

Universiteit
Leiden
The Netherlands

RNA structures regulating nidovirus RNA synthesis

Born, E. van den

Citation
Born, E. van den. (2006, February 1). *RNA structures regulating nidovirus RNA synthesis*. Retrieved from https://hdl.handle.net/1887/4285

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4285

Note: To cite this publication please use the final published version (if applicable).

RNA Structures Regulating Nidovirus RNA Synthesis

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 1 februari 2006
klokke 15.15 uur

door

Erwin van den Born
geboren te Wageningen
in 1977

Promotiecommissie

Promotor:	Prof. Dr. W.J.M. Spaan
Co-promotor:	Dr. E.J. Snijder
Referent:	Prof. Dr. L. Enjuanes (Centro Nacional de Biotecnologia)
Overige Leden:	Prof. Dr. R.W. Goldbach (Wageningen Universiteit)
	Dr. A.A.F. de Vries
	Prof. Dr. C.W.A. Pleij

The work described in this thesis was financially supported by grant CW 99-010 from the Council for Chemical Sciences of the Netherlands Organisation for Scientific Research (NWO-CW). NWO-CW also contributed to the print costs of this thesis.

Table of contents

Chapter 1	General Introduction	**5**
Chapter 2	RNA Elements; Basic Tools for Regulating Key Steps in the Nidovirus Life Cycle (A Literature Review)	**19**
Chapter 3	Sequence Requirements for RNA Strand Transfer During Nidovirus Discontinuous Subgenomic RNA Synthesis	**47**
Chapter 4	Secondary Structure and Function of the 5'-Proximal Region of the Equine Arteritis Virus RNA genome	**69**
Chapter 5	Discontinuous Subgenomic RNA Synthesis in Arteriviruses Is Guided by an RNA Hairpin Structure Located in the Genomic Leader Region	**95**
Chapter 6	Antiviral Activity of Morpholino Oligomers Designed to Block Different Aspects of Equine Arteritis Virus Reproduction in Cell Culture	**123**
Chapter 7	Discussion	**143**
References		**155**
Summary		**173**
Samenvatting		**177**
Curriculum Vitae		**181**

Chapter 1

General Introduction

Higher order structures play a pivotal role in biological processes

Macromolecules form the basis of life. They catalyze chemical transformations, build complex multimolecular structures, generate movement, and store and transmit genetic information. Even though macromolecules are made from a limited repertoire of building blocks, each can have unique and dedicated properties. The macromolecule as a whole adopts a specific three-dimensional structure that depends on the sequence of its building blocks. Each of these macromolecular structures carries specific biological information that can be read or used during its interactions with other molecules, enabling it to perform a precise function in one of the processes mentioned above.

The hereditary information of life is stored in genes that are generally made of the polynucleotide deoxyribonucleic acid (DNA) and has the form of a sequence of nucleotides carrying the bases adenine (A), cytosine (C), guanine (G) or thymine (T). Two strands of DNA can form a double helix on the basis of base-pairing interactions (hydrogen bonds) between the purine A in one chain and the pyrimidine T in the other, or between the purine G in one chain and the pyrimidine C in the opposing strand. Since each strand contains a nucleotide sequence that is exactly complementary to the nucleotide sequence of the partner strand, both strands actually carry the same genetic information. For each organism, it is obviously crucial that its genetic information is precisely copied and passed on to its progeny. To achieve this, DNA is copied ("replicated") using a process in which the two DNA strands are separated and each strand then serves as a template for the production of a new complementary partner strand (Watson & Crick, 1953a; Watson & Crick, 1953b; Meselson & Stahl, 1958; Felsenfeld, 1985).

To a large extent, a cell is made of protein, which can constitute more than half of its dry weight. Proteins determine the shape and structure of the cell and also serve as the main instruments of molecular recognition and catalysis. DNA and the chemically similar RNA (see below) consist of just four building blocks, proteins are made from an assortment of 20 very different amino acids, each with distinct chemical properties. Amino acids can be modified (phosphorylation, glycosylation, acetylation,…), which creates even more variation in the pool of available building blocks. The position and variety of the amino acids within the folded polypeptide makes each protein unique and enables it to bind specifically to other macromolecular surfaces and to certain small molecules. The variation of size and amino acid composition is the basis for the chemical versatility of proteins and the ability to form different structures (Dill *et al.*, 1995). For example, collagen molecules consist of three separate polypeptide chains that form a regular triple helix, which gives them enormous tensile strength. Elastin, on the other hand, can form a rubber-like elastic meshwork. The elasticity is due to the ability of the individual protein molecules to uncoil reversibly whenever a stretching force is applied. This example illustrates once more the added value of higher order structure of macromolecules in biology (Gosline *et al.*, 2002). The versatility of proteins presumably

explains why evolution selected proteins rather than RNA molecules to catalyze the majority of cellular reactions.

As an intermediate step in protein synthesis, DNA genes need to be copied into polynucleotides of a chemically and functionally different type known as ribonucleic acid, or RNA, a process known as transcription. Like DNA, RNA is composed of a linear sequence of nucleotides, but it has two small chemical differences. Firstly, the sugar-phosphate backbone of RNA contains a ribose instead of a deoxyribose sugar, and secondly, the base thymine (T) is replaced by uracil (U), a very closely related base that can also pair with an A. In contrast to DNA, RNA usually is a single-stranded molecule. RNA retains all of the information of the DNA sequence from which it is copied, as well as the base-pairing properties of DNA. The amount of RNA made from a particular region of DNA is controlled by regulatory proteins (transcription factors) that bind to specific sites on the DNA template (promoter elements) close to the protein-coding region. RNA transcripts that direct the synthesis of protein molecules are called messenger RNA (mRNA) molecules, while other RNA transcripts serve as transfer RNAs (tRNAs) or form the RNA components of ribosomes (rRNA) or smaller ribonucleoprotein particles (Darnell, Jr., 1985). The nucleotide sequence in an mRNA molecule is translated into a protein sequence by reading groups of three nucleotides. Each such triplet, called a codon, specifies one amino acid building block of a protein chain. The translation of mRNA into protein depends on adapter molecules, tRNAs, and ribosomes. tRNAs are able to both bind an amino acid and recognize the codons of the mRNA. tRNAs are small RNA molecules of about 80 nucleotides in length and have a highly folded three-dimensional L-shaped conformation that is held together by intrachain base-pairing interactions. Four short segments of the molecule contain a double-helical structure, producing a molecule that looks like a cloverleaf in two dimensions. The 3’ end of the tRNA is covalently attached to a specific amino acid, while an unpaired region in the middle of the molecule forms the so-called anticodon that base-pairs to the complementary codon in an mRNA molecule (Rich & Kim, 1978). The mechanics of ordering the tRNA molecules on an mRNA and making the peptide bonds between the amino acids that they carry is complex and require a ribosome, a complex of more than 50 different proteins associated with several structural RNA molecules (rRNAs).

RNA function associated with its higher order structure

RNA molecules have been viewed as strings of nucleotides with a relative uninteresting chemistry, but this view changed after the discovery in 1982 of a catalytic RNA molecule with the type of sophisticated chemical reactivity that was previously thought to be associated with proteins only. This ribosomal RNA molecule from *Tetrahymena* (a ciliated protozoan) is produced as a large precursor from which an RNA splicing reaction produces one of the rRNAs catalyzed by the RNA molecule itself in the

absence of proteins. During RNA splicing, an internal sequence (intron) is excised from a precursor RNA. For this reaction, the RNA forms a higher order structure to form a complex surface that can function repeatedly, like an enzyme in reactions with other molecules (Kruger *et al.*, 1982). Since the discovery of the *Tetrahymena* self-splicing intron, several other types of catalytic RNAs, or ribozymes, have been discovered (Scott & Klug, 1996; DeRose, 2002). A catalytic RNA sequence also plays an important role in the life cycle of many plant viroids (Gora-Sochacka, 2004). Most remarkably, ribosomes are now suspected to function largely by RNA-based catalysis, with the ribosomal proteins only playing a supporting role for the rRNAs (Steitz & Moore, 2003). Moreover, there are strong indications that self-replicating RNA systems, mixed with other organic molecules including simple polypeptides, played an important role during the early stages of the evolution of life (Paul & Joyce, 2002; Cech, 2002).

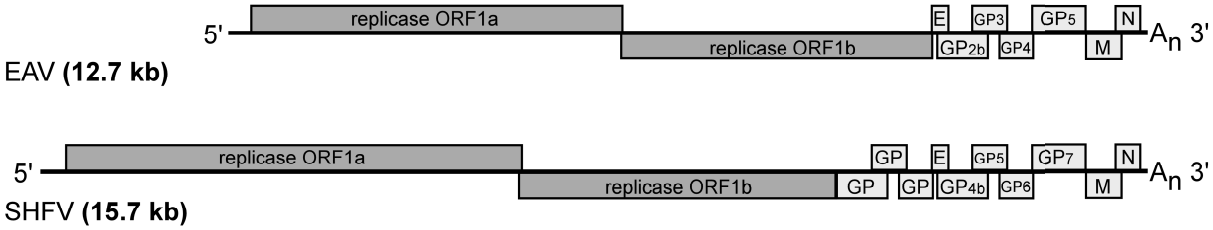
Regulation of RNA virus genome expression by RNA structure

Regulation of replication and transcription of DNA viruses is mainly governed by the primary structure (sequence) of their double stranded DNA genomes. Within the genome, promoter elements are present that are recognized by a polymerase. The polymerase can be associated with accessory proteins to form a polymerase complex. After successful positioning of a DNA or RNA polymerase complex on a promoter element, replication or transcription can be initiated by these complexes. During replication, an exact copy of the DNA genome is synthesized. The new copy can be used as template to make additional copies, it can be incorporated into virus particles, or it can be used as a template for transcription. During the latter process, mRNAs are produced that are translated by the cellular translation machinery to generate the viral proteins. These replication and transcription processes are regulated primarily by controlling the accessibility of the promoter elements via regulatory proteins that can specifically bind to these elements (Latchman, 1999; Miller *et al.*, 2003; Liu *et al.*, 2003a).

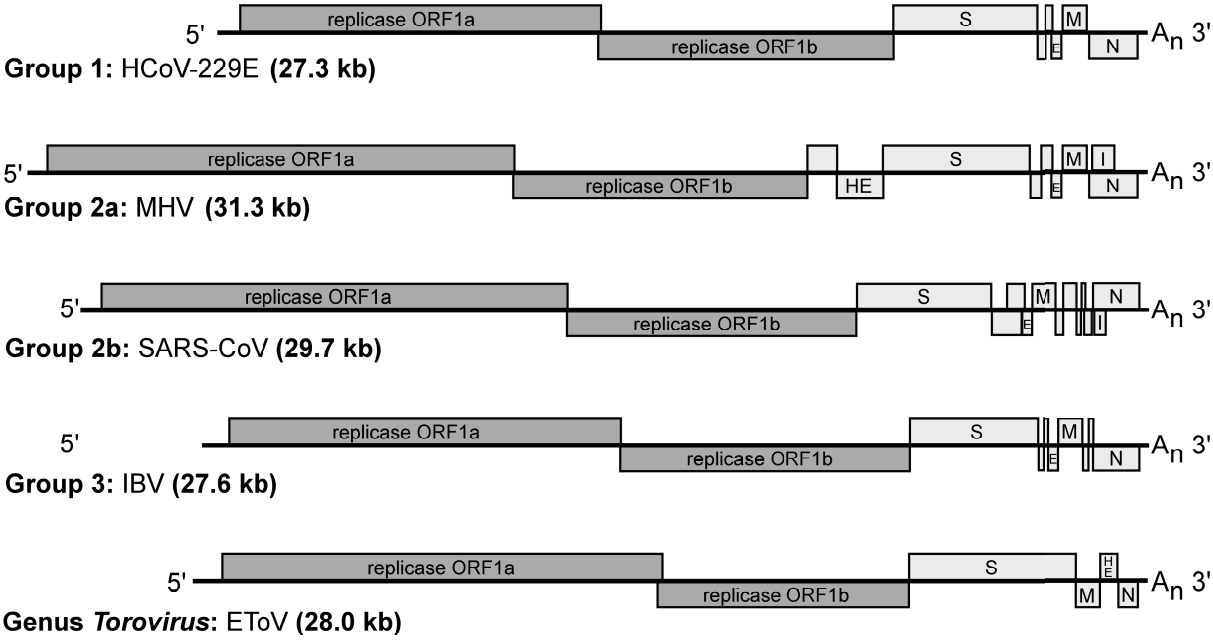
A much larger group of viruses have their genetic information stored in a RNA molecule. In the case of plus-stranded RNA viruses the genome is involved in four different key processes, i.e. genome translation, replication, transcription and encapsidation into progeny virions. Higher order RNA structures in the viral genome play a crucial role in the regulation of these processes. First of all, the genomic RNA is used directly as an mRNA and is translated to produce viral proteins. To be able to act as mRNAs, the virus genome requires a 5' cap structure and a 3' poly(A) tail to ensure its recognition by the cellular translation machinery (Gallie, 1998). Some virus genomes lack either or both of these structures and have developed other means to recruit ribosomes for translation, such as (highly structured) internal ribosome entry sites (IRES) (Vagner *et al.*, 2001; Bol, 2005). Secondly, replication of RNA viruses involves the recognition of RNA signals (or promoters) that promote the association of the 3' end of the viral genome with a

