RNA Biology

RNA-RNA and RNA-protein interactions in coronavirus replication and transcription

Isabel Sola, Pedro A. Mateos-Gomez, Fernando Almazan, Sonia Zuñiga & Luis Enjuanes

To cite this article: Isabel Sola, Pedro A. Mateos-Gomez, Fernando Almazan, Sonia Zuñiga & Luis Enjuanes (2011) RNA-RNA and RNA-protein interactions in coronavirus replication and transcription, RNA Biology, 8:2, 237-248, DOI: 10.4161/rna.8.2.14991

To link to this article: http://dx.doi.org/10.4161/rna.8.2.14991

Published online: 01 Mar 2011.

Submit your article to this journal

Article views: 461

View related articles

Citing articles: 18 View citing articles

Full Terms & Conditions of access and use can be found at http://www.tandfonline.com/action/journalInformation?journalCode=krnb20
RNA-RNA and RNA-protein interactions in coronavirus replication and transcription

Isabel Sola, Pedro A. Mateos-Gomez, Fernando Almazan, Sonia Zuñiga and Luis Enjuanes*

Department of Molecular and Cell Biology; CNB; CSIC; Cantoblanco; Madrid, Spain

Key words: coronavirus, plus-strand RNA virus, replication, transcription, RNA-RNA interactions, RNA-protein interactions

Abbreviations: CoV, coronavirus; sgRNA, subgenomic RNA; UTR, untranslated region; TRS, transcription-regulating sequences; SL, stem loop; DI RNA, defective interfering RNA; ORF, open reading frame; BSL, bulged stem-loop; PK, pseudoknot; CS, conserved core sequence; ΔG, thermodynamic free energy; nsp, non-structural protein; RTC, replication-transcription complex; hNRNP, heterogeneous nuclear ribonucleoprotein; RNP complex, ribonucleoprotein complex

Introduction

The replication and transcription strategies of RNA viral genomes involve intramolecular interactions between physically close and distal domains, or even intermolecular regulatory contacts. The long-range and local RNA-RNA contacts result in tertiary structures that modulate the function of enhancers, promoters and silencers. In addition to the direct interactions within the viral RNA, a wide spectrum of contacts between viral and cellular proteins has been identified that leads to a complex interaction network required for the fine-tuning of virus replication. This network conforms a higher order RNA-protein structure. Nowadays, limited technological procedures have studied compartmentalized aspects of these structures that represent frozen states in a dynamic equilibrium. A widely used strategy to approach RNA structure is the application of predictive algorithms that consider the viral RNA genomes free in solution. Although these theoretical predictions are useful to approach scientific questions, the results withdrawn are necessarily limited to small RNA domains. More refined studies predict RNA structure as it is being synthesized, but the real situation, the RNA in contact with viral and cellular proteins, is only addressed using reduced RNA motifs.

RNA viruses are classified according to the nature of their RNA genomes into plus-strand, minus-strand and double-stranded RNA viruses. Of these three classes, the most abundant are the plus-stranded RNA viruses. The first replication step in these viruses is the translation of their genome that acts as a viral mRNA. Viral RNA replication takes place in replication complexes generally associated to intracellular membranes. Replication and transcription of several plus-strand RNA viruses require RNA motifs frequently located within the non-coding domains of the 5' and 3' ends of the genome. Plus-stranded RNA viruses replicate their genome through a continuous process that starts with the synthesis of an intermediate full-length minus-strand RNA, which then serves as the template for the asymmetric synthesis of progeny plus-strand genomes. During replication, cis acting replication elements (CREs) residing within the coding sequence, or genome cyclization may be needed. The cyclization of RNA viral genomes may be mediated by RNA
RNA Biology
Volume 8 Issue 2

238

Figure 1. CoV genome structure and strategy of sgRNA expression. The upper line represents the TGEV genome structure. Letters above the bar indicate the gene names, L, leader sequence, AAA, poly(A). The shorter lines represent the viral mRNAs. Letters to the right of these lines indicate the gene encoded by each mRNA. Black lines represent the positive stranded RNAs, whereas gray lines represent the negative stranded RNAs. The triangles indicate the positions of the conserved core sequences (CSS) located upstream of each gene.

Many features used by CoVs for genome replication and transcription have been established using natural and artificially constructed defective interfering (DI) RNAs that conserve the cis-acting signals required for their replication in trans by a helper virus.35-41 Recently, reverse genetic systems have been developed for several CoVs, allowing the study of CoV replication and transcription mechanisms in the context of the viral genome.42-47

The minimal sequences required for genome replication have been defined in the three CoV genera by DI RNA deletion mapping analysis and, in all cases these sequences were located at both ends of the genome. For transmissible gastroenteritis virus (TGEV),35,48 from genus α, mouse hepatitis (MHV) from genus β, and infectious bronchitis virus (IBV) from genus γ, these sequences expanded the first 466 to 649 nt at the 5' end of the genome, and the 388 to 493 nt plus the poly(A) tail at the 3' end. These regions of the genome contain secondary and higher-order RNA structures, named cis-acting RNA elements that interact with RNA motifs or proteins, and are located within the highly structured 5' and 3' UTRs of the CoV genome although, sometimes, they are also embedded into the adjacent coding sequences.

