement.

Another important future goal would be to proceed beyond lineage A betacoronaviruses to identify PSs across all four genera of the coronavirus family. It is possible that suitably constructed MHV chimeras could be used to trap heterologous PSs. Although PS identity is genus- or lineage-specific, the precise definition of more of these RNA elements and their interacting protein partners may reveal commonalities that are not currently apparent and shed light on unifying principles of coronavirus PS recognition.

Finally, moving beyond lineage A betacoronaviruses would create the potential for utilization of knowledge about genomic RNA packaging to devise antiviral strategies for pathogenic human coronaviruses. The discovery that a packaging-defective MHV mutant was markedly suppressed by host innate immunity suggests an attractive pathway toward live-attenuated vaccine design (Athmer et al., 2018). The PS itself may also become a candidate for therapeutic intervention as RNA structures are being increasingly developed as small-molecule druggable targets (Anokhina et al., 2019; Ingemarsdotter et al., 2018). Additionally, more precise elucidation of packaging-specific oligomeric interactions could uncover molecular targets that would hinder the escape of drug-resistant viral mutants (Tanner et al., 2014).

References

Anokhina, V.S., McAnany, J.D., Ciesla, J.H., Hilimire, T.A., Santoso, N., Miao, H., Miller, B.L., 2019. Enhancing the ligand efficiency of anti-HIV compounds targeting frameshift-stimulating RNA. Bioorg. Med. Chem. 27, 2972–2977.

Athmer, J., Fehr, A.R., Grunewald, M., Smith, E.C., Denison, M.R., Perlman, S., 2017. In situ tagged nsp15 reveals interactions with coronavirus replication/transcription complex-associated proteins. mBio 8 e02320-16.

Athmer, J., Fehr, A.R., Grunewald, M.E., Qu, W., Wheeler, D.L., Graepel, K.W., Channappanavar, R., Sekine, A., Aldabeeb, D.S., Gale Jr., M., Denison, M.R., Perlman, S., 2018. Selective packaging in murine coronavirus promotes virulence by limiting type I interferon responses. mBio 9 e00272-18.

Bárcena, M., Oostergetel, G.T., Bartelink, W., Faas, F.G., Verkleij, A., Rottier, P.J., Koster, A.J., Bosch, B.J., 2009. Cryo-electron tomography of mouse hepatitis virus: insights into the structure of the coronavirion. Proc. Natl. Acad. Sci. U.S.A. 106, 582–587.

Baric, R.S., Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N., Stohlman, S.A., 1988. Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. J. Virol. 62, 4280–4287.

Barrows, N.J., Campos, R.K., Liao, K.C., Prasanth, K.R., Soto-Acosta, R., Yeh, S.C., Schott-Lerner, G., Pompon, J., Sessions, O.M., Bradrick, S.S., Garcia-Blanco, M.A., 2018. Biochemistry and molecular biology of flaviviruses. Chem. Rev. 118, 4448–4482.

Becker, T., Bhushan, S., Jarasch, A., Armache, J.P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., Westhof, E., Gilmore, R., Mandon, E.C., Beckmann, R., 2009. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. Science 326, 1369–1373.

Bolte, H., Rosu, M.E., Hagelauer, E., García-Sastre, A., Schwemmle, M., 2019. Packaging of the influenza virus genome is governed by a plastic network of RNA- and nucleoprotein-mediated interactions. J. Virol. 93 e01861-18.

Borodavka, A., Tuma, R., Stockley, P.G., 2012. Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proc. Natl. Acad. Sci. U.S.A. 109, 15769–15774.

Bos, E.C., Luytjes, W., van der Meulen, H.V., Koerten, H.K., Spaan, W.J., 1996. The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. Virology 218, 52–60.

Bost, A.G., Prentice, E., Denison, M.R., 2001. Mouse hepatitis virus replicase protein complexes are translocated to sites of M protein accumulation in the ERGIC at late times of infection. Virology 285, 21–29.

Chang, C.K., Hou, M.H., Chang, C.F., Hsiao, C.D., Huang, T.H., 2014. The SARS coronavirus nucleocapsid protein–forms and functions. Antivir. Res. 103, 39–50.

Chang, C.K., Hsu, Y.L., Chang, Y.H., Chao, F.A., Wu, M.C., Huang, Y.S., Hu, C.K., Huang, T.H., 2009. Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J. Virol. 83, 2255–2264.