complex of proteins that includes the RNA-dependent RNA polymerase (RdRp). This RdRp complex initiates the synthesis of full-length minus-strand RNA (or antigenome), which in turn will be the template for the synthesis of novel genome RNA. For the latter process, promoter elements present in the 3' end of the antigenome are required. Both RNA primary and secondary structures in promoter elements are involved in the interaction of the genome with the RdRp. Enhancer elements, which may be located at some distance from the actual promoter, can contribute to promoter activity via binding of host and/or viral factors (Blackwell & Brinton, 1997; Melchers *et al.*, 1997; Sivakumaran *et al.*, 1999; Agrawal *et al.*, 2001; Gowda *et al.*, 2003). A third process, found in several groups of RNA viruses, involves the generation of subgenomic (sg) mRNAs as a strategy to express one or multiple viral proteins (Miller & Koev, 2000). In addition to internal initiation of translation, the production of sg mRNAs provides a mechanism to express genes positioned downstream of the 5'-proximal cistron in polycistronic genomes of eukaryotic RNA viruses. Higher order RNA structures are crucial element of sg promoters and are involved in termination of minus strand synthesis (premature termination model), or initiation of plus strand synthesis (internal initiation model) (Miller & Koev, 2000; White, 2002). Finally, to protect the genome during the extracellular phase of the virus life cycle and to ensure that the viral genome is transferred to other cells or hosts, the genome is packaged into a virus particle. The structural proteins that form the capsid of the particle have to specifically recognize the virus genome in order to exclude that RNA of the host cell is incorporated into a particle. Apart from specific nucleotide sequences, specific RNA structures of the packaging signal play an important role in RNA-protein interactions (Choi *et al.*, 2002b; Koning *et al.*, 2003). The multifunctionality of viral RNA genomes requires a tight control of different processes in which RNA structures are definitely important regulators. For example, without regulation, translating ribosomes migrating from the 5' end to the 3' end of the genome will collide with replicases moving in the opposite direction to synthesize the minus strand. The 5’-proximal region of the plus-stranded RNA genome may contain regulatory RNA elements. In addition to being the initiation site for translation of the viral (poly)proteins, this region may be involved in genome replication, genome encapsidation, and, in the case of nidoviruses, also sg RNA synthesis.

Arteriviridae



Coronaviridae



Roniviridae

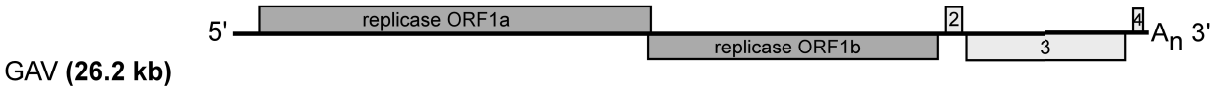


Figure 1-1. Genome organization of selected representatives of the Arteriviridae, Coronaviridae, and Roniviridae families. *The ORFs in the genome are indicated, but only the names of the replicase protein and structural protein genes are given. Note that the arterivirus genome is drawn to a different scale than the genomes of corona-, toro-, and roniviruses. The 3' poly(A) tail is indicated by A_n. EAV, Equine arteritis virus; SHFV, Simian hemorrhagic fever virus; HCoV-229E, Human coronavirus 229E; MHV, Mouse hepatitis virus; SARS-CoV, Severe acute respiratory syndrome coronavirus; IBV, Infectious bronchitis virus; EToV, Equine torovirus; GAV, Gill-associated virus. Adapted from Siddell et al. (2005).*

Nidoviruses

The order *Nidovirales* includes three families, *Coronaviridae* (consisting of the genera *Coronavirus* and *Torovirus*), *Arteriviridae* and *Roniviridae* (consisting of the genus *Okavirus*). The genus *Coronavirus* is further subdivided into groups 1, 2 and 3 (Cavanagh, 1997; González *et al.*, 2003; Spaan *et al.*, 2004; Walker *et al.*, 2004). Several nidoviruses cause respiratory and/or enteric disease in livestock or companion animals and human coronaviruses are recognized as the second most frequent cause of the common cold syndrome (Saif *et al.*, 1988; Albina, 1997; Makela *et al.*, 1998; Cowley *et al.*, 2000; Cook, 2002). Soon after the severe acute respiratory syndrome (SARS) outbreak in 2003, a coronavirus (SARS-CoV) was identified as the etiological agent (Ksiazek *et al.*, 2003; Drosten *et al.*, 2003; Peiris *et al.*, 2003). Subsequently, a few additional novel human coronaviruses were identified (Fouchier *et al.*, 2004; van der Hoek *et al.*, 2004; Woo *et al.*, 2005), and more may follow, now that the significance of this group of viruses has been recognized in the wake of the SARS outbreak. Currently, many aspects of SARS-CoV and the disease that it causes are being characterized at an unprecedented rate (Stadler *et al.*, 2003; Marra *et al.*, 2003; Rota *et al.*, 2003; Thiel *et al.*, 2003; Snijder *et al.*, 2003; Berger *et al.*, 2004; Ziebuhr, 2004). It has to be noted that its pathogenicity for humans necessitates drastic and inconvenient safety measures when SARS-CoV is studied in the laboratory. In the past, the coronavirus *Mouse hepatitis virus* and the arterivirus prototype *Equine arteritis virus* (EAV) have proven to be safe alternatives and suitable tools to study the molecular biology of nidoviruses in general. Future studies on these viruses will certainly continue to contribute to increase our understanding of nidoviruses.

Nidoviruses are enveloped, positive-strand RNA viruses that share a similar genome organization (Fig. 1-1) and similar replication strategy. However, they differ significantly in terms of genome size (12.7 kb to more than 30 kb), virion morphology (Fig. 1-2), host range and various biological properties. Nidoviruses have a single-stranded, linear RNA genome that contains a 5' cap structure and a 3' poly(A) tail (Fig. 1-1). The nidovirus replicase gene is comprised of two large and overlapping open reading frames (ORFs) each encoding a large replicase polyprotein, pp1a and pp1ab, respectively, that are proteolytically processed by three or four virus-encoded proteinases to produce the functional subunits, such as key virus enzymes like helicase and RdRp (Fig. 1-1). Expression of the downstream ORF1b is mediated by programmed –1 ribosomal frameshifting just upstream of the ORF1a termination codon (Ziebuhr *et al.*, 2000; Snijder & Meulenberg, 2001; Ziebuhr, 2005). A nested set of at least two (but usually five to nine) 3′ co-terminal sg mRNAs is produced in the infected cell to express the viral structural proteins (and sometimes also accessory proteins) from genes located in the 3'-proximal third of the genome (the region downstream of ORF1b), which remains inaccessible for ribosomes translating the genome RNA (Fig. 1-3). Generally, only the ORFs contained within the 5′ unique regions of each mRNA, i.e. the regions not found in

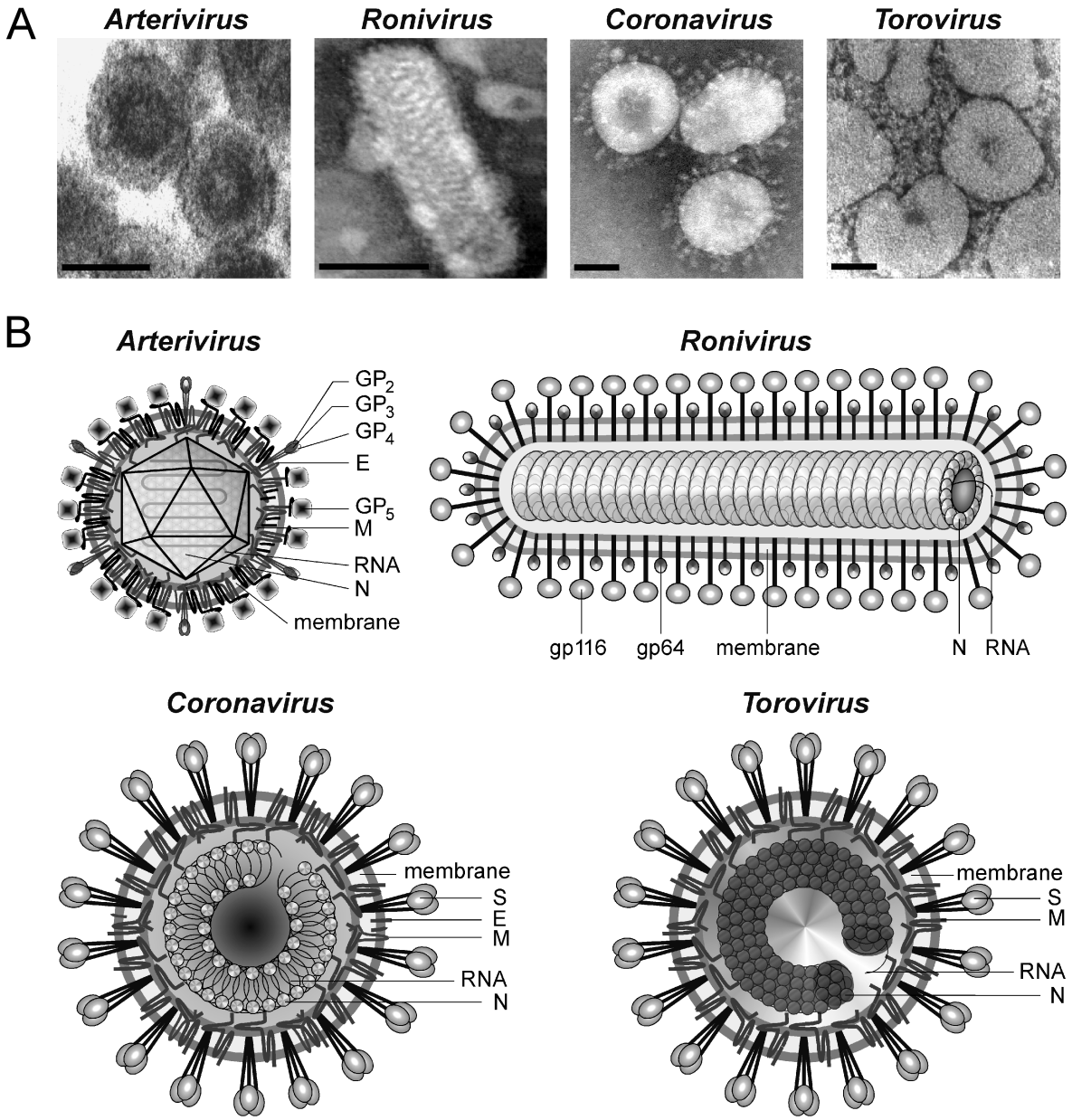


Figure 1-2. Electron micrographs and schematic representations of nidovirus particles. *(A) Characteristic images of negatively stained arterivirus, ronivirus, coronavirus and torovirus particles, as captured with a transmission electron microscope. The diverse size and different structural features of representatives of the four main nidovirus branches are evident. Images were reproduced from Snijder & Meulenberg (1998), arterivirus; Spann et al. (1995), ronivirus; Dr. F. Murphy, Centers for Disease Control and Prevention (CDC), Atlanta, USA, coronavirus; and Weiss et al. (1983), torovirus. The scale bar represents 50 nm. (B) Artist impression of arterivirus, ronivirus, coronavirus and torovirus virion structure. The different symmetries of the nucleocapsid structures and virions are indicated. Viral membrane proteins present in the virion envelope are depicted. Note that some viruses have been found to contain additional envelope proteins, e.g. the HE (hemagglutinin esterase) protein identified in certain coronaviruses and toroviruses. The stoichiometry of the virion components is shown (more or less) arbitrarily. Abbreviations: N, nucleocapsid protein; M, membrane protein; S, spike protein; E, envelope protein; GP or gp, glycoprotein. Adopted from Snijder et al. (2005).*

the next smallest mRNA, are expressed. Arteri- and coronavirus transcription is regulated by transcription-regulating sequences (TRSs) that are located both at the 5' end of the genome and preceding each ORF in the 3'-terminal third of the genome (van Marle *et al.*, 1999a; Zuniga *et al.*, 2004). Transcription of the largest torovirus sg mRNA is presumably also regulated by such a TRS element. Transcription of all other torovirus sg mRNAs and each ronivirus sg mRNA is regulated by TRS-like promoter element (Fig. 1-3) (van Vliet *et al.*, 2002; Cowley *et al.*, 2002).

