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

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

Subgenomic messenger RNAs: Mastering regulation of (+)-strand RNA virus life cycle

Joanna Sztuba-Solińska a, Victor Stollar b, Jozef J. Bujarski a,c,*

a Plant Molecular Biology Center and the Department of Biological Sciences, Northern Illinois University, DeKalb, IL 60115, USA
b Department of Molecular Genetics, Microbiology, and Immunology, Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA
c Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

A R T I C L E   I N F O

Article history:
Received 1 November 2010
Returned to author for revision 14 December 2010
Accepted 4 February 2011
Available online 5 March 2011

Keywords:
Subgenomic RNA (SG RNA)
Internal initiation
Premature termination
Discontinuous transcription
Riboregulators
Copy-choice RNA recombination

A B S T R A C T

Many (+)-strand RNA viruses use subgenomic (SG) RNAs as messengers for protein expression, or to regulate their viral life cycle. Three different mechanisms have been described for the synthesis of SG RNAs. The first mechanism involves internal initiation on a (−)-strand RNA template and requires an internal SGP promoter. The second mechanism makes a prematurely terminated (−)-strand RNA which is used as template to make the SG RNA. The third mechanism uses discontinuous RNA synthesis while making the (−)-strand RNA templates. Most SG RNAs are translated into structural proteins or proteins related to pathogenesis; however other SG RNAs regulate the transition between translation and replication, function as riboregulators of replication or translation, or support RNA–RNA recombination. In this review we discuss these functions of SG RNAs and how they influence viral replication, translation and recombination.

© 2011 Elsevier Inc. All rights reserved.

Introduction: definition and occurrence of SG RNAs

Synthesis of subgenomic (SG) messenger RNAs (mRNAs) by (+)-strand RNA viruses allows the differential expression of specific viral genes, both quantitatively and temporally. SG RNAs as considered in this review, have the following properties: (i) they are made in infected cells but do not interfere with the normal course of viral replication; (ii) the SG RNA sequences are shorter than their cognate genomic RNAs; (iii) their sequences are usually co-terminal with the 3′ genomic sequence but sometimes are co-terminal with the 5′ sequences. Yet other viruses make SG RNAs which contain a 5′ co-terminal leader joined to a 3′ co-terminal sequence; (iv) typically, whether a messenger SG RNA contains only one ORF, or multiple ORFs, with some rare exceptions (Dorokhov et al., 2006), only the 5′ ORF is translated (Pasternak et al., 2006; Yang et al., 2009). Although most SG RNAs function as messengers and are translated, other SG RNAs, generally those with 5′ co-terminal sequences, have other functions.

The production of SG RNAs was initially reported in studies of Brome mosaic virus (BMV) and was followed by the discovery of SG RNA in Tobacco mosaic virus (TMV)-infected leaves (Miller and Koev, 2000) and later in Alphao-, Carmo- and Sobemo- families (Strauss and Strauss, 1994; Rico and Hernandez, 2009; McGavin and MacFarlane, 2009). Indeed, among (+) strand RNA viruses of plants, only viruses of the Potyviridae and Comoviridae as well as the Sequiviridae family do not use this strategy (Zaccomer et al., 1995).

Our goal in this review is to describe the major mechanisms by which SG RNAs are generated and to discuss their roles in the life cycle of (+) strand RNA viruses. Since the last reviews discussing that subject (Koev and Miller, 2000; White, 2002; Miller and White, 2006), these RNAs have been shown to function not only as mRNAs, but also as riboregulators of replication and transcription; as well, SG RNAs may participate in genome rearrangements that help maintain genomic integrity and possibly the acquisition of non-self sequences.

Although replicon RNAs made in the laboratory to study the replication of Hepatitis C virus (HCV) and other Flaviviruses have also been referred to as subgenomic RNAs, since they are not seen in a normal infection, they will not be considered in this review.

Mechanisms of SG RNA synthesis and regulation

Internal initiation on the (−)-strand RNA template

Following the synthesis of the viral RNA-dependent RNA polymerase (RDRP), the (+) strand RNA is copied into a genome-length (−) strand which then serves as a template for the genomic (G) and the SG (+) strand RNAs (Fig. 1A). Thus, the (−) strand RNA contains at least two different promoters, one for the synthesis of G RNA at or near the
3′ end, and one or more internal or subgenomic promoters (SGPs). To synthesize a SG RNA, the viral RDRP recognizes and binds to the SGP and initiates transcription (Koev and Miller, 2000). This mechanism was first found with BMV, a tripartite plant virus. The (−) strand of dicistronic BMV RNA 3 serves as template for transcription of the 3′ co-terminal SG RNA due to de novo internal initiation from the SGP. The BMV SGP has been mapped to the sequence from nt −95 to +16 (Fig. 2A). Its modular composition includes an AU-rich enhancing region, a poly (U) tract, a core region, the +1 C transcription initiation site, and a downstream sequence (Wierzchoslawski et al., 2004). The enhancing region (nts −95 to −20) greatly increases the amount of SG RNA 4. The core region (nts −19 to −1) contains a stem-loop (SL) structure which is responsible for RDRP binding. Adkins and Kao (1998) have suggested an induced fit mechanism whereby the RDRP recognizes key nucleotides, after which nucleotide base-pairing occurs and the SL structure forms. Another model proposes that the SL structure binds to the RDRP complex, after which the key nucleotides stabilize the hairpin (Haasnoot et al., 2000, 2002). Regardless of the mechanism, internal initiation appears to be a multi-step process, which involves both host and viral proteins (Hertz and Huang, 1995a,b; Adkins and Kao, 1998; Haasnoot et al., 2002; Haasnoot et al., 2002).