5' cis-acting RNA elements. Higher-order RNA structures in the 5' end of the CoV genome were first studied in the bovine coronavirus (BCoV), where six stem-loop structures, denoted SL-1 to SL-VI (Fig. 2), were identified and confirmed by enzymatic probing and functional mutational analysis using DI RNAs.38,39-41 The first four stem-loops mapped within the 5' UTR, while the SL-V and -VI mapped into the adjacent ORF 1a coding sequence. A consensus secondary structural model for the first 5' UTR 140 nt, based on the sequence of nine CoVs representing the three genera, has been proposed.42 In this model, the region comprising the original SL-1 contains two stem-loops that within this review have been named SL-1a and SL-1b, in order to maintain the original nomenclature proposed in previous reports for the next stem-loops (Fig. 2). Recently, it has been predicted that stem-loops SL-1a and SL-1b are structurally conserved in all known CoVs,43 and their structures have been confirmed by nuclear magnetic resonance (NMR) spectroscopy in MHV and human CoV (HCoV)-OC43.60-62 A general feature of SL-1a is that it adopts a bipartite structure and, at least in MHV, the upper region of the stem-loop needs to form a double stranded stem to

interactions and also by RNA-protein and protein-protein interactions. Interestingly, long-range and local RNA-RNA interactions may be mutually exclusive, and may function as molecular switches controlling critical aspects, such as changes between replication, transcription and translation.19-22

Nidoviruses include the Coronaviridae family with two subfamilies, Coronavirinae and Torovirinae. The Coronavirinae incorporates three genera: α, β and γ; see talk.ictvonline.org/media/g/vertebrate-2008/default.aspx for official coronavirus taxonomy. The CoVs have the largest known viral RNA genome, a plus-strand RNA of 27 to 31 kb. Both viral genome and sgRNAs share common 5' and 3' termini that could enable these molecules to amplify by a partially characterized mechanism (Fig. 1). CoV genome includes a 5' end CAP and a 3' end poly(A) that may also promote 5' to 3' interactions, mediated by viral and cellular proteins. These interactions may be involved in the timely switches controlling CoV replication and transcription. These activities include the synthesis of minus-strand RNAs associated to double-layered membranes (DMVs).23-25

Replication and transcription processes require the translation and processing of a large protein complex, encoded by the 20 kb replicase gene, into up to 16 non-structural proteins (nsp).7 During nidovirus transcription, a 3' nested set of sgRNAs of increasing size is synthesized (Fig. 1). Each sgRNA contains a 5' leader segment connected to a body sequence that is identical to the 3' end of the genome. CoV transcription requires discontinuous RNA synthesis that involves highly specific long distance interactions promoting a template switch leading to a high frequency recombination.26,27 This process is guided by transcription regulating sequences located at the 3' end of the leader (TRS-L) and preceding the body (TRS-B) of each gene. Accumulating evidence supports a model in which the discontinuous step in sgRNA synthesis occurs during minus-strand RNA synthesis.5,27-31

CoV genome 5' and 3' ends and internal structural motifs involved in virus genome replication, and their interaction with viral and cell proteins are being summarized in this review, in comparison with similar long-distance and local RNA-RNA and RNA-protein interactions necessary for the replication of other RNA virus genomes. This review can be complemented with previously published ones.4,6,32-34

Due to the limited space, we have only reported selected representative examples of the different types of interactions described in plus-strand RNA viruses.
support viral replication. The lower portion of this stem loop has an optimized lability that probably mediates a long-range interaction between the 5’ and 3’ ends of the genome critical for sgRNA synthesis, but not required for genomic minus-strand synthesis.60 SL-Ib is the most conserved structure in the CoV 5’ UTR. It is composed of a 5 bp stem and a highly conserved loop sequence that adopts a YNMG-type tetra-loop conformation, which has an important role in MHV sgRNA synthesis.60,62

The predicted SL-II in the BCoV (named SL-III in other studies)58,59,62 which includes the leader core sequence (CS) into the terminal loop, has a low free energy and is poorly conserved among CoVs.54,62 In fact, SL-II seems non-structured in other CoVs.59,62 Downstream to the leader CS is located SL-III, an RNA motif structurally conserved in all CoVs.59 The presence of a large number of co-variations supports the existence of this stem-loop and particularly the upper half (designed SL-III in the BCoV model),35 which has been confirmed by enzyme structure-probing and is essential for replication of BCoV DI RNA.35 A bulged stem-loop (BSL) SL-IV has been predicted in all CoVs.59 This BSL contains the start codon of ORF 1a preceding the downstream arm of this stem. The existence of this higher-order structure was demonstrated in BCoV by RNase mapping and, similarly to SL-III, the structure rather than the primary sequence was important for DI RNA replication.56 In addition, both SL-III and SL-IV are targets for the binding of viral and cellular proteins, which may play a role in the function of both stem-loops.55,56

Recent studies with BCoV DI RNAs have shown that the 5’-terminal 186 nt of ORF 1a coding region were required for DI RNA replication.59 This region contained two predicted stem-loops, named SL-V and -VI (Fig. 2), with structures confirmed by RNase probing and supported by nucleotide co-variation among closely related genus β CoVs.53,57 In the same study, several cellular proteins binding both stem-loops in vitro were identified, although the functional role of these interactions has not been proven.