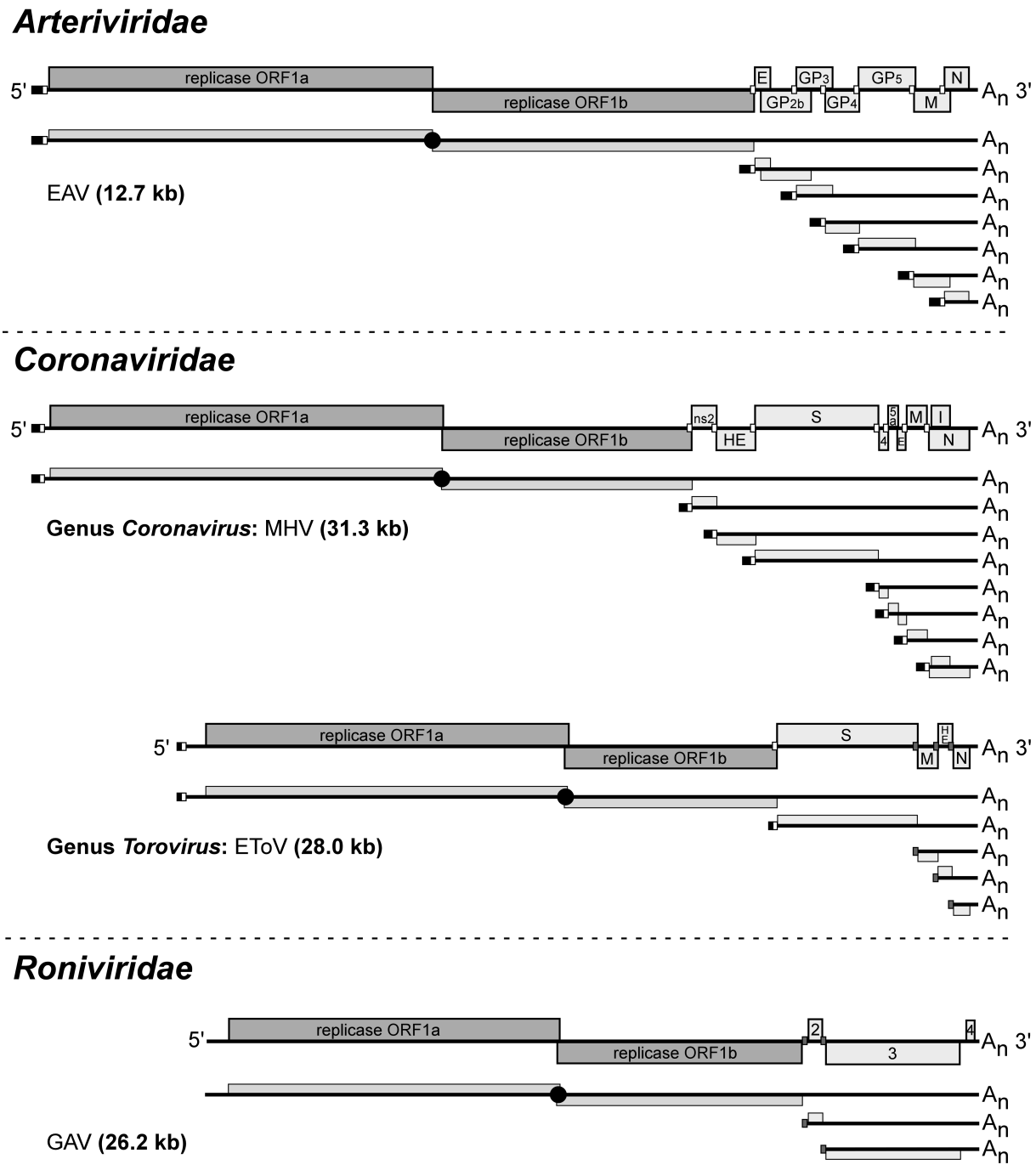


Figure 1-3. Nidoviruses produce a nested set of sg mRNAs. *The genome expression strategies of EAV, MHV, EToV, and GAV. The figure shows the genomic organization of the viruses and the structural relationships of the genome-length and subgenome-length mRNAs. The leader sequence and TRS sequences found at the 5' end of EAV and MHV mRNAs, as well as the genome-length and the largest sg mRNA of EToV are indicated as black and white boxes, respectively. The TRS-like element found at the 5' end of all other EToV sg mRNAs and GAV sg mRNAs is indicated as a dark gray box. The ribosomal frameshifting element found in genome-length mRNA is indicated as a black circle. Only the translated ORFs are indicated for each mRNA. Adapted from Siddell et al. (2005).*

Outline of this thesis

The complex organization of the large nidovirus RNA genome (up to 32 kb) implicates a sophisticated regulation of its different functions. Several regulatory RNA elements have already been identified in nidovirus genomes. RNA secondary structures were assigned to some of these elements and were found to be involved in replication, packaging, or translation. To date, the TRS is the only RNA element that has been rigorously proven to be involved in regulating arteri- and coronavirus sg RNA synthesis. The aim of the work described in this thesis was to further dissect the mechanism of arterivirus sg RNA synthesis, in particular the involvement of RNA sequences and structures located in the 5'-proximal region of the EAV genome.

Chapter 2 presents a literature review summarizing data on *cis*-acting RNA elements that regulate different aspects of the nidovirus lifecycle.

Chapter 3 addresses the duplex formation between the leader TRS and the body TRS. A site-directed TRS mutagenesis approach was applied to make all possible mutations in the hexanucleotide leader TRS and RNA7 body TRS. Besides support for leader TRS-to-body TRS base-pairing, evidence for body TRS-specific functions was obtained. The data constitute additional support for the "discontinuous extension of minus-strand RNA" model for nidovirus sg RNA synthesis.

Chapter 4 describes the structural and functional characterization of the 5'-proximal region of the EAV genome to identify RNA structures involved in sg RNA synthesis. A detailed RNA secondary structure model was established using bioinformatics, phylogenetic analysis, and RNA structure probing. The existence of a prominent hairpin (Leader TRS Hairpin; LTH) was corroborated, which carries the leader TRS located in its loop. Indications for a direct role of the LTH in sg RNA synthesis were obtained through a limited mutagenesis study. Similar LTH structures could be predicted in the 5'-proximal region of all arterivirus and most coronavirus genomes.

Chapter 5 describes experiments to delimit the sequences in the 5'-proximal region of the EAV genome that are required for sg RNA synthesis. A full-length cDNA clone was engineered in which mutations in the 5'-proximal region could be made to monitor their effect on sg RNA synthesis without seriously affecting genome replication and translation. The LTH and its immediate flanking sequences were found to be essential for efficient sg RNA synthesis. The possibility of a RNA conformational switch in the LTH region that regulates the functions of the 5'-proximal region of the arterivirus genome in sg RNA synthesis is discussed.

Chapter 6 presents a study in which the antiviral potential of antisense morpholino oligomers (PMOs), targeting specific EAV RNA sequences, was evaluated. PMOs designed to interfere with several aspects of virus amplification were tested. Using the current knowledge of EAV transcription, in part based on the data presented in Chapters 3, 4, and 5, oligomers were designed that could specifically affect sg RNA synthesis, especially by interfering with leader TRS-to-body TRS interactions. Although this

approach was unsuccessful, targeting of the 3’ terminus of plus or minus strand resulted in a moderate reduction of virus amplification. Complete inhibition of virus replication was achieved when translation was blocked using oligomers targeting the 5’ untranslated region (5’ UTR) of the genome.
Chapter 7 discusses the results described in this thesis in the context of other recent findings for nidovirus RNA synthesis.

Chapter 2

RNA Elements;
Basic Tools for Regulating Key Steps in the Nidovirus Life Cycle

(A literature review)

Replication signals in the RNA genome of *Coronaviridae*

During the past two decades, defective interfering (DI) genomes have been a useful tool to investigate RNA elements involved in coronavirus genome replication. DI genomes are generated due to errors in genome replication and require an intact helper virus in order to replicate. They interfere with the helper virus by replicating at its expense. Cloned cDNA copies of DI RNA genomes could be easily modified by mutagenesis and *in vitro* generated transcripts were transfected into infected cells to study the consequences of such changes. The amplification of these minigenomes relies on functions provided *in trans* by a helper virus. The intracellular presence of the RNA genome of both the DI RNA replicon and the helper virus creates the inevitable risk of (homologous) recombination between both molecules. This is a major disadvantage of the DI RNA system, since it is often difficult to determine to which extent repair of the mutations by recombination with sequences of the helper virus genome have contributed to the observed phenotype. The amplification step that is often required to detect DI RNA replicons, which is commonly achieved by passaging the DI RNA after transfection, is another disadvantage of the DI RNA system. DI RNAs that lack a functional packaging signal cannot be passaged and are therefore not detected, but may still be replication competent. Systems were developed to analyze DI RNA replication without passaging, to reduce the consequences of recombination and to separate packaging from replication. For example, a replicating DI RNA that was expressed in helper virus-infected cells using the *Vaccinia virus* T7 expression system, was readily detected, but was not packaged and could not be retrieved in subsequent passages (Lin & Lai, 1993). The recent development of reverse genetic systems based on cloned full-length cDNA of coronavirus genomes will certainly reduce the use of DI RNA systems to study replication, and further improve our understanding of coronavirus replication signals (Yount *et al.*, 2000; Almazan *et al.*, 2000; Casais *et al.*, 2001; Yount *et al.*, 2002; Yount *et al.*, 2003; Hertzig *et al.*, 2004).

RNA elements involved in replication of Group I coronaviruses

In an elegant study aimed at determining the minimal replication signals required for the group 1 coronavirus *Transmissible gastroenteritis virus* (TGEV), a two-step amplification system was used that allowed for analysis of DI RNA replication without passaging (Izeta *et al.*, 1999). TGEV-derived DI RNA was generated in the nucleus of transfected cells by cellular RNA polymerase II-mediated transcription, and was further amplified, after export from the nucleus to the cytoplasm, by the viral replicase produced by the helper virus. In this system, the maximal sequences required for DI RNA replication were reduced to the 5'-terminal 1,348 nucleotides (nt) and the 3'-terminal 492 nt of the genome. The 5'-terminal region required for replication was further reduced to

649 nt by showing that an artificial viral mRNA containing the GUS reporter gene was amplified when it contained at least this domain (Table 2-1) (Escors *et al.*, 2003).

Table 2-1. Replication elements in the nidovirus genome

family	subgroup	virus	5'-proximal region	3'-proximal region
Coronaviridae	group I coronavirus	TGEV	649	492
	group II coronavirus	MHV	466	436
		BCoV	498	N.D. (1637)
	group III coronavirus	IBV	544	338
	torovirus	BEV	607	242
Arteriviridae	arterivirus	EAV	296	354
Roniviridae	okavirus		N.D.	N.D.

N.D. = not determined

RNA elements involved in replication of Group II coronaviruses

Mouse hepatitis virus (MHV) is by far the best-studied coronavirus, and its replication has been characterized quite extensively. The minimal *cis*-acting 5'-proximal region required for MHV genome replication was determined in several DI RNA systems. Deletion mapping studies reduced this region to approximately 466 nt (Table 2-1) (de Groot *et al.*, 1992; van der Most *et al.*, 1992; Kim *et al.*, 1993; Masters *et al.*, 1994; Luytjes *et al.*, 1996). Deletions within this domain were not tolerated, suggesting that the entire region is required for replication (Lin & Lai, 1993). It was speculated that this region may contain a RNA superstructure, which may also be involved in a long distance RNA-RNA interaction (Luytjes *et al.*, 1996). Roughly the 5' half of this region consists of the 5' UTR and the other half encodes the N-terminal domain of the replicase. A thermodynamically favorable RNA secondary structure was predicted for the 5' UTR of the MHV genome (Fig. 2-1) (Baric *et al.*, 1987; Wang & Zhang, 2000). An alternative conformation of the central portion of the 5' UTR was predicted in another study (Fig. 2-1) (Shieh *et al.*, 1987). Biochemical support for these structures was not provided and functions were not assigned to each of the proposed structures. *Bovine coronavirus* (BCoV) is related to MHV with a nucleotide sequence identity of around 71%. The 5'-terminal 498 nt of the BCoV genome were found to be required for replication of a DI RNA derivative, a finding that is similar to that for MHV (Table 2-1) (Chang *et al.*, 1994). Computer analysis and structure probing of the first 126 nt of this region identified three hairpin structures that were different from those in the corresponding MHV region (Fig. 2-1) (Chang *et al.*, 1996; Raman *et al.*, 2003). Nucleotide substitution mutations that were designed to disrupt and then restore the conformation of hairpin I and hairpin II appeared