Fig. 1. Schematic representation of different mechanisms for subgenomic RNA synthesis. (A) Internal initiation model (Miller and Koev, 2000), (B) Premature termination model showing termination either during (−) and (+)-strand synthesis (White, 2002), (C) Leader-primed transcription model, (D) Discontinuous template synthesis (Pasternak et al., 2006). Genomic and SG RNA (+)-strands are depicted as horizontal yellow boxes, the (−)-strands are depicted as green boxes, the ovals represent RdRp enzymes capable of starting/stopping at the internal initiation SGP promoters that are depicted as blue SGP boxes. The leader TRS (L-TRS) and the body TRS (B-TRS) are represented by dark green and brown boxes, respectively. Refer to the text for details.
Sivakumaran et al., 2004). Other (+)-strand RNA viruses use SGPs with a similar organization (Miller and Koev, 2000; Haasnoot et al., 2000; Morales et al., 2004; Olsthoorn et al., 2004). TMV has a single-stranded non-segmented RNA genome and represents another example of a virus that synthesizes SG RNAs via internal initiation. Three of these SG RNAs are co-terminal with the 3′ terminus of the genomic RNA; the smallest SG RNA has only one ORF, and encodes CP. The middle-sized SG RNA has two ORFs (for MP and CP), but only the 5′ ORF, that for MP, is expressed. The third and largest SG RNA, referred as the I₁, has been detected in TMV-infected tobacco tissue. It contains in addition to ORFs for MP and CP, a 5′ ORF for a 54-kDa protein with a sequence identical to that of the read-through region of the 183-kDa replicase. Interestingly, RNA structures that are able to regulate the activity of TMV promoters can be far removed from the initiation site. Culver et al. (1993) found that the location of the TMV SGP in relation to the 3′ UTR is a crucial factor in enhancing expression. Later, Szecsi et al. (1999) and Shivprasad et al. (1999) associated the positive effect of the 3′ UTR with the three pseudoknots contained therein; these were found to redistribute RDRP activity localizing most of it to the nearest SGP promoter.
Both RNA sequence and secondary structures are generally critical for the activity of an SGP (Koev et al., 1999; Haasnoot et al., 2000). In the case of the TMV promoter for the middle-sized SG RNA (Fig. 2B), computer analysis predicted two stem-loop structures (SL1 and SL2) upstream of the transcription start site. The SL1 secondary structure, rather than its sequence, was critical for promoter activity. On the

---

**Fig. 3.** Long distance RNA–RNA interactions that regulate SG RNA synthesis in RNA viruses. (A) Schematic overview of the genome organization and expression of EAV. The gray boxes in the genomic RNA indicate the positions of the leader TRS and (major) body TRSs. Below is shown the proposed base pairing interaction between the sense leader TRS (L-TRS) and the antisense body TRS (B-TRS) in the 3’ end of the nascent minus strand (van Marle et al., 1999); only the loop and the top of the stem of the predicted hairpin structure are presented. (B) Linear representation of the TBSV RNA genome showing its coding organization. The relative positions of interacting RNA elements involved in SG RNA transcription are shown above the genome and are indicated by arrows. Initiation sites for SG RNA transcription are labeled sg1 and sg2, and corresponding structures of the two SG RNAs are represented by bold arrows below the genome. Below, the long-distance RNA–RNA interactions that regulate SG RNA transcription in TBSV are shown in detail. Relevant sequences of the TBSV genome are presented with corresponding genomic coordinates. The AS1/RS1 base-pairing interaction is essential for the efficient transcription of SG RNA1 (Choi and White, 2002), while the AS2/RS2 and DE-A/CE-A base-pairing interactions promote SG RNA2 transcription (Choi et al., 2001; Lin and White, 2004); (C) Representation of genomic RNA1 of Flock house virus (FHV). Interaction between the distal subgenomic control element (DSCE) and the proximal subgenomic control element (PSCE) is required for RNA3 synthesis. The internal replication element (INTRE) and 3’ replication element (3’ RE) are required for RNA1 replication. Below, potential base pairing of the DSCE to regions proximal to the subgenomic region start site. Gray boxes mark PSCE residues that potentially base pair to the DSCE. Putative helices 1 and 2 are bracketed. Arrow, RNA3 start site at nt 3178 (Lindenbach et al., 2002); (D) Schematic of the RCNMV genome showing the relative positions of the in trans interacting RNA elements: the loop portion of a stem-loop in RNA2 (termed trans-activator or TA) and a complementary sequence in RNA1 (termed TA binding site or TABS) located just upstream from the initiation site for SG mRNA transcription (Guenther et al., 2004); (E) Representation of CLSV genome and the proposed AS2 and RS2 interaction during regulation of SG RNA2 synthesis (Xu and White, 2008); (F) Overview of the BYDV genome and the cis interaction between genomic 5’-UTR stem-loop (BCL) and the genomic 3’-BTE, and the in trans interaction between the genomic 3’ BTE and the 5’ UTR of SG RNA1 (Rakotondrafara et al., 2006). Numbers depict the nucleotide positions in the viral genomic RNA. The nucleotides participating in the long-range base-pairing are joined by lines.
other hand, removal of most of the SL2 region increased accumulation of the SG RNA four-fold (Grdzelishvili et al., 2000). Also, the three SGP promoters of Barley yellow dwarf virus (BYDV) fold into different size stem-loops downstream of their respective initiation sites (Koev and Miller, 2000). In the case of promoters of Barley stripe mosaic virus (BSMV) SG RNAs β1, β2, and γ, the conservation of both the sequences and the secondary structures appear to be important for their activity (Johnson et al., 2003). Here the substantial differences between SG RNAs β1, β2, and γ promoter sequences were postulated to explain competition for the viral RDRP, coordination of the temporal expression and abundance of the proteins, and constitutive expression of the γ b protein.

Viruses producing multiple SG RNAs often contain homologous sequences within their SGP promoters (Koev and Miller, 2000). Also, certain elements in genomic promoters for the SG strand synthesis can share similarities with elements in the internal promoters; for example, a stem-loop C (SLC) in BMV (Haasnoot et al., 2002) and the triloop hairpin (hpE) in Alfalfa mosaic virus (AMV) (Olsthoorn et al., 2004). Initiation from the 3′ end of the sense genomic BMV RNA obeys different rules than does the initiation of SG RNA (Stawicki and Kao, 1999), likely due to additional factors (Diez et al., 2000) that can adjust the properties of the RDRP (Adkins and Kao, 1998; Ranjith-Kumar et al., 2003; Haasnoot et al., 2002; Sivakumaran et al., 2004).