3’ cis-acting RNA elements. The cis-acting RNA elements of the CoV 3’ UTR have been extensively studied in the genus β CoVs MHV and BCoV, and three higher-order structures have been demonstrated by chemical and enzymatic probing and by genetic studies with DI RNAs and recombinant viruses (Fig. 2). The first structure is a BSL of around 68 nt, beginning immediately downstream of the N gene stop codon, which is essential for MHV viral RNA replication.63,64 Adjacent to the BSL there is a hairpin-type pseudoknot (PK) of around 54 nt that overlaps with the BSL in 5 nt and is also required for viral RNA replication.22,65 Despite considerable primary sequence divergence, overlapping BSL and PK sequences are conserved in all genus β CoVs and are functionally equivalent.55,56,66,67 However, genus α CoVs contain a highly conserved PK but not a detectable BSL,65 and CoVs from genus γ have only a conserved and functionally essential BSL.39 These data indicate that CoVs from genera α and γ must have developed alternative mechanisms to those controlling genus β CoV replication.3’-Downstream of the PK there is a hypervariable region (HVR) that is highly divergent both in primary sequence and secondary structure even among closely related CoVs like MHV and BCoV. However, the HVR contains an octanucleotide sequence, 5’-GGA AGA GG-3’, that is universally conserved among CoVs. In the case of MHV, the HVR consists in a complex multiple stem-loop structure expanding the last 160 nt of the genome (Fig. 2).68 Mutational analysis of MHV HVR indicated that this stem-loop was required for DI RNA replication.68

Figure 2. Cis-acting RNA elements at the 5’- and 3’- genome ends of genus β CoVs. The higher-order RNA structures indicated in the diagram are mainly based on studies done with BCoV and MHV. SL, stem-loop. The core sequence within the leader TRS is shown as a black box on the top of SL-II. SL-1a and SL-1b correspond to SL-I and SL-II described in MHV.60,62 BSL, bulged stem-loop; PK, pseudoknot; S1 and L1, stem 1 and loop 1 of the pseudoknot; HVR, hypervariable region; Oct, 5’-GGA AGA GG-3’ conserved octanucleotide.
Short-Distance RNA-RNA Interactions in CoV RNA Synthesis

The cis-acting elements in viral RNA genomes have been conventionally viewed as static structures. However, recent research has changed this point of view providing evidence for cooperative activity involving short and long-distance RNA-RNA interactions, as well as conformational changes of these structures to control viral replication and transcription.\(^{73,74}\)

However, further genetic analysis in the context of the full-length genome showed that the deletion of the entire HVR, including the octanucleotide, had only a modest effect on genomic RNA replication in cell culture, although an effect in pathogenesis was observed.\(^{69}\) This discrepancy could be explained by the competitive nature of DI replication that could increase the effect of mutations that only have a moderate impact in the context of the intact genome. Finally, the 3' poly(A) tail acts as a cis-replication signal for both MHV and BCoV DI RNAs.\(^{70}\) In this study, a correlation between DI RNA replication and the ability to bind poly(A)-binding protein (PABP) is reported.

In other plus-strand RNA viruses, similarly to CoVs, evidence for higher-order structural elements in the 5' and 3' UTRs, or even in coding sequences far away from genome ends have been reported. These structural elements are involved in RNA replication, transcription, translation and encapsidation and have been recently reviewed in reference \(^{32}\).

A new kind of RNA secondary and higher order structures, named genome-scale ordered structures (GORS), has also been identified in the coding sequences of picornavirus, calicivirus and flavivirus genomes. These RNA structures were not associated with translation or replication, but were correlated with the ability of these viruses to persist in their natural host.\(^{71,72}\)

Short-distance RNA-RNA interactions and RNA conformational changes have been reported at the 3' UTR of MHV. In addition to the PK described above, a predicted interaction between PK loop 1 (L1) and the 3' end of the genome has been reported (Fig. 3).\(^{75}\) In the same study, evidence of an interaction of the replicase products nsp8 and nsp9 with the PK is presented. With this information, a model for the initiation of CoV minus-strand RNA synthesis has been proposed. According to this model, in the initial structure of the 3' UTR, the BSL is folded and the 3' end of the genome is annealed to the base of PK loop \((L1)\) (Fig. 3A). The formed stem-loop binds to a protein complex that includes nsp8 and nsp9 causing a conformational shift that releases the 3' end of the genome and disrupts the lower stem of the BSL, leading to the formation of the PK (Fig. 3B). The new conformation is probably recognized by the viral RdRp and associated factors, promoting minus-strand RNA synthesis. The two postulated structures cannot simultaneously fold up, leading to the proposal that the BSL and the PK are components of a molecular switch that regulates different steps of RNA synthesis.\(^{22,63}\)