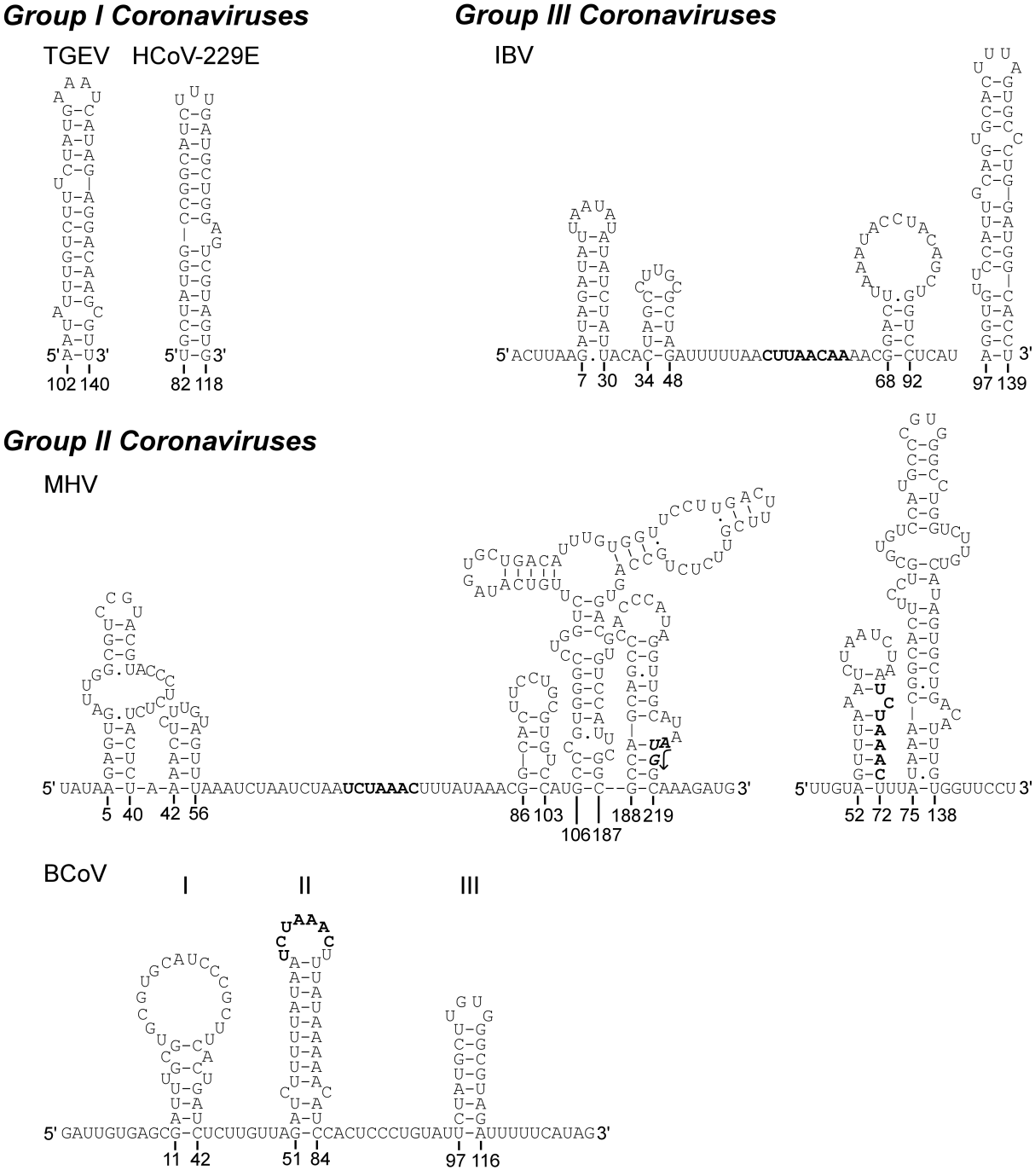
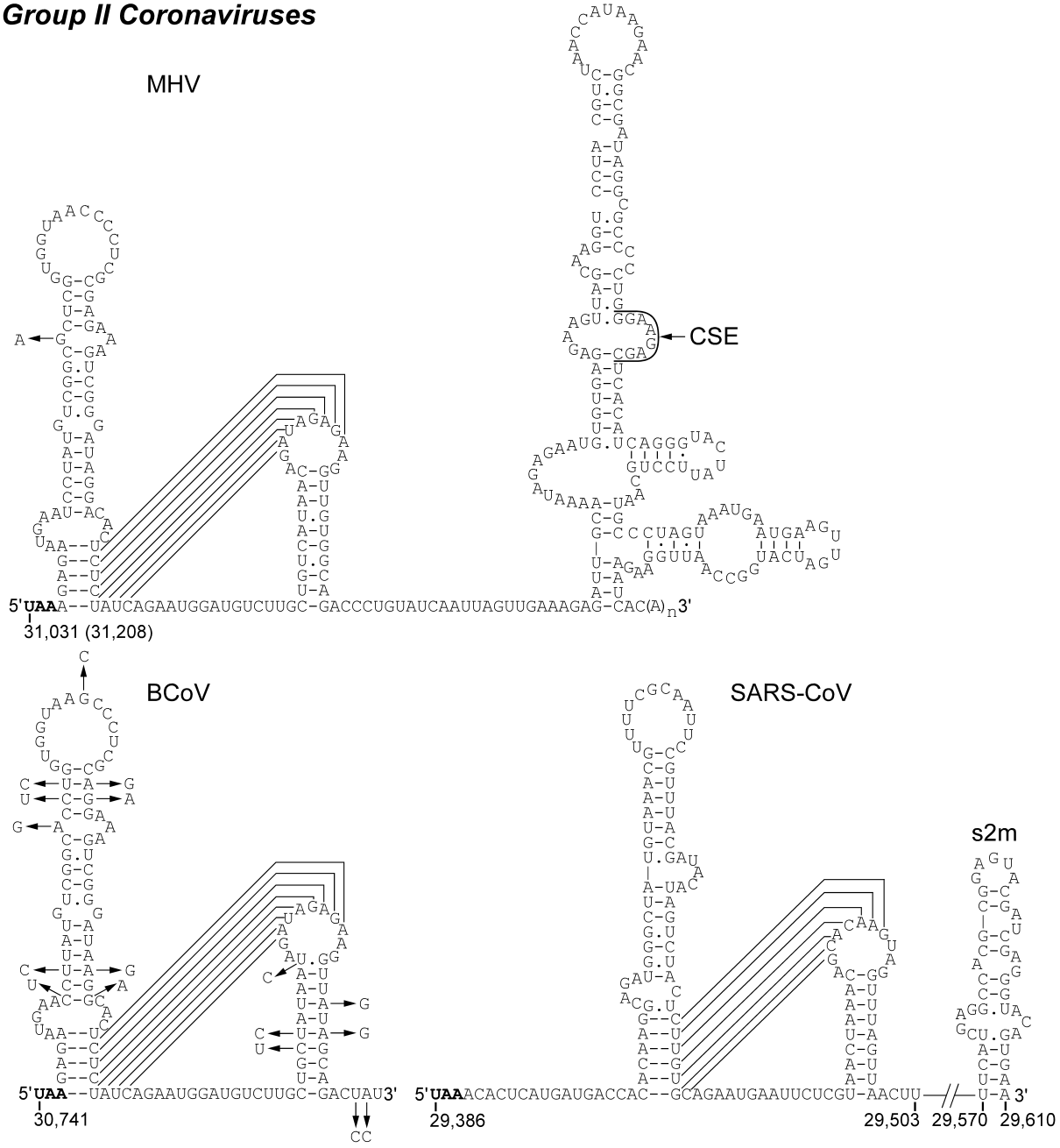


Figure 2-1. RNA secondary structure elements in the 5'-proximal region of the genome of viruses belonging to the Coronavirus genus. *The TGEV (Purdue strain), HCoV-229E and BCoV (Mebus strain) structures, and the most 3' located hairpin of the IBV (Beaudette strain) structure, were adapted from Raman et al. (2003). The rest of the IBV structure was adapted from Stirrups et al. (2000). Two alternative RNA structure conformations are depicted for MHV (JHM strain). The left structure was adapted from Wang & Zhang (2000), whereas the right structure was adapted from Shieh et al. (1987). The replicase translation initiation codon is indicated in bold and italic font and marked with an arrow. The leader TRS is indicated in bold font. Nucleotides are numbered corresponding to their position in the genome.*

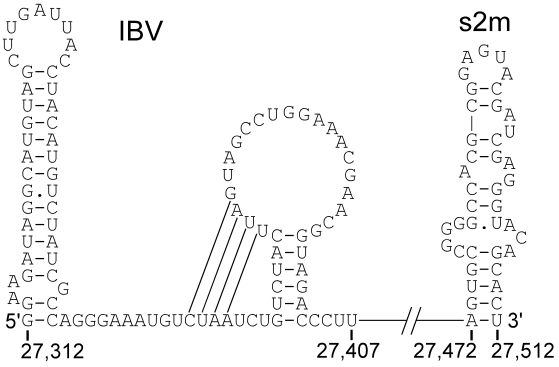
to rapidly revert back to the wild-type sequence, indicating that these mutations have an effect on BCoV DI RNA replication. Since a deletion of hairpin I destroyed the replication ability of a DI RNA, it was postulated that this structure is a *cis*-acting element in BCoV DI RNA replication (Chang *et al.*, 1994; Chang *et al.*, 1996). In another study, the integrity of hairpin III was shown to be required for BCoV DI RNA replication, suggesting that also this structure is an essential *cis*-acting RNA element. Hairpin III appeared to be phylogenetically conserved among group II coronaviruses, which supports its importance in the life cycle of this virus group. Potential hairpin III homologs may also be present in the 5' UTR of the genome of TGEV and *Human coronavirus 229E* (HCoV-229E), group I coronaviruses, and IBV, a group III coronavirus (Fig. 2-1) (Raman *et al.*, 2003).

A combination of studies revealed that the 3'-terminal 436 nt of the MHV genome are required for DI RNA replication (Table 2-1) (de Groot *et al.*, 1992; van der Most *et al.*, 1992; Kim *et al.*, 1993; Lin & Lai, 1993; Masters *et al.*, 1994; van der Most *et al.*, 1995). Initiation of MHV minus strand synthesis requires only the 3'-terminal 55 nt plus the poly(A) tail and no sequence elements from the 5'-proximal region of the genome. This finding suggests that at least a large portion of this region is necessary for initiation of plus strand synthesis (Lin *et al.*, 1994). Although there is no data available about the minimal 3'-proximal region required for BCoV replication, a naturally occurring BCoV DI RNA was described that contains the 3'-terminal 1,637 nt (Table 2-1) (Chang *et al.*, 1994). Several (predicted) RNA structures located within the 3'-terminal half kilobase of the genome of group II coronaviruses were identified as replication mediators. First, a RNA pseudoknot structure was initially identified in the BCoV 3' UTR and its existence was supported by enzymatic structure probing (Fig. 2-2). This structure has a counterpart in all coronavirus subgroups and is highly conserved in location and higher order structure, but not at the sequence level. Mutagenesis experiments demonstrated that this RNA pseudoknot is involved in BCoV replication (Williams *et al.*, 1995; Williams *et al.*, 1999). Secondly, a 68-nt bulged hairpin is located at the 5' end of the 3' UTR directly upstream of the pseudoknot mentioned above and directly downstream of the coding region of the nucleocapsid (N) protein (Fig. 2-2). This structure appeared to be essential for replication. Because the 3'-terminal 55 nt suffice for minus-strand RNA synthesis, and this hairpin is located outside this region, it was suggested that it functions during plus-strand RNA synthesis (Hsue & Masters, 1997; Hsue & Masters, 1999). In a more extensive study to address the role of this hairpin, its presence was supported by biochemical experiments, and a mutagenesis study demonstrated that three of its stem regions could not be disrupted without loss of replication. Because the base of the hairpin partly overlaps with the 5' side of the pseudoknot, the authors speculated that both structures might be involved in a RNA switch (Hsue *et al.*, 2000). Support for this molecular switch came with the generation of a triple mutation involving the base stem region of the hairpin and the interacting loop of the pseudoknot (Goebel *et al.*, 2004a). Finally, a large hairpin was identified in the 3' UTR

Group II Coronaviruses



Group III Coronaviruses



of the MHV genome (Fig. 2-2), and subsequent mutagenesis experiments indicated that this structure has a role in DI RNA replication (Liu *et al.*, 2001). The hairpin contains an octanucleotide motif (GGAAGAGC) that is conserved among coronaviruses and may be functionally important (Fig. 2-2) (Hsue & Masters, 1997; Goebel *et al.*, 2004b). In addition, an 11-nt sequence element positioned in the hairpin was identified approximately 30 nt from the 3' end of the genome. Nucleotide changes in this sequence were not tolerated and affected DI RNA replication (Yu & Leibowitz, 1995). To further characterize this 11-nt sequence, minus-stranded DI RNA transcripts lacking this sequence were synthesized in MHV-infected cells. Since positive-stranded DI RNAs did not accumulate efficiently, it was concluded that the 11 nt sequence was necessary for plus-strand MHV RNA synthesis (Repass & Makino, 1998b).