In the case of Sindbis virus (SV) (the prototype of the genus Alphavirus, family Togaviridae), the minimal sequence of the SG strand RNA needed for the SG activity extends from nt −98 to +14 increases promoter activity at least 6 folds. Most of this increase could be accounted for by strongly conserved sequences extending from −40 to −20, and from +6 to +14 (Wielgosz et al., 2001) (Fig. 2C).

Three classes of mutations down regulate the synthesis of the SV SG RNA: (i) promoter mutations, (ii) nsP2 protein mutations, and (iii) a single mutation in nsP3. SV makes four nonstructural proteins; nsP1 is responsible for capping and methylation of the G and SG RNAs; nsP2 has both a protease activity and an RNA helicase activity; the function of nsP3 is not known, and nsP4 is the viral RDRP. Promoter mutations which affect the synthesis of SV SG RNA have been described by Hertz and Huang (1995a, 1995b) and Lin et al. (2002). These effects are often cell-dependent, suggesting a role for cellular factors in the synthesis of SG RNA.

Mutations in nsP2 of SV and SFV, which mapped to its protease domain, decreased protease activity and thus slowed cleavage of the nonstructural polyproteins, a step necessary for the efficient synthesis of SG RNA. Also, a single insertional mutation in nsP3 lowered the level of the SV SG RNA synthesis, but the mechanism responsible is not known (LaStarza et al., 1994).

By incubating a labeled SGP sequence with an SV RNA transcriptase/replicase complex, distinct protein sites on nsP4 were found to be responsible for the recognition of the SGP and G promoters (Li and Stollar, 2004, 2007). It was shown that Arg to Ala changes at positions 331 or 332 knocked out the in vitro synthesis of SG but not G RNA (Li and Stollar, 2007). Conversely, by changing the Arg residue at 545, 546, or 547 to Ala, the synthesis of G RNA but not SG RNA was knocked out (Li et al., 2010).

Viruses in the family Caliciviridae also make SG RNAs by internal initiation on the (+) strand (Morales et al., 2004). In at least one case, these SG RNAs are packaged into viral particles (Neill, 2002).

### Premature termination

In a second mechanism for making a SG mRNA, the RDRP complex, instead of copying the (+) strand RNA genome into a full length (−) strand RNA, terminates “prematurely” at a specific STOP signal, and synthesizes a shortened (−) strand of SG RNA, which then serves as a template for the synthesis of a SG (+) strand RNA (Fig. 1B). These SG (+) strand RNAs, like those made by internal initiation, have sequences which are co-terminal with the 3′ sequence of the viral genome. Premature termination (PT) can, however, also occur during synthesis of (+) strand RNA (White, 2002), thus generating 5′ co-terminal SG RNAs (Fig. 1B) (Wierzchoslawski et al., 2006). Premature termination (PT) supporting formation of SG RNAs is seen with various RNA viruses including Toroviruses (van Vliet et al., 2002), Roniviruses (Cowley et al., 2002), Betanodaviruses (Iwamoto et al., 2002).
**Flock house virus (FHV)**, the prototype virus of *Nodaviridae*, uses PT to make a SG RNA. FHV has a bipartite (+) strand RNA genome but both RNAs are packaged into a single viral particle. RNA 1 encodes a replicase protein A, while RNA 2 encodes protein α, the precursor of the capsid proteins. FHV infection also gives rise to a 387 nt SG RNA (RNA 3), the sequence of which is 3′-co-terminal with that of RNA 1. RNA 3 has two overlapping ORFs, B1 and B2. The B1 ORF encodes the C-terminal 102 amino acids of the A protein. No function has been associated with this polypeptide. The B2 protein is a suppressor of RNA interference, both in plants and cultured insect cells. RNAs 1, 2, and 3 all have a 5′ cap, but none are polyadenylated.

In studies of FHV RNA replication in yeast (Lindenbach et al., 2002), two cis-acting elements were found in RNA 1 (Fig. 3C): (i) a proximal subgenomic control element (PSCE) extends from nt 2282 to nt 2777 (the start site for the synthesis of SG RNA 3 is at nt 2721), and (ii) a distal subgenomic control element (DSCE) from nt 1229 to nt 1239. Base pairing between the nt sequence 1229 to 1239 in the DSCE and both a 6 nt sequence and a 4 nt sequence (almost 200 nt apart) in the PSCE is required for the synthesis of RNA 3. For the most efficient synthesis of RNA 3, the PSCE sequence between nt 2302 and 2777 was needed; however some synthesis was seen with the 5′ limit of the PSCE at nt 2518. Thus, the region from nt 2302 to 2518 serves as an enhancer. Disruption of PSCE/DSCE base pairing increased the amount of RNA 1 made, but then neither (+) nor 5′-strand RNA 3 was made. Apparently, the long distance interaction between DSCE and PSCE gives rise to a secondary or tertiary structure that results in premature termination when 5′-strand RNA 1 is made. In addition, work by Eckel et al. (2003), strongly suggests that the (+) and 5′-strands of RNA 3 participate in a full-fledged RNA 3 replication even in the absence of RNA 1. Furthermore, as with the synthesis of RNA 1 and RNA 2, the synthesis of (+) strand FHV RNA 3 great exceeds that of 5′-strand RNA. Finally, in this system, RNA 3 is required for the synthesis of RNA 2, and once RNA 2 is made, it suppresses the synthesis of RNA 3.