A molecular switch has also been described in the equine arteritis virus, the arterivirus prototype.\(^{76,77}\) A PK interaction between the arterivirus loop SL-5, located at the 3' UTR and the hairpin of SL-4, located just upstream of SL-5, at the 3' terminus of the N gene, are essential for viral RNA synthesis. Furthermore, it has been proposed that the formation of this PK may constitute a molecular switch that could regulate the specificity and timing of viral RNA synthesis. Similar PK interactions near the 3' end have been predicted for all known arteriviruses.\(^{77}\) In the porcine reproductive and respiratory syndrome virus (PRRSV), an additional kissing loop between SL-4 and an upstream hairpin located within the N gene is required for replication.\(^{79}\) However, a similar kissing loop interaction has not been predicted in other arteriviruses. Similar short-distance RNA-RNA interactions and
RNA conformational changes required for RNA synthesis have been observed in other RNA virus families as indicated above.

**Long-Distance RNA-RNA Interactions Involved in CoV RNA Synthesis**

Long-distance RNA-RNA interactions mediate CoV RNA synthesis. CoV discontinuous transcription process implies a premature termination during the synthesis of the minus stranded RNAs and a template switch of the nascent minus RNA to the leader region. This switch requires long distance RNA-RNA interactions, probably assisted by RNA-protein complexes that would bring in close proximity the 5' end TRS-L and the TRS-B located 5' upstream of each gene (Fig. 4). These complexes, formed previously to template switch, might contribute to the stop of the minus RNA synthesis at TRS-B sequences.3,5,31

After this template switch takes place, a copy of the leader sequence is added to the nascent RNA by a discontinuous synthesis step.29,79,80 The requirement of base pairing between the newly synthesized minus RNA and the TRS-L to facilitate transcription has been described in arterviruses27,30 and CoVs.25,27,33,81,82 The TRSs include a conserved CS and 5' and 3' variable flanking sequences. The template switch takes place after copying the CS sequence. A three RNA strand intermediate could be formed before template switching takes place, including the RNA template (TRS-B), the nascent negative stranded RNA (cTRS-B) and the leader region (TRS-L). The G value associated to the formation of TRS-L and cTRS-B duplex is above a minimum threshold, the template switch may take place.3,31

CoV transcription resembles a similarity assisted copy-choice RNA recombination mechanism,26,27 being the TRS-L the acceptor RNA and cTRS-B the donor RNA. This recombination mechanism requires sequence homology between the donor and acceptor RNAs and a hairpin structure present in the acceptor RNA.83-85 Recently, it has been reported that a template switch may also take place within CoV sgRNAs in the absence of the standard TRS-L structure, an innovative result that may require additional assessment.86 In torovirus, the discontinuous synthesis is also required for sgRNA2 synthesis. In this case, sequence similarity and the presence of a hairpin structure are needed.87

During TGEV transcription, a good correlation has been observed between the sgRNA levels and the G values of the duplexes between TRS-L and each cTRS-B, with only one exception in the synthesis of N gene sgRNA. This sgRNA was the most abundant, despite its lower G value.88 This exception led to the discovery in our laboratory of a cis-acting element upstream of N gene TRS-B that specifically regulated N gene transcription by a long-distance RNA-RNA interaction (Fig. 5).88 The complementarity between the proximal element (pE) located 6 nt upstream of CS-N and the distal element (dE), located 449 nt upstream of TRS-N core sequence (CS-N), was essential to increase the level of sgRNA-N. Interestingly, the G value of this interaction was more relevant than the pE and dE primary sequences. This novel transcription regulatory mechanism is present in the CoV genus α, in which the N gene is not the most 3' one at the end of viral genome. This RNA-RNA interaction most likely promotes an increase in the transcription complex stop at the TRS-B located preceding gene N, during the synthesis of the minus-strand RNA (Fig. 5). This interruption would specifically enhance the frequency of the template switch to the leader during sgRNA N synthesis. This observation is reminiscent of that previously reported by White's group for transcription in tombusviruses by a premature termination mechanism.33,74,89 A similar long-distance RNA-RNA interaction between two genomic RNAs promoting transcription was previously observed in red clover necrotic mosaic virus (RCNMV).90 In flock house virus (FHV) RNA production is also regulated by an intramolecular RNA-RNA interaction.91

Long-distance RNA-RNA interactions between RNA sequence motifs mapping at distal positions of the genome (800 nt or more) that regulate replication and transcription, have also been described in other RNA viruses. In fact, potato virus X replication requires the involvement of RNA motifs, such as a hexa-nucleotide at the 3' end of the genome and a central conserved sequence, into different long distance interactions.92,93 In

---

*Figure 4.* Postulated template switch during CoV transcription. Nascent minus RNA strand (gray) is synthesized by the RdRp using genomic RNA (black) as a template. After template switch (indicated by the dashed arrow), nascent minus RNA strand hybridizes with the TRS-L, to resume synthesis of the subgenomic RNA. Different viral and cellular proteins may be involved in this process. N protein is required for an efficient transcription, and may help template switch by acting as an RNA chaperone.
addition, plus-strand genomic and sgRNA synthesis also require interactions between an octa-nucleotide located at the genome 5' end, and the central conserved sequences indicated above. In hepatitis C virus (HCV), genome replication involves RNA-RNA interactions between a cis-acting element located within the coding sequences and other structural motifs inserted up and downstream to this element.