In an attempt to identify common RNA elements in the BCoV and MHV 3' UTR, the 3' UTR of the MHV genome was successfully replaced with its BCoV counterpart, underlining the close relationship between these viruses (Hsue & Masters, 1997). Moreover, it was demonstrated that MHV could serve as a helper virus for the replication of BCoV DI RNA. Thus, 5'-proximal as well as 3'-proximal *cis*-acting replication elements in BCoV DI RNA can be recognized by the MHV RNA replication apparatus (Wu *et al.*, 2003). With the functional replacement of the MHV genomic 3' UTR by the corresponding RNA element of SARS-CoV, additional biological support was provided for the classification of SARS-CoV in coronavirus group II (Gorbalenya *et al.*, 2004). In contrast, the 3' UTR of the group I coronavirus TGEV and the group III coronavirus *Infectious bronchitis virus* (IBV) could not functionally replace the MHV 3' UTR. Indeed, RNA secondary structure analysis of the SARS-CoV genomic 3' UTR identified a hairpin structure that overlaps with a pseudoknot (Fig. 2-2) and is similar to the structures discovered in the corresponding region of the MHV and BCoV genomes (Goebel *et al.*, 2004b). These findings are in agreement with recent phylogenetic studies on the basis of the most conserved part of the replicase gene, which placed SARS-CoV closest to the group II coronaviruses (Snijder *et al.*, 2003). The classification of SARS-CoV is of importance, because a good SARS-CoV model system is very useful for molecular biology and biosafety studies. An attempt to replace the 5' UTR of the SARS-CoV genome with that of a BCoV DI RNA was unsuccessful, even when this DI RNA was supplied with both the 5' and 3' UTR of the SARS-CoV genome. It was suggested that the SARS-CoV 5' UTR is more group I like, underlining the complexity of the SARS-CoV genome and its classification (Nixon *et al.*, 2004).

Figure 2-2. RNA secondary structure elements in the 3'-proximal region of the genome of viruses belonging to the Coronavirus genus. *The MHV (JHM strain) structure was combined and adapted from Liu et al. (2001), and Goebel et al. (2004a). The single nucleotide variation with the MHV A59 strain is indicated. Also, the conserved sequence element (CSE) is indicated. The BCoV (Mebus strain) and SARS-CoV (Urbani strain) RNA structures are adapted from Williams et al. (1999) and Goebel et al. (2004a), respectively. Sequence variation among BCoV strains is indicated. The IBV (strain Beaudette) and SARS-CoV s2m hairpins are taken from Robertson et al. (2005). The most 5' positioned hairpin of the IBV structure and the pseudoknot were adapted from Dalton et al. (2001) and Williams et al. (1999), respectively. The translation stop codon (UAA) of the N protein ORF is indicated in bold face.*

Besides replication elements located at the genome termini, a MHV-JHM derived DI RNA required a 135-nt internal replication signal for its amplification by a MHV-A59 helper virus (Kim *et al.*, 1993; Lin & Lai, 1993). A 55-nt hairpin structure, located around 3.2 kb from the 5' end of the MHV genome, appeared to be involved, and its structure was supported by probing experiments (Fig. 2-3). Mutagenesis studies of this structure suggested that it exerted its replication function in the plus polarity only (Kim & Makino, 1995; Repass & Makino, 1998a). In contrast, for the closely related MHV-A59 strain, a similar internal replication signal was not required for replication of a DI RNA derivative (van der Most *et al.*, 1991; de Groot *et al.*, 1992; Masters *et al.*, 1994; Luytjes *et al.*, 1996). It was suggested that the cell types used might explain the observed difference (Luytjes *et al.*, 1996).

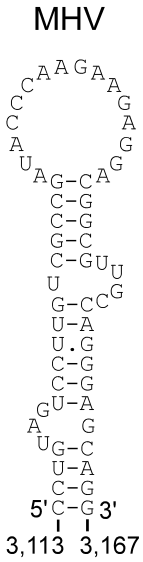


Figure 2-3. RNA secondary structure of the 'internal' replication signal of a DI RNA genome derived from the MHV JHM strain. *Adapted from Kim & Makino (1995).*

RNA elements involved in replication of Group III coronaviruses

The minimal sequence requirements for replication of the group III coronavirus *Avian infectious bronchitis virus* (IBV) were determined using a modified DI RNA that contained the CAT gene under control of a body TRS (DI-61CAT). Using a CAT ELISA, protein expression from the DI RNA could be detected in transfected cells. The results indicated that the 5'-terminal 544 nt and 3'-terminal 338 nt contained the necessary signals for DI RNA replication (Table 2-1). Secondary structure analysis of the first 100 nt of this 5'-terminal region of the IBV genome identified three hairpin structures (Fig. 2-1). The most 5'-proximal hairpin showed a high degree of covariance amongst the IBV strains providing phylogenetic support for the importance of this structure (Stirrups *et al.*, 2000). Within its 3'-proximal region, which is part of the 3' UTR, a hairpin was identified that is highly conserved among group III coronaviruses. Moreover, deletion of this structure abolished DI RNA replication, suggesting that it has a function in replication (Dalton *et al.*, 2001). This coronavirus group III-specific hairpin is located 10 nt upstream of the pseudoknot structure that is conserved throughout the coronavirus genus (Fig. 2-2). However, the IBV pseudoknot is structurally not very convincing and displays only little resemblance with the group I and II coronavirus structures (Williams *et al.*, 1999). It has to be noted that group III coronaviruses contain a conserved so-called "s2m motif" in the 3' UTR of their genomes, which is also present in picorna- and astroviruses (Jonassen *et al.*, 1998; Jonassen *et al.*, 2005). Remarkably, this RNA element is also found in the 3' UTR of SARS-CoV, which is considered to occupy a position distant from group III. The s2m motifs of both IBV and SARS-CoV may adopt similar hairpin structures (Fig. 2.2) (Robertson *et al.*, 2005). Recently, also the three-dimensional structure of the SARS-CoV s2m motif was determined by NMR studies (Robertson *et al.*, 2005).

RNA elements involved in replication of Toroviruses

Two DI RNA genomes of *Berne virus* (BEV) of approximately 1.0 and 1.4 kb were isolated and characterized after serial undiluted passaging of this virus in cell culture. The smallest DI RNA contained the 5'-terminal 607 nt fused to the 3'-terminal 242 nt (Table 2-1) (Snijder *et al.*, 1991). An analysis of the (predicted) higher-order RNA structures in these sequences has not yet been published.

Protein-binding elements and replication

It is well-established that replication of viral genomes involves the recognition of a RNA template by the replicase, which includes the RNA-dependent RNA polymerase

(RdRp) subunit in addition to other viral or host proteins. Many of the above-mentioned RNA primary and secondary structures that were shown to be involved in coronavirus replication may be required to recruit the RdRp complex. Specific and functional interactions of proteins with protein binding elements (PBEs) in the RNA template are an interesting and extensively studied topic in virus research. Studies on PBEs and their associated proteins have been limited mainly to group II coronaviruses with a focus on the 3' UTR of their genomes.

In order to identify host proteins interacting with the BCoV 3' UTR, gel mobility shift assays were performed, and specific interactions were observed. The proteins responsible for this interaction were subsequently studied in UV cross-linking experiments and they appeared to have molecular masses of 99, 95, and 73 kDa, with the latter being Poly(A) Binding Protein (PABP). Competition assays with the MHV 3' UTR demonstrated that the interactions are conserved in these two viruses. Subsequently, the requirement for the poly(A) tail in replication was assessed. A shortened poly(A) tail reduced the interaction with PABP, and this correlated with a decreased replication competence of both a MHV and a BCoV DI RNA, indicating that the poly(A) tail is an important *cis*-acting signal for coronavirus replication (Spagnolo & Hogue, 2000). In another study, several host proteins were identified as binding partners of a PBE located in the 3'-terminal 42 nt of the MHV genome. Some evidence was obtained that these proteins, m-acotinase, HSP40, 60 and 70, form a stable RNA-protein complex (Nanda *et al.*, 2004). However, only for m-acotinase a correlation was found between its expression level and virus growth (Nanda & Leibowitz, 2001). Another PBE was identified in the 3' UTR of the MHV genome (nt 129 to 166 from the 3' end) and it interacted with several proteins. Mutations in this region that dramatically reduced protein binding were not recovered in DI RNA, indicating that these mutations strongly affected replication and were repaired by RNA recombination with the helper virus genome (Liu *et al.*, 1997). Furthermore, a strong binding site for heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) was identified in a region 90 to 170 nt from the 3' end of the MHV genome. A second, weak hnRNP A1 binding site was mapped between nt 260 to 350 from the genomic 3' end. These binding sites are complementary to sites on the minus-strand RNA that bind another cellular protein, polypyrimidine tract binding protein (PTB). Mutations in the RNA sequence that reduced PTB-binding to the minus strand also affected hnRNP A1-binding to the plus strand, indicating a possible relationship between these two binding events. DI RNAs containing a mutated hnRNP A1 binding site have reduced transcription and replication activities (Lai, 1998; Huang & Lai, 2001).

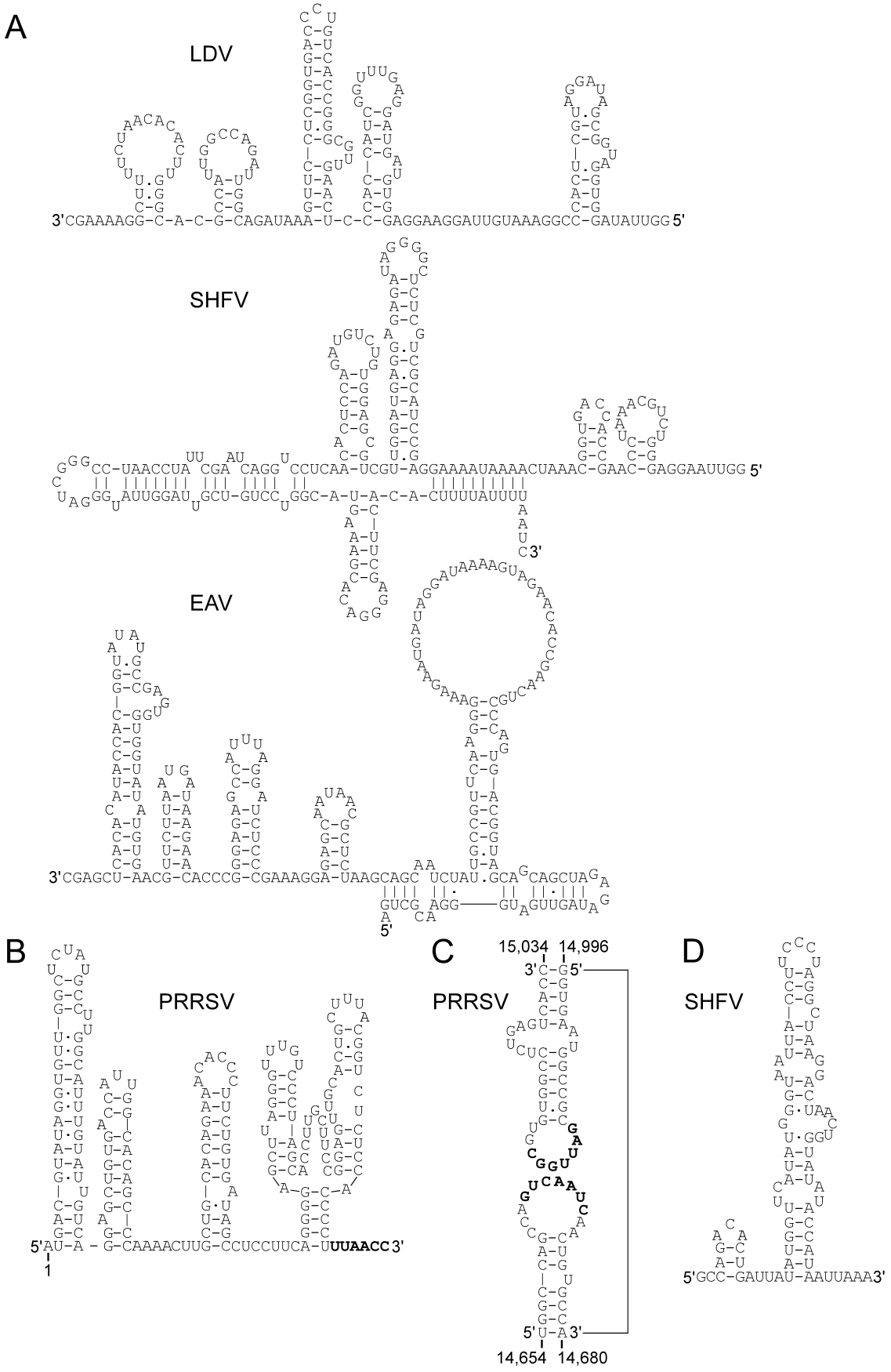
The 5'-proximal region of the MHV genome interacts with hnRNP A1 and the complementary strand with PTB. It was demonstrated that *in vitro* both proteins are able to form a RNP complex with RNA fragments representing the 5'- and 3'-proximal regions of the MHV genome. These studies suggest that an interaction between hnRNP A1 and PTB may provide a molecular mechanism for cross-talk between the 5' and 3' ends of the MHV RNA, which may be important for translation, replication and/or transcription (Lai,

1998; Huang & Lai, 2001). The finding that overexpression of hnRNP A1 accelerated the kinetics of viral RNA synthesis appeared to provide a biological basis for the importance of hnRNP A1 in viral synthesis. This support was strengthened by showing that a dominant negative hnRNP A1 mutant inhibited viral RNA synthesis (Shi *et al.*, 2000). However, conflicting information was derived from a study using cells lacking hnRNP A1, in which MHV replication was not affected at all, indicating that the protein is nonessential for viral RNA synthesis. Whether redundancy in hnRNP A1 protein function is responsible for this observation was not evaluated (Shen & Masters, 2001).