**Tomato bushy stunt virus (TBSV)** uses a similar strategy to make SG RNAs (Fig. 3B) (White and Nagy, 2004; Zhang et al., 2004). While making 5′-strand RNA, the RDRP complex terminates at the promoter regions for (+) sense SG RNAs, generating the subviral-length (−) sense 3′-SG RNAs (White, 2002), which then act as templates to amplify the (+) sense 3′ SG RNA. Substitution of the transcription-initiating nucleotides for SG RNA 1 or 2 inhibits the synthesis of the SG RNA requires a bipartite genome with SG RNA being synthesized from RNA1. The synthesis of the SG RNA requires a trans-activation by an RNA hairpin structure located in RNA2 (see below) (Guenther et al., 2004). Specifically, this activation involves the loop portion of an SL structure in RNA2 (trans-activator or TA) which base-pairs with a complementary sequence in RNA1 (TA binding site or TABS) that is located upstream from the SG RNA transcription initiation site (Fig. 3D) (Guenther et al., 2004). It has been suggested that protein factors may bind to and stabilize the bimolecular RNA–RNA contact. Interestingly, the disruption of the RNA–RNA interaction does not affect replication of RCNMV genomic components (Tatsuta et al., 2005). Thus, intermolecular communication secures the switching between replication and transcription, resolving the problem of their mutual interference.

Also, TCV has recently been found to use a premature termination mechanism for making SG RNA2 (bearing CP ORF) (Wu et al., 2010). Analyses revealed that 5′-strand SG RNA2 accumulation can be uncoupled from that of its (+) strand counterpart. An extended SL RNA structure positioned 5′ to the initiation site for SG RNA2 was found to mediate PT mechanism by functioning in the (+) strand of the viral genome. As with other viruses that use a PT mechanism, the high degree of identity between the SGP promoter for (+) strand genome and that for SG RNA2 transcription, support the idea that TCV uses a PT mechanism for SG RNA2 transcription.

As noted above, in addition to making the 3′ co-terminal SG RNA, RNA 4, BMV also makes a 5′-co-terminal SG RNA, RNA 3a. In this case, the RDRP likely pauses during (+) strand synthesis at the internal oligo U tract resulting in formation of the 5′-co-terminal 3′polyadenylated (+) SG RNA3a (Fig. 1A) (Wierczloslawski et al., 2006). The binding of another RDRP molecule to the near-by SGP promoter core hairpin which initiates the SG RNA4 transcription might pose an additional obstacle to the progression of the RDRP during (+) strand synthesis (Wierczloslawski et al., 2006; Sztuba-Solińska and Bujarski, 2008).

Other examples of elements in SGPs that act as road blocks for the progressing RDRP have been reported for *Citrus leaf blotch virus* (CLBV) (Vives et al., 2002), *Grapevine viitivirus A* (GAV) (Galiakparov et al., 2003), *Citrus tatter leaf virus* (CTLV) (Tatieni et al., 2009), and *Citrus tristeza virus* (CTV) (Che et al., 2001). CTLV contains two overlapping 5′ ORFs, the expression of which requires the production of two 5′ (+) strand co-terminal SG RNAs, which result from premature termination. The 3′-ends of both SG RNAs terminate at two SL structures: SL1 and SL2, with the transcription initiation site for the 3′ SG RNAs located in the loop region of SL2 (Tatieni et al., 2009). Likewise, CTV terminates one of its 5′ SG RNAs (referred as low-molecular-weight tristeza LMT1) at the two SL structures located upstream of the internal initiation site for the 3′ SG RNA (Gowda et al., 2003).

**Discontinuous transcription**

Discontinuous transcription (DT) as a mechanism for the synthesis of 3′ co-terminal SG RNAs is exemplified by viruses in the families...
Coronaviridae (prototype virus: *Murine hepatitis virus* or MHV) and *Arteriviridae* (prototype virus: *Equine arteritis virus* or EAIV) of the order *Nidovirales* (Pasternak et al., 2001; Sawicki and Sawicki, 2005). A special feature of these (+)-strand RNA viruses is that they encode a nested set of SG RNAs which vary in size, but all of which have a 5′ sequence co-terminal with the 3′ end of the G RNA. These SG RNAs encode the structural proteins and in the case of the *Coronaviruses*, also several proteins accessory to the replicate proteins. Except for the smallest SG RNAs, all contain multiple open ORFs, but in each case only the 5′ ORF is translated. Thus, although most of these SG RNAs are structurally polycistronic, functionally they are monocistronic.

Also of note, in contrast to the SG RNAs made by internal initiation or by premature termination (see above), those generated by MHV or EAV contain, in addition to coding sequences derived from the 3′ portion of the G RNA, a short sequence that is identical to a 5′ leader sequence of the G RNA. The explanation for how in a SG RNA a 5′ leader is joined to a 3′ terminal sequence remained controversial for some years. Models were proposed that suggested discontinuous synthesis of RNA transcripts, but this then raised the question as to whether the discontinuous synthesis occurred during (+) or (−)-strand synthesis (Pasternak et al., 2006). The first model, known as the leader-primed transcription model (Fig. 1C), proposed that transcription would start by copying the 3′ end of the G RNA (−)-strand RNA thereby giving rise to a 5′ leader sequence; it followed that the discontinuous step would occur during the synthesis of (+)-strand RNA. However, the finding of multiple species of (−)-strand RNAs, the sizes of which corresponded to the sizes of the SG RNAs, provided strong support for the proposal by Sawicki and Sawicki (1985) that discontinuous synthesis of SG RNA occurred at the level of (−)-strand RNA synthesis (Fig. 1D). These (−)-strand RNAs would then serve as the templates for the synthesis of the same-length (+)-strand SG mRNAs and/or the production of shorter internally nested SG RNAs (Wu and Brian, 2010). The latter mechanism would likely contribute to the greater abundance of the 3′-co-terminal SG RNAs.