Genome cyclization is the most distal intramolecular interaction possible within two domains in a viral genome. This type of contact has been described in several RNA virus systems, including CoVs, as a prior step to genome replication and transcription. Previous studies on MHV DI RNA replication, indicate that the last 55 nt of the 3' end plus the poly(A) tail are sufficient to promote the synthesis of the negative-strand, while both the 5' and 3' ends of the genome are required for plus-strand RNA synthesis.

This observation led us to postulate an interaction between the 5' and 3' ends of the genome for CoV replication. A direct RNA-RNA interaction between genome ends has been predicted for MHV, TGEV and severe acute respiratory syndrome CoV (SARS-CoV) using computer programs designed to analyze full-length genomes. Genetic evidence supporting a 5'-3' end cross-talk that stimulates sgRNA synthesis came from a study on MHV, in which it was suggested that the base of stem loop SL-1a, located at the 5' UTR (Fig. 2), mediates a long range interaction with the 3' end. In addition, experimental evidence reported for MHV and TGEV (Almazan F, Galan C and Enjuanes L, unpublished results) suggests that the interaction between 5' and 3' ends may also be mediated by RNA-protein interactions as described below.

Genome cyclization mediated by RNA-RNA, RNA-protein and protein-protein interactions has also been described in other positive stranded RNA viruses. In poliovirus genome cyclization the PABP binds to the 3' poly(A) tail and interacts with the protease-polymerase precursor (3CD) bound to the 5' cloverleaf RNA. This circularized ribonucleo-protein is required to initiate the synthesis of the negative stranded RNA. In contrast, the binding of the cellular poly(rC) binding protein (PCBP) to the 5' cloverleaf RNA promotes the translation. Similarly, in pestiviruses such as bovine viral diarrhea virus (BVDV), genome circularization is also mediated by three proteins of the NFR group. This circularization promotes the switch from translation to replication. Foot and mouth disease virus (FMDV) uses specific higher-order RNA structures located at the termini of viral genome to regulate the translation and viral RNA synthesis. The replication process requires direct 5' to 3' end RNA-RNA interactions mediated by the S region preceding the IRES, the IRES 5' end and the 3' end of the genome. In addition, RNA-protein interactions also play an important role connecting both ends.

In flaviviruses, genome cyclization is mediated by direct long-distance RNA-RNA interactions between complementary sequence motifs present at both ends of the genome. Complementarity disruption impairs RNA synthesis, providing evidence for the requirement of circularization during the replication process.

The long-distance RNA-RNA interactions described until now probably represent the tip of the iceberg of a complex network involving complete RNA virus genomes. These long-distance interactions probably lead to a tertiary structure of the full-length virus genome. The best paradigm of this theory is probably reflected in the pioneering studies on tomato bushy stunt virus (TBSV), for which a global RNA structure has been proposed to regulate replication, transcription and translation.

**RNA-Protein Interactions Involved in CoV RNA Synthesis**

The identification of viral proteins participating in coronivirus RNA synthesis was traditionally based on the analysis of temperature-sensitive mutant viruses. This approach led to the identification of several replicase-encoded nsps as key factors in replication and transcription. Cell factors involved in RNA synthesis have also been identified by studying their binding to selected regions of the genome, or to replicate proteins. Recently, high-throughput assays have been used for the identification of viral and cellular factors affecting CoV replication. These technologies include genome-wide two-hybrid screenings and proteomic analyses that identify interactions between replicase components, probably relevant in CoV replication. Alternatively, strategies similar to the ones applied to other plus-stranded RNA virus genomes may also be useful to study CoV replication. This is the case of high-throughput functional
assays using host cell mutants or small interfering RNAs (siRNAs) affecting viral RNA amplification and expression applied to HCV\textsuperscript{115,116} and human immunodeficiency virus (HIV).\textsuperscript{115,116}

One of the largest limitations in the identification of factors affecting replication or transcription is to distinguish between direct and indirect effects. In addition, factors affecting RNA synthesis frequently also affect translation.\textsuperscript{117} Therefore, it becomes essential to study the role of factors in defined functional assays, such as in vitro replication or transcription systems. The purification of in vitro active CoV replication-transcription complexes (RTC) has been described in reference \textsuperscript{118}. This system could be used to analyze the role of viral and cell factors in RNA synthesis. Although both viral and cellular proteins are most likely part of the purified active complexes, the precise components that reproduce the activity still need to be defined. An alternative is to develop in vitro systems based on defined components. The key difficulty for the development of these CoV in vitro systems is the requirement of a large number of components.