Replication signals in the RNA genome of *Arteriviridae*

Although reverse genetics systems for the full-length genomes of the arteriviruses *Equine arteritis virus* (EAV) and *Porcine reproductive and respiratory syndrome virus* (PRRSV) were developed prior to those for coronaviruses (van Dinten *et al.*, 1997; Meulenberg *et al.*, 1998), the functional dissection of arterivirus genome replication lags behind that of coronaviruses.

A number of studies were carried out that may shed initial light on the signals required for EAV genome replication. A DI RNA was obtained by serial undiluted passaging of EAV in cell culture and its sequence analysis revealed that the 5'-terminal 1057 nt and 3'-terminal 4,175 nt were present. Subsequently, a deletion analysis delimited the sequences required for DI RNA replication to the 5'-terminal 589 nt and the 3'-terminal 1,066 nt. In addition to these 5'- and 3'-terminal sequences, a 583-nt sequence from the central part of replicase ORF1b (nt 8566 to 9149 in the EAV genome) was found to be essential for DI RNA replication (Molenkamp *et al.*, 2000b). Curiously, full-length genome replication required only the 3'-terminal 354 nt, which is a shorter region than the above mentioned 3'-terminal 1,066 nt required for DI RNA replication (Table 2-1) (Molenkamp *et al.*, 2000c). The 5'-terminal sequences were delimited to 296 nt by showing that an EAV RNA lacking nt 297 to 1004 was able to replicate to wild-type levels (Table 2-1) (Tijms *et al.*, 2001). An unequivocal RNA secondary structure could not be obtained for the 5'-proximal two-third of this region (Kheyar *et al.*, 1998). Several proteins were identified that interact *in vitro* with the first 206 nt of the 5'-proximal domain and its complement, but the functional relevance of these interactions was not addressed (Archambault *et al.*, 2005). In another study, the complement of the same region was found to bind several cellular proteins (Hwang & Brinton, 1998). Four of these proteins also interacted with the corresponding regions of the *Simian hemorrhagic fever virus* (SHFV) and *Lactate dehydrogenase-elevating virus* (LDV) genomes. However, RNA secondary structure predictions for this region revealed the absence of conservation between these viruses (Fig. 2-4A) (Hwang & Brinton, 1998). In an attempt to study the significance of the PRRSV 5'-proximal 190 nt, its RNA secondary structure was predicted


A
LDV
3'CGAAAAGG-C-A-C-G-CAGAUAAA-U-C-C-GAGGAAGGAUUGUAAAGGCC-GAUAUUGG 5'
SHFV
EAV
3'CGAGCU-AACG-CACCCG-CGAAAGGA-UAAGCAGC UCUAU.GCA CAGCUA A
5'
B
PRRSV
5'AU-A - G-CAAAACUUG-CCUCCUUCA-UUUAACC 3'
1
C
PRRSV
15,034 14,996
3'C-G 5'
5'U-A 3'
14,654 14,680
D
SHFV
5'GCC-GAUUAU-AAUUAAA 3'

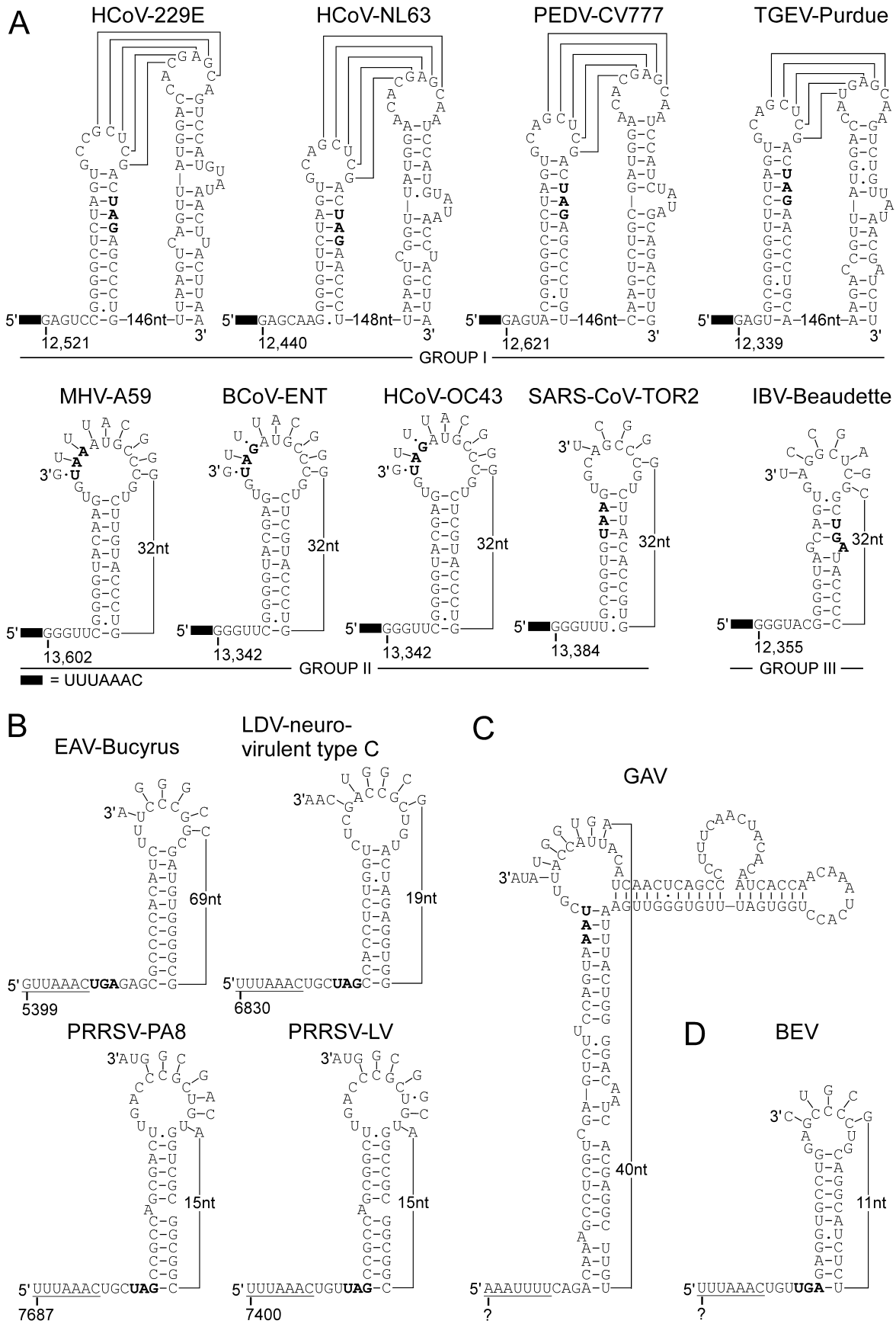
by computer analysis (Fig. 2-4B). Although the authors recognized this region as an important regulatory element in replication and transcription, a function was not attributed to any of the structures identified (Tan *et al.*, 2001). In a more interesting study, the presence of essential replication elements in the structural protein-coding region of the PRRSV genome was investigated. Deletion analysis showed that a stretch of 34 nucleotides within ORF7, which encodes the nucleocapsid protein, is essential for replication. The 34-nucleotide stretch is highly conserved among PRRSV isolates and is predicted to fold into a hairpin. Sequences within the loop of this structure were shown to base-pair with a sequence present in the loop of a hairpin located in the 3' noncoding region, resulting in a so-called kissing interaction (Fig. 2-4C). It was confirmed that this kissing interaction is required for replication (Verheije *et al.*, 2002). The 3'-terminal 79 nt of the SHFV genome were predicted to fold into two hairpin structures, which were supported by structure probing (Fig. 2-4D). Two cellular proteins, PTB and fructose biphosphate aldolase A, bound specifically to a probe containing this 79-nt region of the SHFV genome (Maines *et al.*, 2005).

The ribosomal frameshift-regulating RNA pseudoknot

Nidoviruses produce two large replicase polyproteins, designated pp1a and pp1ab. Translation of ORF1a terminates at a translation stop codon in a region where ORF1a and ORF1b briefly overlap. Translation can either be terminated at the ORF1a stop codon or continued following a –1 ribosomal frameshift just upstream of this termination site. The C-terminally extended pp1ab polyprotein is expressed to a specific, lower level relative to pp1a (15-40%, as estimated from expression studies). Since the ORF1b-encoded protein encodes key replicative domains like the RdRp and helicase, the conserved frameshift mechanism used to down-regulate their expression must be of crucial importance for the regulation of the nidovirus life cycle.

Eukaryotic ribosomal frameshift signals generally contain two element, a heptanucleotide slippery sequence (XXXYYYN) and a frameshift-promoting RNA secondary structure, often an RNA pseudoknot, located 5-9 nt downstream. In general, the slippery sequence consist of triplets of A, U or G residues, followed by the tetranucleotide UUUA, UUUU or AAAC (Jacks *et al.*, 1988; ten Dam *et al.*, 1990; Brierley *et al.*, 1992; Giedroc *et al.*, 2000). This type of –1 ribosomal frameshift signals have been identified in

Figure 2-4. RNA secondary structure elements in the arterivirus. *(A) Depicted are the RNA structures located in the antileader region of the anti-genome of three arteriviruses, EAV (Bucyrus strain), SHFV (LVR 42-0/M6941 strain), and LDV (neuro-virulent type C strain), taken from Hwang & Brinton (1998); (B) RNA secondary structure model of the PRRSV (VR-2332 strain) leader region, taken from Tan et al. (2001). Leader TRS is indicated in bold face; (C) Model of the kissing loop interaction identified in the 3'-proximal region of the PRRSV (Lelystad strain) genome, adapted from Verheije et al. (2002); (D) Two hairpin structures are predicted in the 3'-terminal 79 nt of the SHFV genome. Adapted from Maines et al. (2005).*


A
HCoV-229E
HCoV-NL63
PEDV-CV777
TGEV-Purdue
5' GAGUCC-G–146nt–U-A 3'
12,521
5' GAGCAAG.U–148nt–U-A 3'
12,440
5' GAGUA-U–146nt–C-G 3'
12,621
5' GAGU-A–146nt–A-U 3'
12,339
GROUP I
MHV-A59
BCoV-ENT
HCoV-OC43
SARS-CoV-TOR2
IBV-Beaudette
32nt
5' GGGUUC-G
13,602
13,342
13,342
5' GGGUUU.G
13,384
5' GGGUACG-C
12,355
GROUP II
GROUP III
= UUUAAAC
B
EAV-Bucyrus
LDV-neuro-virulent type C
69nt
5' GUUAAACUGAGAGC-G
5399
19nt
5' UUUAAACUGCUAGC-G
6830
PRRSV-PA8
PRRSV-LV
15nt
5' UUUAAACUGCUAG-C
7687
5' UUUAAACUGUUAG-C
7400
C
GAV
40nt
5' AAAUUUUCAGA-U
?
D
BEV
11nt
5' UUUAAACUGUUGA-U
?

a number of virus groups, such as retroviruses (Jacks & Varmus, 1985; Chamorro *et al.*, 1992), double-stranded RNA viruses of yeast (Dinman *et al.*, 1991; Tzeng *et al.*, 1992; Lopinski *et al.*, 2000), astroviruses (Marczinke *et al.*, 1994; Lewis & Matsui, 1996), and luteoviruses (Prufer *et al.*, 1992; Brault & Miller, 1992).