Chen, S.C., van den Born, E., van den Worm, S.H., Pleij, C.W., Snijder, E.J., Olsthoorn, R.C., 2007. New structure model for the packaging signal in the genome of group IIa coronaviruses. J. Virol. 81, 6771–6774.

Chen, S.C., Olsthoorn, R.C., 2010. Group-specific structural features of the 5'-proximal sequences of coronavirus genomic RNAs. Virology 401, 29–41.

Chou, Y.Y., Vafabakhsh, R., Doganay, S., Gao, Q., Ha, T., Palese, P., 2012. One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. Proc. Natl. Acad. Sci. U.S.A. 109, 9101–9106.

Cologna, R., Hogue, B.G., 2000. Identification of a bovine coronavirus packaging signal. J. Virol. 74, 580–583.

Cologna, R., Spagnolo, J.F., Hogue, B.G., 2000. Identification of nucleocapsid binding sites within coronavirus-defective genomes. Virology 277, 235–249.

Dalton, K., Casais, R., Shaw, K., Stirrups, K., Evans, S., Britton, P., Brown, T.D., Cavanagh, D., 2001. Cis-acting sequences required for coronavirus infectious bronchitis virus defective-RNA replication and packaging. J. Virol. 75, 125–133.

Deng, X., Baker, S.C., 2018. An "old" protein with a new story: coronavirus endoribonuclease is important for evading host antiviral defenses. Virology 517, 157–163.

de Wit, E., van Doremalen, N., Falzarano, D., Munster, V.J., 2016. SARS and MERS: recent insights into emerging coronaviruses. Nat. Rev. Microbiol. 14, 523–534.

Einav, S., Gerber, D., Bryson, P.D., Sklan, E.H., Elazar, M., Maerkl, S.J., Glenn, J.S., Quake, S.R., 2008. Discovery of a hepatitis C target and its pharmacological inhibitors by microfluidic affinity analysis. Nat. Biotechnol. 26, 1019–1027.

Escors, D., Izeta, A., Capiscol, C., Enjuanes, L., 2003. Transmissible gastroenteritis coronavirus packaging signal is located at the 5' end of the virus genome. J. Virol. 77, 7890–7902.

Escors, D., Ortego, J., Laude, H., Enjuanes, L., 2001. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. J. Virol. 75, 1312–1324.

Fosmire, J.A., Hwang, K., Makino, S., 1992. Identification and characterization of a coronavirus packaging signal. J. Virol. 66, 3522–3530.

Gui, M., Liu, X., Guo, D., Zhang, Z., Yin, C.C., Chen, Y., Xiang, Y., 2017. Electron microscopy studies of the coronavirus ribonucleoprotein complex. Protein Cell 8, 219–224.

Hofmann, M.A., Sethna, P.B., Brian, D.A., 1990. Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. J. Virol. 64, 4108–4114.

Hsieh, P.K., Chang, S.C., Huang, C.C., Lee, T.T., Hsiao, C.W., Kou, Y.H., Chen, I.Y., Chang, C.K., Huang, T.H., Chang, M.F., 2005. Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. J. Virol. 79, 13848–13855.

Hsin, W.C., Chang, C.H., Chang, C.Y., Peng, W.H., Chien, C.L., Chang, M.F., Chang, S.C., 2018. Nucleocapsid protein-dependent assembly of the RNA packaging signal of Middle East respiratory syndrome coronavirus. J. Biomed. Sci. 25, 47.

Hsu, J.C., Reid, D.W., Hoffman, A.M., Sarkar, D., Nicchitta, C.V., 2018. Oncoprotein AEG-1 is an endoplasmic reticulum RNA-binding protein whose interactome is enriched in organelle resident protein-encoding mRNAs. RNA 24, 688–703.

Hu, H., Jung, K., Vlasova, A.N., Chepngeno, J., Lu, Z., Wang, Q., Saif, L.J., 2015. Isolation and characterization of porcine deltacoronavirus from pigs with diarrhea in the United States. J. Clin. Microbiol. 53, 1537–1548.

Hurst, K.R., Kuo, L., Koetzner, C.A., Ye, R., Hsue, B., Masters, P.S., 2005. A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. J. Virol. 79, 13285–13297.

Ingemarsdotter, C.K., Zeng, J., Long, Z., Lever, A.M.L., Kenyon, J.C., 2018. An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. Retrovirology 15, 25.

Jiang, P., Liu, Y., Ma, H.C., Paul, A.V., Wimmer, E., 2014. Picornavirus morphogenesis. Microbiol. Mol. Biol. Rev. 78, 418–437.

Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2007. Crystal structure of a monomeric form of severe acute respiratory syndrome coronavirus endonuclease nsp15 suggests a role for hexamerization as an allosteric switch. J. Virol. 81, 6700–6708.

Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., Bolden, N.C., McCowin, S., Case, D.A., Johnson, B.A., Salemi, M., Telesnitsky, A., Summers, M.F., 2015. Structure of the HIV-1 RNA packaging signal. Science 348, 917–921.

Keane, S.C., Van, V., Frank, H.M., Sciandra, C.A., McCowin, S., Santos, J., Heng, X., Summers, M.F., 2016. NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. Proc. Natl. Acad. Sci. U.S.A. 113, 13033–13038.

Kim, D.Y., Firth, A.E., Atasheva, S., Frolova, E.I., Frolov, I., 2011. Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. J. Virol. 85, 8022–8036.

Kuo, L., Koetzner, C.A., Masters, P.S., 2016b. A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. Virology 494, 100–107.

Kuo, L., Masters, P.S., 2002. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. J. Virol. 76, 4987–4999.

Kuo, L., Masters, P.S., 2013. Functional analysis of the murine coronavirus genomic RNA packaging signal. J. Virol. 87, 5182–5192.

Kuo, L., Hurst-Hess, K.R., Koetzner, C.A., Masters, P.S., 2016a. Analyses of coronavirus assembly interactions with interspecies membrane and nucleocapsid protein chimeras. J. Virol. 90, 4357–4368.

Kuo, L., Koetzner, C.A., Hurst, K.R., Masters, P.S., 2014. Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid protein. J. Virol. 88, 4451–4465.

Lau, S.K., Woo, P.C., Yip, C.C., Fan, R.Y., Huang, Y., Wang, M., Guo, R., Lam, C.S., Tsang, A.K., Lai, K.K., Chan, K.H., Che, X.Y., Zheng, B.J., Yuen, K.Y., 2012. Isolation and characterization of a novel betacoronavirus subgroup A coronavirus, rabbit coronavirus HKU14, from domestic rabbits. J. Virol. 86, 5481–5496.

Law, M.J., Lee, D.S., Lee, C.S., Anglim, P.P., Haworth, I.S., Laird-Offringa, I.A., 2013. The role of the C-terminal helix of U1A protein in the interaction with U1hpII RNA. Nucleic Acids Res. 41, 7092–7100.

Makino, S., Yokomori, K., Lai, M.M., 1990. Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. J. Virol. 64, 6045–6053.

Masters, P.S., 2006. The molecular biology of coronaviruses. Adv. Virus Res. 66, 193–292.

Masters, P.S., Perlman, S., 2013. Coronaviridae. p 825-858. In: Knipe, D.M., Howley, P.M., Cohen, J.I., Griffin, D.E., Lamb, R.A., Martin, M.A., Racaniello, V.R., Roizman, B. (Eds.), Fields Virology, vol. 1 Lippincott Williams & Wilkins, Philadelphia, PA.

Mendes, A., Kuhn, R.J., 2018. Alphavirus nucleocapsid packaging and assembly. Viruses 10, E138.

Méndez, A., Smerdou, C., Izeta, A., Gebauer, F., Enjuanes, L., 1996. Molecular characterization of transmissible gastroenteritis coronavirus defective interfering genomes: packaging and heterogeneity. Virology 217, 495–507.

Molenkamp, R., Spaan, W.J., 1997. Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. Virology 239, 78–86.

Morales, L., Mateos-Gomez, P.A., Capiscol, C., del Palacio, L., Enjuanes, L., Sola, I., 2013. Transmissible gastroenteritis coronavirus genome packaging signal is located at the 5' end of the genome and promotes viral RNA incorporation into virions in a replication-independent process. J. Virol. 87, 11579–11590.

Narayanan, K., Makino, S., 2001. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. J. Virol. 75, 9059–9067.

Narayanan, K., Chen, C.J., Maeda, J., Makino, S., 2003. Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. J. Virol. 77, 2922–2927.

Narayanan, K., Maeda, A., Maeda, J., Makino, S., 2000. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. J. Virol. 74, 8127–8134.

Nemoto, M., Oue, Y., Murakami, S., Kanno, T., Bannai, H., Tsujimura, K., Yamanaka, T., Kondo, T., 2015. Complete genome analysis of equine coronavirus isolated in Japan. Arch. Virol. 160, 2903–2906.