The leader and body TRSs do not show complete sequence conservation. For instance, with *Murine hepatitis virus* (MHV) DI RNAs, it was found that there were 2–4 copies of the LTRS with the core sequence 5′-UCUAAC-3′, whereas the different B-TRSs had a core sequence centering around 5′-AUUAAC-3′. More important, it has been demonstrated that formation of a duplex between the (−)-strand LTRS in the genome and the (+)-strand copy of a BTRS is required for synthesis of the SG RNA having that TRS at its 5′ end (Figs. 1D and 3A).

The relative contributions of the LTRS and the BTRS to the different 5′-TRS in the SG RNA are variable. In some cases, the entire sequence of the junction TRS in the SG RNA is derived from the BTRS; in other instances, both the L- and the BTRS contribute to the junction TRS. In an interesting experiment, it was found that when cells were infected with two different strains of MHV, many of the SG RNAs had a leader sequence of the co-infecting virus, indicating that the BTRS of one strain can base pair with the LTRS of the second strain (Pasternak et al., 2006).

How is it determined at which BTRS the synthesis of the (−) strand RNA stalls, leading to duplex formation with an LTRS? The level of SG RNA production by both *Arteriviruses* and *Coronaviruses* was shown to depend upon the efficiency of interactions between LTRS and BTRS (Pasternak et al., 2001, 2003; Zúñiga et al., 2004; Sola et al., 2005). Furthermore, it has been proposed that the L-BTRS junctions occur at multiple sites, with a preference for 3′ proximal nucleotides within the BTRS (van der Most et al., 1994). With *Transmissible gastroenteritis coronavirus* (TGEV), it has been shown that sequences flanking the core TRS influence transcription of the SG RNAs (Curtis et al., 2004; Sola et al., 2005). Among *Nidovirales*, the 3′ end of the G RNA is critical for SG RNA transcription. In MHV, the 300 nt 3′ UTR promotes transcription (Lin et al., 1996), but only the 3′ 55 nts are required for (−) strand RNA synthesis. It was also shown that mutations disrupting the U-turn motif of the 5′ UTR stem-loop 2 (SL2) affected SG RNA synthesis, suggesting that SL2 mediates specific interactions with viral and/or cellular proteins involved in the synthesis of SG RNAs (Liu et al., 2007). In general, the 5′ UTRs of *Coronavirus* RNAs fold into similar secondary structures containing three to four SL structures that include a highly conserved 5 nt hairpin loop SL2 with a U-turn motif; SL1 and SL2 are close to each other (shown in Liu et al., 2007). The polypyrimidine-tract binding (PTB) protein hRNP1 may play a role in the regulation of SG RNA transcription due to its ability to interact with short pyrimidine-rich tracts, e.g., UCU, or CUCU. Since the 5′ UTR SL2 contains UCUAA repeats, PTB binding might assist circularization of the viral genome and aid
template switching during discontinuous SG RNA synthesis (Choi et al., 2002; Oberstrass et al., 2005).

Related to the discontinuous mechanism is the question of how the synthesis of SG RNAs is regulated, and how the ratios of the different SG RNAs to each other are maintained. To address this problem, studies were carried out on the nsp1 protein of both EAV and MHV (Tijms et al., 2001; Donaldson et al., 2007). This protein contains a zinc-binding domain that assists in RNA-protein interactions and regulates the replication/transcription balance (Tijms et al., 2001). The EAV nsp1 protein interacts with p100-binding polypeptide (p100BP) which co-functions with the RNA polymerase II transcription factor c-Myb. It has been speculated that the nsp1-p100BP interaction is important for SG RNA synthesis, either directly or by recruiting another protein to the viral RdRP complex. Alternatively, nsp1 might modulate transcription in the infected cell, explaining its targeting to the nucleus (Tijms and Snijder, 2003). Similarly, a deficiency in nucleocapsid protein (N) of Human coronavirus (HCoV-229E) impaired RNA replication, but not transcription, demonstrating that N protein regulates the equilbrium between these two processes (Schelle et al., 2005).

Also, it has been proposed that a long range interaction regulates TGEV SG RNA transcription (Moreno et al., 2008b). Here, a 9-nucleotide (nt) sequence located 449 nt upstream of the N gene TRS core sequence (CS-N) interacts with a complementary sequence immediately upstream of CS-N. The complementarity between these two 9-nt elements in TGEV was functionally relevant in the transcriptional activation of the N gene. Moreover, a positive correlation between the predicted stability of the base-pairing interaction and the accumulation levels of SG RNA N was observed (Moreno et al., 2008b).

Functions of subgenomic RNAs

Role in translation

Most of the SG RNAs generated by (+)-strand RNA viruses function as mRNAs. This is consistent with their composition, i.e. the presence of a 5′ cap (van Vliet et al., 2002) and a 3′ polyA tail (e.g. Wierzchoslawski et al., 2006). Some SG RNAs lack the 5′ cap but contain an IRES structure. For instance, the uncapped SG RNAs in TMV strain U1 (Grdzelishvili et al., 2006). The high recombination frequencies in TGEV was functionally relevant in the transcriptional activation of the N gene. Moreover, a positive correlation between the predicted stability of the base-pairing interaction and the accumulation levels of SG RNA N was observed (Moreno et al., 2008b).

With other viruses, e.g. Turnip crinkle virus (TCV), or Barley yellow dwarf virus (BYDV), the 3′ UTR cap-independent translation elements (CITEs) are important for launching translation of SG RNAs. In some cases a CITE acts cooperatively with an IRES. Among eight distinct structural classes of CITEs described to date, all contain an SL structure which base-pairs to a 5′ UTR sequence. The 3′ location of CITEs favors the translation of SG RNAs over that of genomic RNAs (Qu and Morris, 2000; Scheets and Redinbaugh, 2006; Shen et al., 2006).

Although the 5′-UTR of Potato leafroll virus (PLRV) SG RNA lacks a 3′ translational enhancer (Juszczuk et al., 2000), an IRES signal for translation of replication-associated protein 1 (Rap1) has been identified internally within the PLRV RNA genome 1500 nt downstream of its 5′ end (Jaag et al., 2003). The presence of CA-rich motifs, which increase the length but reduce the secondary structure of the SG RNA 5′UTR boost the protein expression by the SG RNAs of TBSV, Carnation mottle virus (CMV), Cardamine chlorotic fleck virus (CCFV) (Skotnik et al., 1993), and TCV (Qu and Morris, 2000).