**Viral proteins involved in CoV RNA synthesis.** Among the viral proteins required in CoV RTC, it has been proposed that the RdRp (nsp12), the helicase (Hel, nsp13) and N protein are essential.\textsuperscript{3} In addition, other viral proteins probably contribute to the formation of this complex and the regulation of its activity. In fact, several replicase proteins such as nsp3, nsp4 and nsp6 have mainly scaffolding and membrane rearrangement functions.\textsuperscript{7,119,121} In addition, nsp3 and nsp5 act as proteases for the correct processing of replicase polyproteins.\textsuperscript{7}

Arterivirus nsp1 binds RNA and has been involved in transcription,\textsuperscript{122,123} probably by modulating the relative abundance of subgenomic and genomic RNAs.\textsuperscript{124} In CoVs, the binding of nsp1 to a BCoV cis-acting replication signal has been reported in reference 57. Nevertheless, nsp1 protein has not been involved in RNA synthesis. In line with this observation, the replicase of IBV, a member of CoV genus γ, does not encode a nsp1,\textsuperscript{7,127} suggesting that CoV nsp1 is probably dispensable for CoV RNA synthesis.

Nsp2 is recruited to the RTC.\textsuperscript{126} Although nsp2 is dispensable for genus β CoV replication, a 50% reduction in viral RNA synthesis is observed when this protein is deleted.\textsuperscript{127} Therefore the role of nsp2 in CoV RNA synthesis still remains to be fully established.

Nsp3 is a multifunctional protein that binds RNA.\textsuperscript{111,128} The role of nsp3 in CoV RNA synthesis remains unclear, as most of its domains are accessory for RTC activity, although they may play a role in pathogenesis.\textsuperscript{129,130} The interaction between nsp3 and N protein has been recently described, and it has been proposed that nsp3 is involved in the location of CoV RNA genome within the RTC complex at early stages of RNA synthesis.\textsuperscript{131} SARS-CoV E protein binds the N-terminal acidic domain of nsp3, but its specific role in CoV replication remains to be established.\textsuperscript{132}

It has been proposed that nsp5 to 10 form a functional cassette involved in virus RNA synthesis.\textsuperscript{133} Nsp7 and nsp8 form a hexadecameric complex that has structural properties compatible with that of a processivity factor for RdRp.\textsuperscript{134} Nsp8 has RdRp activity and probably produces the primers required for nsp12 mediated RNA synthesis.\textsuperscript{135,136} Nsp9 is a single-stranded RNA binding protein that may stabilize viral RNAs during RNA synthesis and processing.\textsuperscript{137,138} Nsp9 forms a homodimer and its oligomerization is required for viral replication.\textsuperscript{139} MHV nsp8 and nsp9 interact with RdRp,\textsuperscript{140} and genetic evidence suggest that they also interact with the RNA PK at the 3’ end of the viral genome.\textsuperscript{73} Nsp10 contains two Zn-finger motifs, binds nucleic acids non-specifically,\textsuperscript{141,142} and has been implicated in minus-strand RNA synthesis.\textsuperscript{143}

Nsp14, 15 and 16 encode RNA-processing activities that are unique to the RNA virus world, as exonuclease (ExoN), endoribonuclease (EndoU) and methyltransferase (MT) activities, respectively.\textsuperscript{7} All these activities are required for efficient CoV RNA synthesis.\textsuperscript{133,144,146} It has been proposed that the presence of these activities, especially nsp14, allows the maintenance of functional CoV genomes of large size by acting as a proofreading system.\textsuperscript{125,147,148} CoV nsp14 also encodes a MT activity at its C-terminus, required for RNA synthesis.\textsuperscript{149} It is possible that MT activities of nsp14 and nsp16 act coordinately for viral mRNA capping.\textsuperscript{150}

CoV N protein forms a ribonucleoprotein complex with genomic RNA. In addition to its structural role, this protein has also a prominent role in RNA synthesis. N protein colocalizes with the RTC.\textsuperscript{151,152} Initial observations suggested that N protein plays a role in RNA synthesis\textsuperscript{153-158} and additional studies confirmed that this was the case.\textsuperscript{159-161} These results are in agreement with the significant increase of infectious virus rescue from different CoV cDNA clones by providing N protein in trans.\textsuperscript{47,160,162} The N-terminal domain of CoV N protein binds specifically to TRS-L sequences and facilitates unwinding of TRS duplexes.\textsuperscript{163} CoV N protein also has RNA chaperone activity\textsuperscript{164} that enhances template switch in vitro.\textsuperscript{165} As a consequence of these activities, it has been postulated that N protein may facilitate the CoV template switch step during CoV transcription.\textsuperscript{163,165} In fact, it has been recently reported that N protein is required for an efficient transcription but is not essential for CoV RNA replication.\textsuperscript{165}

**Cellular proteins involved in CoV RNA synthesis.** Cellular proteins associated with CoV RNA synthesis are reviewed below according to their binding to 5’ UTR, internal TRS-BS, 3’ UTR or to viral replica components. Protein binding to 5’ or 3’ UTR might be involved in viral replication and also in transcription, translation and viral RNA stability. Binding to internal TRS-BSs might be related to transcription or post-transcriptional steps.