Neuman, B.W., Adair, B.D., Yoshioka, C., Quispe, J.D., Orca, G., Kuhn, P., Milligan, R.A., Yeager, M., Buchmeier, M.J., 2006. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. J. Virol. 80, 7918–7928.

Neuman, B.W., Joseph, J.S., Saikatendu, K.S., Serrano, P., Chatterjee, A., Johnson, M.A., Liao, L., Klaus, J.P., Yates 3rd, J.R., Wüthrich, K., Stevens, R.C., Buchmeier, M.J., Kuhn, P., 2008. Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. J. Virol. 82, 5279–5294.

Neuman, B.W., Kiss, G., Kunding, A.H., Bhella, D., Baksh, M.F., Connelly, S., Droese, B., Klaus, J.P., Makino, S., Sawicki, S.G., Siddell, S.G., Stamou, D.G., Wilson, I.A., Kuhn, P., Buchmeier, M.J., 2011. A structural analysis of M protein in coronavirus assembly and morphology. J. Struct. Biol. 174, 11–22.

Nugent, C.I., Johnson, K.L., Sarnow, P., Kirkegaard, K., 1999. Functional coupling between replication and packaging of poliovirus replicon RNA. J. Virol. 73, 427–435.

Onafuwa-Nuga, A.A., Telesnitsky, A., King, S.R., 2006. 7SL RNA, but not the 54-kd signal recognition particle protein, is an abundant component of both infectious HIV-1 and minimal virus-like particles. RNA 12, 542–546.

Sarute, N., Ross, S.R., 2017. New world arenavirus biology. Annu Rev Virol 4, 141–158.

Sethna, P.B., Hung, S.L., Brian, D.A., 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. Proc. Natl. Acad. Sci. U.S.A. 86, 5626–5630.

Song, Y., Gorbatsevych, O., Liu, Y., Mugavero, J., Shen, S.H., Ward, C.B., Asare, E., Jiang, P., Paul, A.V., Mueller, S., Wimmer, E., 2017. Limits of variation, specific infectivity, and genome packaging of massively recoded poliovirus genomes. Proc. Natl. Acad. Sci. U.S.A. 114, E8731–E8740.

Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martínez-Sobrido, L., García-Sastre, A., Weber, F., Kochs, G., 2007. The intracellular sites of early replication and budding of SARS-coronavirus. Virology 361, 304–315.

Sturman, L.S., Holmes, K.V., Behnke, J., 1980. Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. J. Virol. 33, 449–462.

Tanner, E.J., Liu, H.M., Oberste, M.S., Pallansch, M., Collett, M.S., Kirkegaard, K., 2014. Dominant drug targets suppress the emergence of antiviral resistance. Elife 3 e03830.

van der Most, R.G., Bredenbeek, P.J., Spaan, W.J., 1991. A domain at the 3' end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. J. Virol. 65, 3219–3226.

Verma, S., Bednar, V., Blount, A., Hogue, B.G., 2006. Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. J. Virol. 80, 4344–4355.

Verma, S., Lopez, L.A., Bednar, V., Hogue, B.G., 2007. Importance of the penultimate positive charge in mouse hepatitis coronavirus A59 membrane protein. J. Virol. 81, 5339–5348.

Xu, X., Zhai, Y., Sun, F., Lou, Z., Su, D., Xu, Y., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M., Rao, Z., 2006. New antiviral target revealed by the hexameric structure of mouse hepatitis virus nonstructural protein nsp15. J. Virol. 80, 7909–7917.

Yabukarski, F., Lawrence, P., Tarbouriech, N., Bourhis, J.M., Delaforge, E., Jensen, M.R., Ruigrok, R.W., Blackledge, M., Volchkov, V., Jamin, M., 2014. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. Nat. Struct. Mol. Biol. 21, 754–759.

Zhang, J., Guy, J.S., Snijder, E.J., Denniston, D.A., Timoney, P.J., Balasuriya, U.B., 2007. Genomic characterization of equine coronavirus. Virology 369, 92–104.

Zhang, L., Li, L., Yan, L., Ming, Z., Jia, Z., Lou, Z., Rao, Z., 2018. Structural and biochemical characterization of endoribonuclease nsp15 encoded by Middle East respiratory syndrome coronavirus. J. Virol. 92 e00893-18.

Zhao, X., Shaw, K., Cavanagh, D., 1993. Presence of subgenomic mRNAs in virions of coronavirus IBV. Virology 196, 172–178.

Ziebuhr, J., 2005. The coronavirus replicase. Curr. Top. Microbiol. Immunol. 287, 57–94.