Role in replication

One of the first demonstrations that an SG RNA could control viral RNA replication was shown with FHV RNA3, a SG RNA derived from RNA1 (see above) (Eckerle and Ball, 2002). RNA1 mutants deficient in the SG RNA3 synthesis failed to make RNA2. However, when RNA3 was supplied in trans, synthesis of RNA 2 resumed (Eckerle and Ball, 2002; Eckerle et al., 2003). Precisely how RNA3 exerts its effect is not understood. However, the RNA3-dependent replication signal has been mapped to the RNA2 3′ end (Albariño et al., 2003).

The influence of an SG RNA on viral replication is also seen with BYDV (Shen and Miller, 2004). BYDV makes three 3′ co-terminal SG RNAs by internal initiation. SG RNA3 has no coding capacity, and its function is not known. Translation of BYDV G RNA and SG RNA1 is mediated by BTE: barley yellow dwarf (like) translation elementin the 5′end of the 3′ UTR (Fig. 3F) (Shen et al., 2006; Rakotondrafara et al., 2006). BTEs as cap-independent translation elements can be characterized by two structural features: (i) a conserved 17 nt sequence, that includes a stem loop (SL-I) with a GN RNA loop motif, and (ii) a loop (not in SL-I) that can base pair to a loop in the 5′UTR of the RNA (Kneller et al., 2006). It has been proposed that the premature addition of SG RNA2, which contains the 5′ UTR BTE, regulates BYDV replication by inhibiting translation of the viral polymerase from genomic RNA (G RNA). Thus, the G RNA ceases to function as a messenger and instead functions as a template for replication (Shen and Miller, 2004).

A role in separation of translation from replication has been proposed for the 5′ co-terminal SG RNAs of CTV, BMV and CTLV (Gowda et al., 2003; Wierzchoslawski et al., 2006; Tatineni et al., 2009). The 5′ SG RNA can serve as a template for expression of viral protein, e.g. movement protein in BMV, which sets the genomic RNA free for more efficient replication, recombination and packaging (Gowda et al., 2003). As with BYDV SG RNA 2, the early addition of the 5′ co-terminal SG RNA3a reduced BMV RNA replication in barley protoplasts (J. Sztuba-Solilża and J.J. Bujarski, unpublished). In contrast, the in planta experiments showed a dose–dependent bell-shape response in which high concentrations of SG RNA3a reduced virus yield. We speculate that the translation of minute amounts of initial SG RNA3a leads to an excess of MP which facilitates the viral spread. At higher doses the SG RNA3a may, however, act as a molecular decoy that sequesters translational factors, reducing translation of replicate proteins (J. Sztuba-Solilża and J.J. Bujarski, unpublished).

Role in recombination

RNA recombination can salvage damaged or mutated viral RNAs and can contribute to genome variability (Cheng and Nagy, 2003; Chetverin et al., 2005). The exchangeable subgenomic components can facilitate the production of rearranged viral RNA genomes. For instance, the transcription mechanism in Nidovirales, involving leader–body TRS duplex formation together with distinct cis–acting signals that affect the nascent strand transfer during SG RNAs production, resembles the copy-choice RNA recombination (Figs. 1C and D) (van Marle et al., 1999; Pasternak et al., 2006). The high recombination frequencies in Coronavirus and Arteriviruses have been associated with the highly structured 3′ UTRs in the genomes of these viruses (Molenkamp et al., 2000; Pasternak et al., 2000, 2001). Here, the base-pairing between the donor and acceptor molecules at consensus TRS motifs plays a role in the production of Nidoviral SG RNAs, suggestive of similarity-assisted recombination (Yuan et al., 2004; Pasternak et al., 2006). Apparently, these viruses frequently use a transcription strategy for recombination.

The integrity of non-segmented genomes of Closteroviridae is supported by SG RNA-assisted recombination. In the case of CTV defective interfering (DI) RNAs the junction sites coincide with the SG RNA transcription initiation sites (Yang et al., 1997). This suggested that CTV DI RNAs emerged by recombination of SG RNA with a 5′ region of the G RNA (Bar-Joseph et al., 1997). The proposed model emphasizes the role of intergenic AU-rich sequence located between two SL structures, between ORF10 and ORF11 that might induce the SG RNA1 premature termination followed by template switching (Yang et al., 1997; Bar-Joseph et al., 1997). A similar mechanism might explain the acquisition by viruses in the family Closteroviridae of non-self-sequences, either from co-infecting virus or the host (Cuellar et al., 2008).
It is likely that Norovirus (NoV) recombinant isolates originated from SG RNA-mediated rearrangements (Bull et al., 2005). Here, the high frequency recombination could result from RDRP stalling at the SL structure of the ORF1/ORF2 overlap (Rohayem et al., 2005), forcing the enzyme to hop across to either (-) strand SG RNA or G RNA species. As a result, the recombinant Norovirus isolates can acquire new ORF2 and ORF3 sequences (Bull et al., 2005).

The multipartite viruses in the family Bromoviridae also utilize SG RNAs for the modular swap of their genomes. The recently described BMV 5′ co-terminal SG RNA3a can prime recombination events on the (−) and (+) RNA3 strands (Wierzchoslawski et al., 2006; Sztuba-Solińska et al., 2011). These and the following studies revealed the presence of several recombination hot-spots within RNA3 including the 5′ UTR, the upstream encapsidation signal (packaging element called PE), the B-box motif and the intergenic polyU track (Sztuba-Solińska et al., 2011; Sztuba-Solińska and Bujarski, unpublished results). It was proposed that the RDRP associated with other protein factors, e.g. coat protein, might bridge over the RNA3 and SG RNA3a molecules in cis and/or in trans at highly structured RNA elements such as B box-like motif, PE element, or 3′ TLS. This can facilitate template switching during (−) and (+) strand synthesis. Apparently, SG RNAs can act as building components that contribute to genomic rearrangements of complete viral genes.