The RNA elements responsible for ribosomal frameshifting during nidovirus genome translation were first identified for IBV (Brierley *et al.*, 1987; Brierley *et al.*, 1989). Its frameshift signal consists of a heptanucleotide UUUAAAC stretch, which is highly conserved in coronaviruses, followed by an RNA pseudoknot (Fig. 2-5A). It was found that the primary structure of this pseudoknot apparently is not a determinant in the frameshifting mechanism. As long as the overall structure is maintained, frameshifting is highly efficient (Brierley *et al.*, 1991; Napthine *et al.*, 1999). When the IBV pseudoknot was placed in the middle of an mRNA coding region, a translational intermediate of the expected size appeared, indicating that ribosomal pausing was induced by this structure. However, ribosomal pausing by itself does not suffice and additional events are required, because a hairpin with the same energetic stability as the pseudoknot was also able to stall ribosomes without inducing a frameshift. Thus, the pseudoknot must have certain properties important for its role in frameshifting (Somogyi *et al.*, 1993; Kontos *et al.*, 2001). It was suggested that the coaxially stacked stems 1 and 2 are kinked, which may be such a functional determinant (Liphardt *et al.*, 1999).

Frameshift-inducing pseudoknots similar to that of IBV were identified in several other coronaviruses (Fig. 2-5A) (Bredenbeek *et al.*, 1990; Lee *et al.*, 1991; Thiel *et al.*, 2003; Ramos *et al.*, 2004; Baranov *et al.*, 2005). There are some indications that a RNA hairpin can be formed within the loop that connects the two stems of the pseudoknot, and that this property may contribute to the frameshifting signal (Ramos *et al.*, 2004; Plant *et al.*, 2005). Although, the group I coronavirus frameshift signal may consist of a H-type pseudoknot structure like that of group II coronaviruses, including the presence of a hairpin structure in the connecting loop (Ramos *et al.*, 2004), it was proposed that group I coronavirus genomes possess an “elaborated pseudoknot" which is required for high frequency frameshifting (Herold & Siddell, 1993; Eleouet *et al.*, 1995; Kocherhans *et al.*, 2001; Baranov *et al.*, 2005). In fact, this type of "pseudoknot" involves a kissing loop interaction between two hairpins (Fig. 2-5A). A slippery sequence and downstream pseudoknot was also identified in the genome of the torovirus BEV, and was found to be

Figure 2-5. RNA secondary structure models of nidovirus replicase frameshifting signals. *(A) Frameshift-inducing pseudoknots of the Coronavirus genus taken from Baranov et al. (2005). The slippery sequence is indicated with a black bar.; (B) Frameshift-inducing pseudoknots of the Arterivirus genus. Depicted are the RNA structure of EAV, adapted from Den Boon et al. (1991), LDV, taken from Godeny et al. (1993), the North-American PRSSV isolate PA8, adapted from Wootton et al. (2000), and the European PRRSV isolate LV, adapted from Meulenberg et al. (1993a). Slippery sequences are underlined; (C) Frameshift pseudoknot of the Ronivirus genus member Gill-associated virus (GAV), taken from Cowley et al. (2000); (D) Frameshift pseudoknot of the Torovirus genus member BEV, taken from Snijder et al. (1990a).*

similar to that of group II and III coronaviruses, but with a relatively short connecting loop (Fig. 2-5D) (Snijder *et al.*, 1990a). It has to be noted, that in a proposed alternative structure the connecting loop of the pseudoknot may adopt a hairpin conformation (Snijder *et al.*, 1990a), which is reminiscent of the structure postulated for coronaviruses (Plant *et al.*, 2005).

Ribosomal frameshift signals similar to those of group II and III coronaviruses were predicted in arterivirus genomes (Fig. 2-5B) (den Boon *et al.*, 1991; Godeny *et al.*, 1993; Meulenberg *et al.*, 1993a; den Boon, 1996; Wootton *et al.*, 2000). Ronivirus genomes may contain a more complex pseudoknot structure downstream of their AAAUUUU heptanucleotide sequence compared to that of other nidoviruses (Fig. 2-5C) (Cowley *et al.*, 2000).

Internal ribosome entry site in nidoviruses

Different virus groups employ an internal ribosome entry site (IRES) element for the translation initiation of their replicase ORF, sometimes as part of the strategy to specifically promote translation of viral over cellular mRNAs (Vagner *et al.*, 2001). Most likely, this is not the case for the nidovirus replicase, which is expressed via cap-dependent translation initiation and ribosome scanning (Lai & Stohlman, 1981; Lai *et al.*, 1982a; Sagripanti *et al.*, 1986; van Vliet *et al.*, 2002) and this thesis). On the other hand, there is some evidence that internal ribosomal entry occurs on some of the coronavirus subgenomic (sg) mRNAs to express downstream ORFs. In theory, ORFs located downstream of the replicase gene could then also be expressed from the genomic RNA, making the process independent from sg RNA synthesis. In addition, every larger mRNA having this ORF and putative IRES element may be translated to express such a gene.

The first indications for a nidovirus IRES came from work on sg mRNA 3 of IBV, which is functionally tricistronic, having the capacity to encode three small proteins (3a, 3b and 3c) from separate ORFs. The 3c product, the structural protein E, is the most abundant of the three, which makes it unlikely that leaky scanning is involved in its production (Liu *et al.*, 1991). It was indeed confirmed that ORF 3c translation is initiated via a cap-independent mechanism (Liu & Inglis, 1992). A thorough RNA secondary structure analysis on the 3a and 3b coding region upstream of the 3c ORF identified five hairpin structures. Some tertiary (pseudoknot) interactions contribute to the formation of a compact folded structure (Fig. 2-6A). It was suggested that these proposed structures share features with the IRES elements in picornavirus genomes. Immediately upstream of the initiator AUG of the 3c ORF, sequences complementary to the 18S rRNA were identified, similar to those in picornaviruses (Le *et al.*, 1994). MHV sg mRNA 5 has the capacity to encode two proteins from ORF 5a (encoding an accessory protein of unknown function) and ORF 5b (encoding the envelope protein E), which overlap by five nucleotides and are in different reading frames. Following sequence analysis of the MHV

genome, it was already suggested that ORF5b is translated from mRNA5 by a mechanism involving internal initiation of protein synthesis (Skinner *et al.*, 1985). Although, this putative IRES within mRNA 5 was further characterized and the boundaries of this element were determined, a RNA secondary structure model was not proposed (Thiel & Siddell, 1994; Jendrach *et al.*, 1999). It should be noted that an additional sg mRNA, designated sg mRNA5-1, of which the synthesis is driven by a noncanonical body TRS, may be responsible for ORF5b translation (Zhang & Liu, 2000). The translation of ORF 3b of TGEV is probably also mediated by an internal entry mechanism. Upstream of this ORF several RNA pseudoknot structures were identified (Fig. 2-6B), however, upon a deletion analysis most of the structures could be removed without profound effects on ORF 3b translation (O'Connor & Brian, 2000).

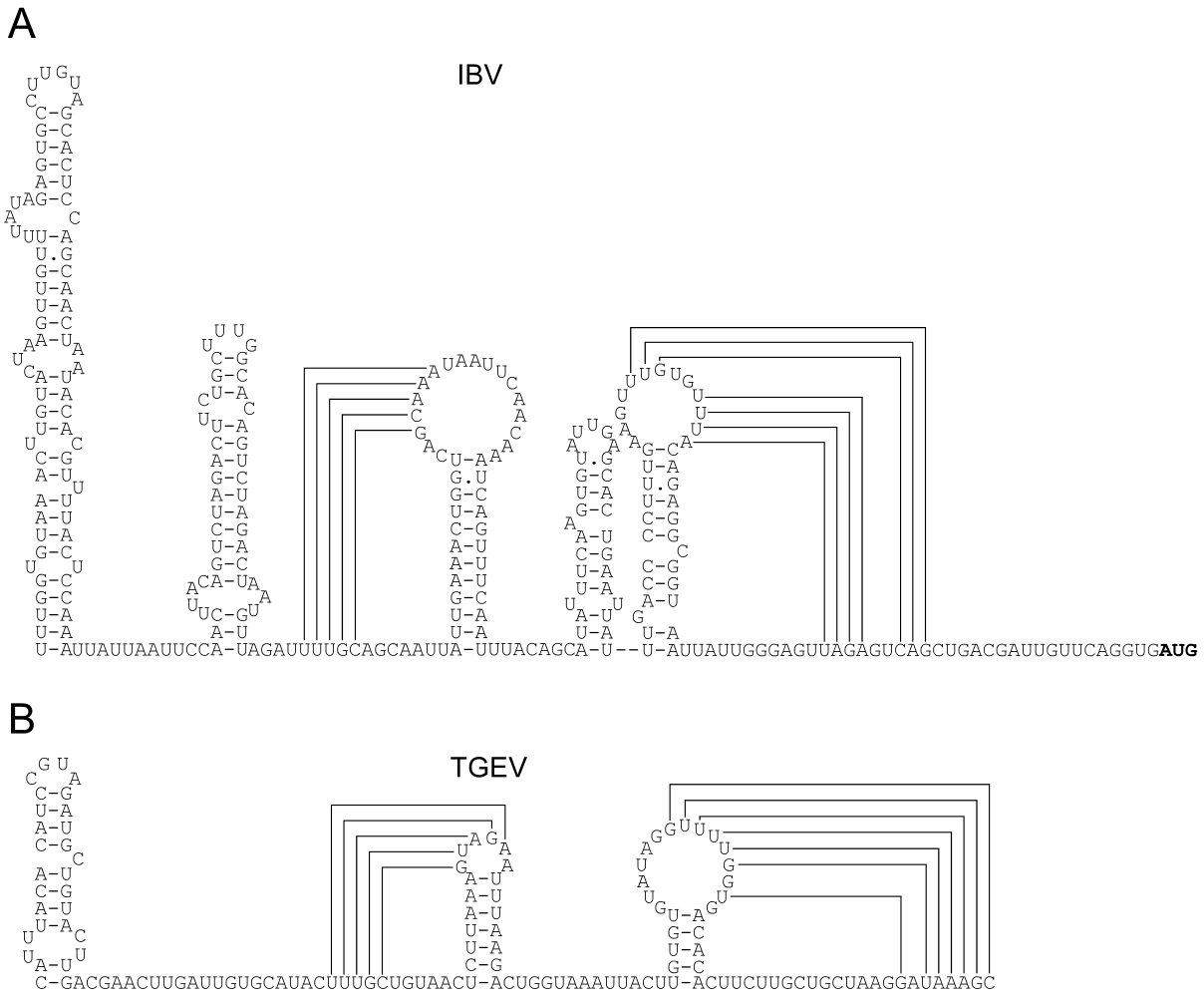


Figure 2-6. Putative IRES elements in the genome of members of the Coronavirus genus. *RNA secondary structure of the putative IRES element in (A) sg mRNA3 of IBV (Portugal/322/82 strain), adapted from Le et al. (1994), and (B) sg mRNA3 of TGEV (Purdue strain), adapted from O'Connor & Brian (2000).*

It was demonstrated that the EAV structural proteins E and G_s are both translated from two overlapping ORFs, ORF 2a and ORF 2b, on sg mRNA 2. However, no details on the mechanism of expression of the downstream ORF were unravelled. It was suggested that the homologous genes at corresponding positions in the genome of the three other arteriviruses, PRRSV, LDV, and SHFV, are also expressed from functionally bicistronic (and structurally polycistronic) sg mRNAs (Snijder *et al.*, 1999). Furthermore, it was hypothesized that sg mRNA 2 of SHFV is another putative functionally bicistronic mRNA, directing the synthesis of both the ORF 2a and ORF 2b products (Godeny *et al.*, 1998).

Packaging

After its successful amplification, the nidovirus genome has to be encapsidated into a new virus particle (virion). The virion protects the genome from environmental influences and transports it safely to another susceptible host. In addition, the virion mediates entry into the cell after successful binding to receptors on the cell surface.

The process of encapsidation starts with the formation of a nucleocapsid structure that consists of the viral RNA and a dedicated nucleocapsid (N) protein. For a number of plus-stranded RNA viruses a specific packaging signal (PS) in the genome was identified. Ultimately, when such a PS would be all that is needed to package an RNA, it should be able to direct the encapsidation of heterologous RNAs carrying the PS, as was indeed demonstrated for a number of viruses (Adam & Miller, 1988; Hayashi *et al.*, 1992; Woo *et al.*, 1997; Cologna & Hogue, 2000; Choi & Rao, 2003). Apart from the sequence, the RNA secondary structure of the PS plays a major role. Nucleotide-specific contacts between protein and RNA generally involve nucleotides in single-stranded loops and bulges of RNA structures (Varani & Pardi, 1994). With regard to genome packaging of nidoviruses, RNA encapsidation signals in coronavirus genomes have been identified and studied in some detail, while our knowledge concerning arterivirus encapsidation is still very limited.