**Cellular proteins binding CoV 5’ UTR.** In MHV, belonging to CoV genus β, poly pyrimidine-tract binding (PTB) protein, also known as hnRNP I, binds pentanucleotide repeats of leader TRS, located in the 5’ UTR (nt 56–112). Deletion of these leader repeats in DI RNAs significantly inhibited RNA transcription, suggesting that PTB might be involved in transcription regulation.\textsuperscript{166} Functional studies showed delayed virus production and reduction of viral RNA synthesis in MHV infected cells overexpressing PTB. This protein interacts with viral N protein both in vitro and in vivo, suggesting an additional contribution to the ribonucleoprotein (RNP) complex formation.\textsuperscript{167} PTB also binds to the TRS-L sequences at the 5’ end of TGEV genome\textsuperscript{106} (Sola I and Enjuanes L, unpublished results). hnRNP A1 specifically binds to MHV RNA consensus sequences present in viral genome negative RNAs cTRS-L and cTRS-B, complementary to TRS-L and TRS-B RNA motifs, respectively.\textsuperscript{168} Another
member of hnRNP family, SYNRIP or hnRNP Q, binds to MV 5' UTR or to its complementary sequence 5' UTR. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was also identified in association with TGEV TRS-L (Table 1). In addition, the knockdown of the GAPDH gene produced a positive effect on CoV-replication both in a TGEV replicon and full-length infectious virus systems, suggesting a potential role in the balance of cellular factors that positively influence TGEV RNA synthesis. This hypothesis is supported by the formation of a complex including GAPDH, the bifunctional aminoacyl-tRNA synthetase (EPRS), hnRNP Q and the ribosomal protein L13a, in an inflammatory response.

**Cellular proteins binding CoV TRS-B.** hnRNP A1 binds to the complementary RNA strand of MV TRS-L (cTRS-L) and of several TRS-Bs (cTRS-Bs). Mutagenesis of TRS-B sequences in a DI RNA reporter system resulted in reduced transcriptional activities that correlated with relative binding affinities of the cTRS-B sequences to hnRNPs A1, suggesting that this protein may be a transcription factor in CoV RNA synthesis. Nevertheless, it cannot be ruled out that the observed effect was due to changes in the extent of the complementarity between the newly synthesized cTRS-B and TRS-L. In an in vitro reconstitution assay, hnRNP A1 mediates the formation of an RNP complex with the minus strand cTRS-L and cTRS-B. Furthermore, hnRNP A1 interacted both in vitro and in vivo with MV N protein, that forms part of the RTC.

We have used TGEV CoV as a model to identify candidate cellular proteins participating in CoV transcription. Leader and body TRSs (TRS-L and TRS-B, respectively) with either positive or negative polarity, as well as RNA duplexes mimicking TRS-L/cTRS-B hybrids formed during transcription template switch, were used in RNA affinity chromatography assays. Mass spectrometry analysis led to the identification of around 40 cellular proteins interacting with TGEV TRSs. A major proportion of these proteins was associated with cellular RNA metabolism, including pre-mRNA processing (45%), transcription (8%) and translation (20%) (Fig. 6). Members of hnRNP family or the DEAD box family of cellular RNA helicases (DDX and DHX) were extensively represented. hnRNP I was found in association with the plus-stranded RNA of TRS-L and several TRS-B and, therefore, PTB may participate in the regulation of sgRNA synthesis (Sola I and Enjuanes L, unpublished results). Other identified proteins were related to cytoskeleton (16.7%) or cell metabolism processes (3.3%) (Fig. 6 and Table 1). Some of the identified proteins have already been involved in the life cycles of other plus-stranded RNA viruses and may provide novel insights into the protein machinery utilized by CoVs in RNA synthesis.

**Cellular proteins binding CoV 3' UTR.** UV-crosslinking experiments showed binding of hnRNP A1 to MV genome 3' UTR sites located at nt 90–170 (mapping within the HVR domain) and nt 260–350 numbered from the 3' end of the genome (partially overlapping with the BSL). Overexpression of hnRNP A1 or of a dominant-negative mutant derived from this protein, accelerated or delayed, respectively, viral RNA synthesis, suggesting that hnRNP A1 participates in the formation of the RTC. hnRNP A1 is translocated from the nucleus to the cytoplasm of infected cells, probably as a result of specific interaction with MV RNA. Other member of the hnRNP A2/B1 multigene family can substitute hnRNP A1 in MV replication. hnRNP A1 binding sites are complementary to those on the negative-strand RNA that bind PTB. Partial deletion of hnRNP A1 and PTB binding sites abolished sgRNA synthesis by MV DI RNAs. hnRNP A1 and PTB also bind to the complementary strands at the 5' end of MV RNA and mediate the in vitro formation of a RNP complex involving the 5' and 3' end fragments of MV RNA.