Further perspectives

The activity of transcriptional regulatory elements, their general structure and sequence context, as well as their interactions with protein factors, all affect the production of SG RNA. It is not known, however, how these structures cause stalling/detachment/reattachment of an actively copying RDRP, or how they affect the timing of SG RNA synthesis. Since RDRPs frequently encounter base-paired regions while copying a template (Ng et al., 2008), explaining how RNA structures affect the progress of an RDRP would greatly improve our understanding of regulation of SG RNA synthesis. Moreover, work with in vitro systems is needed to shed light on both cis- and trans-acting regulatory signals and their cognate factors that affect the production of SG RNA (Nagy and Pogany, 2000; Li et al., 2005a). One problem is that the replication complexes are membrane bound (Mackenzie, 2005; Denison, 2008). However, by using cell-free extracts from evauculated plant protoplasts Komoda et al. (2003) and Okamoto et al. (2008) have carried out preliminary studies on replication, transcription and translation mechanisms with Tomato mosaic virus (ToMV), BMV, TCV, and RCMV.

One question concerning SG RNAs that remains to be answered concerns our understanding of how the ratio of genomic to SG RNA synthesis is maintained. Besides the relative strength of the promoters the concentrations of the various NTPs plays a role in determining this ratio (Li et al., 2008, 2010). A related question is whether there is competition between the synthesis of G RNA and SG RNA. With respect to viruses with multiple SG RNAs, why are some RNAs made in larger amounts than others? Much also remains to be learned about the particular mechanisms of SG RNAs synthesis and why certain SG RNAs are more robust messengers than other SG RNAs or the G RNA. A better understanding of SG RNAs and their functions may provide us with new targets for antiviral therapy, specifically concerning important diseases such as SARS, caused by a Coronavirus.

The idea that SG RNAs may be a tempting target for antiviral therapy is supported by the identification of plant-derived compounds (Picard et al., 2005; Li et al., 2007) that inhibit plant and animal viruses, e.g. in Alphaviruses, at nanomolar concentrations. These compounds appear to act by preferential inhibition of synthesis of SG RNAs. Although the exact mechanism is not known, one possibility is that they bind to sites on the RDRP that recognize the SGs.

Also, the ability to exogenously regulate SG RNAs transcription, e.g., by using a small-molecule ligand and aptamer technology (Wang et al., 2008), offers a promising way to modulate the timing and the levels of SG RNA transcription which, in turn, would provide a means to control viral protein expression. This approach has already found use in regulating the function of the higher-order component of the attenuation signal allowing the control of BSV SG RNA transcription (Wang and White, 2007; Wang et al., 2008) and it offers further promising ways for the analysis and modulation of viral processes at either the level of SG RNA transcription or translation.

Since SGPs have been identified as recombination hot spots (Suzuki et al., 2003; Wierzchoslawski et al., 2004; Coyne et al., 2006), the understanding of the mechanism of SG RNA transcription may provide insights into the origin and evolution of viruses. The continual emergence of new recombinant strains, e.g. HCoV-NL63 and HCoVHKU1 isolated in the wake of SARS (van der Hoek et al., 2004), and SARS-like Coronaviruses isolated from animal reservoirs (Li et al., 2005b), raises the question of the role of SG RNAs in the reshuffling of genome sequences of Corona- and other viruses. Clearly, by incorporating foreign sequences into their genomes, certain viruses acquire new surprising properties (Pasternak et al., 2006; Moreno et al., 2008a). For example, the Closterovirus SG RNA-expressed HSP70h, that facilitates viral assembly and cellular movement, has likely been captured from the host (Prokhnevsky et al., 2005). The Coronavirus SG RNA-expressed hemagglutinin-esterase (HE), acting as a receptor-binding fusion protein, likely originated from Orthomyxoviruses (Zeng et al., 2008). Viruses that generate SG RNAs are increasingly being used as expression vectors. For example, with Alphaviruses a foreign gene can be placed under the control of the SGP promoter, replacing the genes encoding the structural protein (Rayner et al., 2002; Rausalu et al., 2009), RNAs expressed from this promoter are translated in large amounts, producing proteins which can be used in vaccines, or as therapeutics. Also, by incorporating an RNA sequence encoding an antibody chain that recognizes a ligand on a cell surface, into an SG RNA, one can produce proteins targeted to a specific cell type.

Acknowledgments

This work was supported by grants from National Science Foundation (MCB-0920617) and National Institutes of Health (G1A62203) and by the Plant Molecular Biology Center at Northern Illinois University to JJB, and by grants from the National Institutes of Health (AI49273) and (AI070728) to VS.

References

Adkins, S., Kao, C.C. 1998. Subgenomic RNA promoters dictate the mode of recognition by bromoviral RNA-dependent RNA polymerases. Virology 252, 1–8.

Albariño, C.G., Eckerle, L.D., Ball, L.A. 2003. The cis-acting replication signal at the 3′ end of Flock House virus RNA2 is RNA3-dependent. J. Virol. 76, 1293–1309.

Alonso, S., Izeta, A., Sola, I., Enjuanes, L., 2002. Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. J. Virol. 76, 1293–1309.

Bar-Joseph, M., Yang, G., Gafny, R., Mawassi, M., 1997. Subgenomic RNAs: the possible building blocks for modular recombination of Coronaviridae genomes. Semin. Virol. 8, 113–119.

Bull, R.A., Hansman, G.S., Clancy, L.E., Tanaka, M.M., Rawlinson, W.D., White, P.A., 2005. Norovirus recombination in ORF1/ORF2 overlap. Emerg. Infect. Dis. 11, 1079–1085.