Other cellular proteins bind the CoV 3' end genome and also affect virus replication. Thus, PABP binds to BCoV and TGEV (Marquez-Jurado S, Enjuanes L and Almazan F, unpublished results) poly(A). Furthermore, a positive effect of PABP in TGEV RNA synthesis has been observed in reference 106. Other 3'-interacting host factors such as the hnRNP Q and the EPRS also positively contribute to TGEV RNA synthesis, as shown in functional analysis with siRNAs using both a TGEV derived replicon and the infectious virus itself. Transcriptional coactivator p100 (SNOD1) specifically binds the 3' end of the TGEV and also the TRS-L and TRS-B RNA motifs (Table 1) (Sola I and Enjuanes L, unpublished results). P100 protein interacts with the nspl of EAV, specifically involved in mRNA synthesis. However, no functional studies have shown the role of p100 in arterivirus transcription. Mitochondrial proteins aconitase, heat shock protein (HSP)40, HSP60 and HSP70 bind to MV 3'-most 42 nt of RNA genome, although their requirement for CoV replication was not directly addressed. Specific mutations in this 3' UTR sequence led to a virus with a defect in sgRNA synthesis, supporting the relevance of this 3' end sequence domain in virus transcription.

**Cellular proteins binding CoV replicase subunits.** An alternative approach to identify host factors that participate in CoV RNA synthesis relies on their specific binding to viral replicase subunits. Several members of the cellular DEAD box family of helicases such as DDX5 and DDX1 have been associated with CoV RNA synthesis. These helicases have been involved in several biological processes related to RNA transcriptional regulation, pre-mRNA processing, RNA degradation, RNA export, ribosome assembly and translation. The specific interaction of cellular DDX5 protein with the SARS-CoV helicase protein (nsp13) has been shown using yeast and mammalian

---

**Table 1.** Illustrative examples of cellular proteins binding CoV TRSs identified by RNA-affinity chromatography and mass spectrometry

<table>
<thead>
<tr>
<th>Biological process</th>
<th>Protein name</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA splicing</td>
<td>hnRNP A1, hnRNP A2/B1, hnRNP A3, hnRNP I, hnRNP Q</td>
</tr>
<tr>
<td>Transcriptional regulation</td>
<td>ILF2, ILF3, SNDI</td>
</tr>
<tr>
<td>mRNA binding and processing</td>
<td>Nucleolin, DDX1, DHX15</td>
</tr>
<tr>
<td>Translational regulation</td>
<td>elf1α, elf3S10</td>
</tr>
<tr>
<td>Cytoskeleton-movement</td>
<td>Tubulin, annexin A2, moesin</td>
</tr>
<tr>
<td>Metabolic processes</td>
<td>GAPDH</td>
</tr>
</tbody>
</table>

*aGene ontology classification of biological processes.
two-hybrid assays and co-immunoprecipitation. Silencing of DDX5 expression by siRNAs significantly reduced viral RNA replication and virus titers.107 DDX1 interacted with IBV and SARS-CoV exoribonuclease (np14), as demonstrated by yeast two-hybrid and coimmunoprecipitation assays. The effect of DDX1 expression knockdown in CoV RNA replication and transcription indicated that DDX1 might be a cofactor essential for these activities.108 Interestingly, RNA helicase DDX1 has also been identified in association with TGEV TRSs (Sola I and Enjuanes L, unpublished results) and might perform common functions in CoV replication.

Perspectives

Recent advances in understanding the molecular mechanisms of replication and transcription in positive-stranded RNA viruses indicate that regulatory elements in viral genomes are dynamic complexes involving short and long-distance RNA-RNA interactions and also RNA-protein complexes including viral and cellular proteins. Transition between two structural conformations may regulate the switch between essential activities within virus cycle. Future challenges in this field rely on the development of novel methodologies to study RNA structure formation in vivo as a consequence of RNA-RNA and RNA-protein interactions in the cellular environment. In addition, in vitro reconstitution of replication-transcription complexes with defined composition will be critical to analyze the contribution of viral and cell factors to RNA synthesis.

Acknowledgements

This work was supported by grants from Ministry of Science and Innovation of Spain (BIO2007-60978 and BIO2010-16705), US National Institutes of Health (ARRA-W000151845) and the European Community Seventh Framework Programme (FP7/2007-2013) under the projects “EMPERIE” (EC Grant Agreement number 223498) and “PoRRSCon” (EC Grant Agreement number 245141). I.S. and S.Z. received contracts from the indicated EU projects, respectively. P.A.M. received a fellowship from the Ministry of Science and Innovation of Spain.


Lai MMC, Cav. 2002; 76:1293-308.


Pardesy B, Towner JS, Bly LB, Ehrenfeld E, Sements BL. Poly (rC) binding protein 2 forms a ternary complex with the 5' poly(A) tail of poliovirus RNA and the viral 3CD proteinase. RNA 1997; 3:1124-34.


Nedialkova DD, Gorbalenya AE, Snijder EJ. Arterivirus Nlp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs. PLoS Pathog 2010; 6:1000772.