Che, X., Piestun, D., Mawassi, M., Yang, G., Satyanarayana, T., Gowda, S., Dawson, W.O., Bar-Joseph, M., 2001. 5′-coterminal subgenomic RNAs in Citrus tristeza virus-infected cells. Virology 283, 374–381.

Cheng, C.P., Nagy, P.D., 2003. Mechanism of RNA recombination in carmop- and tombusviruses: evidence for template switching by the RNA-dependent RNA polymerase in vitro. J. Virol. 77, 12033–12047.

Chetverin, A.B., Kopein, D.S., Chetverina, I.V., Demidenko, A.A., Urgazov, V.I., 2005. Viral RNA-directed RNA polymerases use diverse mechanisms to promote recombination between RNA molecules. J. Biol. Chem. 280, 8748–8755.

Choi, J.R., White, K.A., 2002. An RNA activator of subgenomic mRNA1 transcription in tomato bushy stunt virus. J. Biol. Chem. 277, 3760–3766.

Choi, J.R., Ostrovsky, M., Zhang, G., White, K.A., 2001. Regulatory activity of distal core RNA elements in tombus virus subgenomic mRNA2 transcription. J. Biol. Chem. 76, 41761–41768.
Kim, K.H., Hemenway, C.L., 1999. Long-distance RNA interactions but not translation of mouse hepatitis virus RNA. Virology 303, 58–68.

Covey, J.A., Zimmock, C.M., Walker, P.J., 2002. Gli-associated noddus of Pervanodon prawns transcribes 3′-coterminal subgenomic mRNAs that do not possess 5′-leader sequences. J. Gen. Virol. 83, 927–935.

Coyne, K.P., Reed, F.C., Porter, C.J., Dawson, S., Gaskell, R.M., Radford, A.D., 2006. Recognition of Flock house calicivirus within an endemically infected cat colony. J. Gen. Virol. 87, 921–926.

Cuellar, W.J., Tairo, F., Kreuze, J.F., Valkonen, J.P.T., 2008. Analysis of gene content in sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in its capsid. J. Gen. Virol. 89, 2093–2097.

Curtis, K.M., Yount, B., Sims, A.C., Baric, R.S., 2004. Reverse genetic analysis of the replication regulatory region of the coronavirus transmissible gastroenteritis virus. J. Virol. 78, 6061–6066.

Denison, M.R., 2008. Seeking membranes: positive-strand RNA virus replication complexes. PLoS Biol. 6 (e270), 2098–2100.

Diez, J., Ishikawa, M., Kaido, M., Ahlquist, P., 2000. Identification and characterization of a host protein required for efficient template selection in viral RNA replication. Proc. Natl Acad. Sci. USA 97, 10911–10916.

Donaldson, E.F., Sims, A.C., Graham, R.L., Denison, M.R., Baric, R.S., 2007. Murine coronavirus RNA3 transcribes a 3′-terminal and positive-stranded 5′-terminal RNAs. Virology 370, 95–108.

Gallakpour, N., Goszczynski, D.E., Che, X., Batuman, O., Bar-Joseph, M., Mawassi, M., 2003. Two classes of subgenomic RNA of Grapevine virus A produced by internal controller elements. Virology 312, 434–448.

Gowda, S., Satyanarayana, T., Ayillo, M.A., Albach-Marti, M., Mawassi, M., Rabindran, S., Garsey, S.M., Dawson, W.O., 2001. Characterization of the cis-acting elements controlling subgenomic mRNAs of citrus tristeza virus: production of positive and negative-stranded 3′-terminal and positive-stranded 5′-terminal RNAs. Virology 286, 134–151.

Gowda, S., Ayillo, M.A., Satyanarayana, T., Bar-Joseph, M., Dawson, W.O., 2003. Translation strategy in a closterovirus: a novel 5′-terminal controller element of citrus tristeza virus produces 5′- and 3′-terminal subgenomic RNAs and differs from 3′-open reading frame controller elements. J. Virol. 77, 340–352.

Grzeličkivští, V.Z., Chapman, S.N., Dawson, W.O., Lommel, S.A., 2004. Structural characterization of an intermolecular cis-acting element of Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA replication. J. Virol. 77, 6774–6784.

Grzeličkivští, V.Z., Chapman, S.N., Dawson, W.O., Lommel, S.A., 2004. Structural characterization of an intermolecular cis-acting element of Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA replication. J. Virol. 77, 6774–6784.

Hahm, H.O., Kinsey, P.J., Eapen, S.P., Saha, A., Parinaud, Y., Gourdon, M., 2000. Murine coronavirus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. J. Virol. 74, 3915–3919.

Haasnoot, P.C.J., Brederode, F.Th., Olsthoorn, R.C.L., Bol, J.F., 2000. A conserved hairpin structure in Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA replication. J. Virol. 74, 3915–3919.

Hahn, S.M., Vogel, S., Ascher, P., Li, M.L., Wang, H., Stollar, V., 2010. Detailed kinetic analysis of intracellular nucleoside triphosphate concentrations. J. Virol. 84, 2739–2749.

Hahn, S.M., Zúñiga, S., Enjuanes, L., Sola, I., 2008b. Identification and characterization of a host protein required for efficient template selection in viral RNA replication. Proc. Natl Acad. Sci. USA 105, 19079–19084.

Hahn, S.M., Zhao, Y., Kwan, Y.J., Zhang, X., Wu, R.C., Lai, M.M.C., 1996. The 3′ untranslated region of coronavirus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. J. Virol. 70, 7236–7245.

Hahn, S.M., Zúñiga, S., Enjuanes, L., Sola, I., 2008b. Identification and characterization of a host protein required for efficient template selection in viral RNA replication. Proc. Natl Acad. Sci. USA 105, 19079–19084.

Hahn, S.M., Zúñiga, S., Enjuanes, L., Sola, I., 2008b. Identification and characterization of a host protein required for efficient template selection in viral RNA replication. Proc. Natl Acad. Sci. USA 105, 19079–19084.