A putative packaging signal was mapped near the 3' end of the second replicase ORF (ORF1b) in the genome of MHV, a group 2 coronavirus (Makino *et al.*, 1990; van der Most *et al.*, 1991). Subsequent deletion analysis of this region using a MHV DI RNA replicon system, identified a 190-nt element necessary for DI RNA packaging (genome position 20.2-20.4 kb). A computer-based structure prediction of the corresponding sequence in the genomes of the JHM and A59 strains of MHV revealed the presence of a common stable hairpin of 69 nt within this region (Fig. 2-7). Mutagenesis showed that the minimum PS consisted of 61 nt and that the integrity of its RNA structure was important for packaging (Fosmire *et al.*, 1992). However, the 69-nt PS displayed an approximately five-fold lower packaging efficiency than the 190-nt element. Furthermore, *in vitro* binding studies demonstrated that the ORF1b sequences that flank the 69-nt element were

required for efficient interaction with the N protein, whereas the 69-nt PS alone was not sufficient for such an interaction. It was suggested that these flanking sequences are necessary to force the 69-nt hairpin into a specific structural conformation (Molenkamp & Spaan, 1997; Narayanan & Makino, 2001). Nevertheless, the insertion of the 69-nt element in MHV DI RNAs that were not packaged or into non-MHV RNAs was sufficient to incorporate these molecules into MHV particles, albeit sometimes with low efficiency (Lin & Lai, 1993; Woo *et al.*, 1997). A MHV-A59 sg mRNA transcript could also be packaged when it contained the PS, but the sg mRNA was packaged approximately six times less efficient than the DI RNA from which it was produced, again indicating that additional factors determine packaging efficiency (Bos *et al.*, 1997). Subsequently, near the 3' end of the replicase ORF, a region homologous to that of MHV was identified in the genome of BCoV, which appeared to be conserved in terms of both sequence and function. Two alternative RNA secondary structures were predicted for this region (Fig. 2-7). RNAs were incorporated into virions when this region was appended to a nonrelated RNA (Cologna & Hogue, 2000). Despite the fact that the above mentioned PS elements of BCoV and MHV were shown to be important for packaging, several MHV and BCoV DI RNAs lacking this ORF1b-derived sequence were found to be packaged (Makino *et al.*, 1988a; Makino *et al.*, 1988b; Chang *et al.*, 1994; Chang & Brian, 1996). This may indicate that several regions in the DI RNA genome contribute to packaging.

The 5' UTR and/or the 3'-proximal 293 nt of the 3' UTR of a IBV DI RNA genome were specifically required for its packaging, although rescue of DI RNA-containing virions was poor. Replicase ORF1b sequences, but not from a specific part, enhanced the packaging efficiency, suggesting that small RNAs are not packaged due to size restrictions (Dalton *et al.*, 2001). A similar finding was made for the group I coronavirus TGEV. Besides the genome termini, DI RNAs required either one of two regions of the replicase ORF for optimal packaging. It was suggested that multiple PSs might contribute to efficient packaging. Alternatively, the unnatural arrangement of RNA sequences in DI RNAs may affect higher order RNA structures, thereby influencing packaging (Izeta *et al.*, 1999). In another study, a PS was identified in the 5'-terminal 649 nt of the TGEV genome, and this region inserted into a sg mRNA led to its specific encapsidation (Escors *et al.*, 2003).

It has to be noted that in some reports sg mRNA packaging was described (Sethna *et al.*, 1989; Hofmann *et al.*, 1990; Zhao *et al.*, 1993; Cologna & Hogue, 2000), but the presence of these sg mRNAs in virion preparations may have been due to contamination of virion preparations with mRNAs from infected cells (Escors *et al.*, 2003).

The mechanism of specific encapsidation of the coronavirus genome into virions and the role of the structural proteins is poorly understood. *In vitro* studies demonstrated that the MHV N protein interacted with a small PS-containing RNA and that this interaction depended on the presence of the 69-nt hairpin and flanking ORF1b-derived sequences (Molenkamp & Spaan, 1997). Multiple sites that contributed to N protein

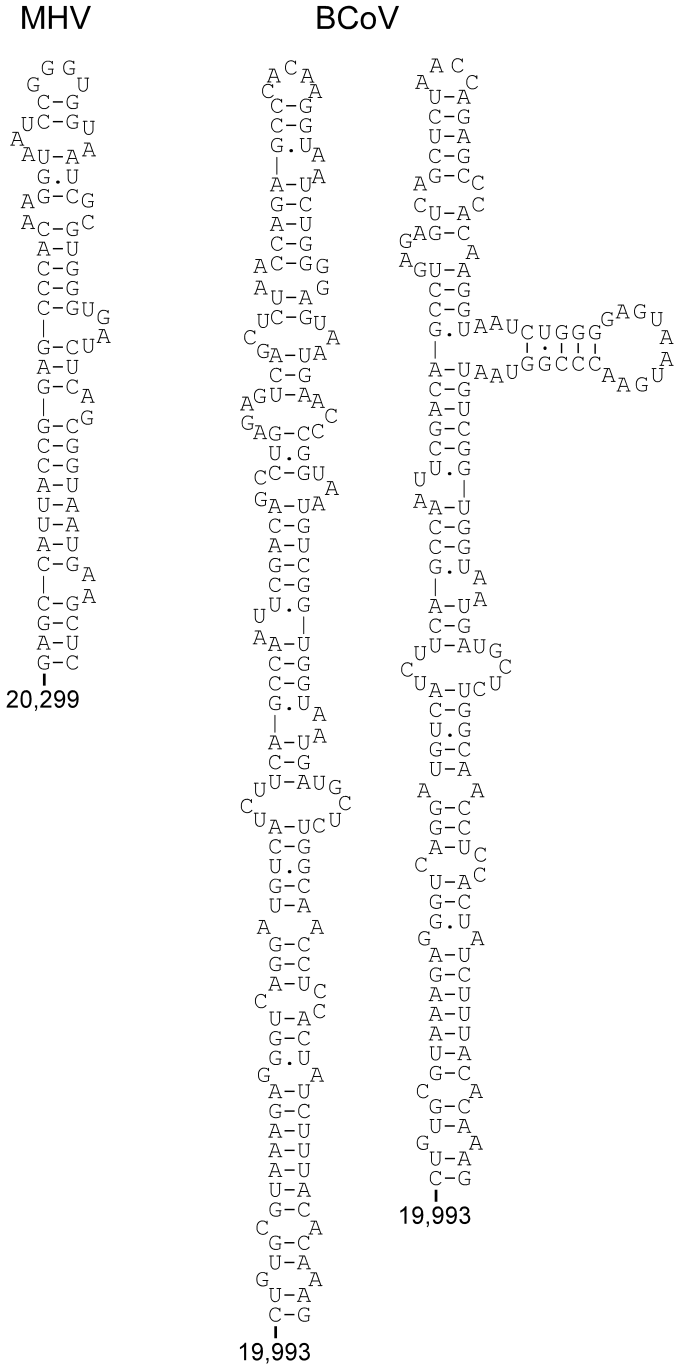


Figure 2-7. *RNA packaging signals of two group II coronaviruses. The structure of MHV (A59 strain) was taken from Fosmire et al. (1992), and two alternative structure conformations of BCoV (Mebus strain) were taken from Cologna & Hogue (2000).*

binding were found in MHV and BCoV DI RNA genomes. One high-efficiency N-binding site is located in the 3' region of the N ORF, but it was not sufficient to effect packaging of a nonviral RNA, indicating that this region may act as an enhancer element only (Cologna *et al.*, 2000). A model for MHV RNA packaging was proposed based on data from several studies ((Narayanan *et al.*, 2003) and references therein). After the N protein specifically binds to the viral genome to form a ribonucleoprotein (RNP) complex, the membrane protein M binds to the PS present in the genomic RNA (or the PS already in complex with N). Thus M binding determines the selective packaging of genomic RNA and excludes the packaging of sg mRNAs, which lack the PS. Subsequently, the M protein-RNP complex engages in budding in concert with the E protein.

Regulatory elements involved in sg mRNA synthesis

The synthesis of one or multiple sg mRNAs is a mechanism that many positive-strand RNA viruses have evolved to selectively express structural and accessory proteins from ORFs other than those encoding the major replicase subunits (Miller & Koev, 2000). Three different mechanisms have been proposed to explain sg mRNA production by positive-stranded RNA viruses (Miller & Koev, 2000; White, 2002). The most common mechanism is internal initiation of sg plus strand synthesis from the anti-genome by the RdRp, which is recruited by promoter sequences in the viral minus-strand RNA, as was initially described for *Brome mosaic virus* (BMV) (Miller *et al.*, 1985). The core promoter for basal level transcription and enhancer elements required for full sg promoter activity were identified in RNA domains both 3' and 5' of the transcription initiation site. The second model, premature termination of minus strand synthesis, proposes that minus strand synthesis is terminated when the RdRp encounters a sg promoter element in the plus strand. Subsequently, the attenuated minus-strand transcript serves as template for the production of sg plus strands. The third model, discontinuous extension of minus strand synthesis, is uniquely employed by two nidovirus subgroups, arteriviruses and coronavirus, and shares characteristics with the premature termination model described above. Nidovirus sg mRNA synthesis ensures that ORFs located downstream of the replicase ORFs 1a and 1b can be expressed. The 5'-proximal part of each sg mRNA encompasses a specific ORF (or sometimes several ORFs) from the 3'-proximal genome region, which thus becomes accessible for translation by host ribosomes (Fig. 1-3).

Nidoviruses produce a 3' co-terminal nested set of 2 to 10 mRNAs, depending on the particular virus species. However, only members of the *Coronavirus* and *Arterivirus* genera produce sg mRNAs that consist of two RNA segments that are noncontiguous in the virus genome. These sg mRNAs possess a so-called common leader, a 5' sequence element that is colinear with the 5'-proximal region of the genome, which is fused to sequence elements (sg mRNA bodies) derived from the 3'-proximal region of the genome (Fig. 1-3) (Baric *et al.*, 1983; Spaan *et al.*, 1983b; Lai *et al.*, 1984; de Vries *et al.*, 1990).

Over the years, different models have been proposed to explain the discontinuous step in arteri- and coronavirus sg RNA synthesis (van der Most & Spaan, 1995; Brian & Spaan, 1997; Lai & Cavanagh, 1997; Snijder & Meulenberg, 1998; Snijder & Meulenberg, 2001; Lai & Holmes, 2001). In recent years, the discontinuous minus strand extension model, which was originally proposed by (Sawicki & Sawicki, 1995), has gained considerable experimental support from both biochemical and genetic studies (van Marle *et al.*, 1999a; Baric & Yount, 2000; Sawicki *et al.*, 2001; de Vries *et al.*, 2001; Pasternak *et al.*, 2001; Sawicki & Sawicki, 2005). This model postulates that the fusion of the sg RNA body to the common leader sequence involves the discontinuous extension of minus-strand RNA synthesis, which presumably yields sg minus-strand templates that are used to produce the sg plus strands (Fig. 7-1) (Sawicki & Sawicki, 1995). Thus, as usual, minus-strand RNA synthesis is thought to initiate at the 3' end of the genome template and elongation of RNA synthesis is thought to be attenuated at one of the conserved transcription-regulating sequences (body TRSs), that precede almost every ORF in the 3'-proximal part of the genome. The full-length complement of the genome (anti-genome) is produced whenever the RdRp complex passes all TRSs without attenuation occurring. After attenuation, the nascent minus strand, with the body TRS at its 3' end, is thought to be translocated to the leader TRS that is located within the 5' UTR of the genome, a step that is guided by antisense body TRS-to-sense leader TRS base-pairing. Subsequently, minus strand synthesis is resumed to add the complement of the leader sequence (anti-leader), thus completing the minus-strand template, which can be used to generate the sg mRNAs.

It was proposed that arteri- and coronavirus discontinuous transcription may mechanistically resemble similarity-assisted copy-choice RNA recombination (Chang *et al.*, 1996; Brian & Spaan, 1997; Pasternak *et al.*, 2001). As in the case of RNA recombination, the higher order structure of the template (and possibly also of the nascent strand) may play an important role, in particular in the regions where leader TRS and body TRSs reside. Many details of the mechanism that regulates the fusion of sequences that are noncontiguous in the genome during sg RNA synthesis of arteri- and coronaviruses remain to be elucidated. The involvement of RNA secondary structures is likely, but these remain to be identified. In this paragraph, the involvement of RNA elements described in the literature will be discussed in the context of the model for discontinuous extension of minus-strand RNA synthesis, as it is currently believed to operate in coronaviruses and arteriviruses.

The attenuation step during discontinuous sg RNA synthesis has been extensively studied using coronavirus DI RNA systems as well as full-length arteri- and coronavirus genomes. It was found that the conserved part of the body TRS (or core TRS), without its flanking sequences, was sufficient to initiate RNA synthesis when introduced into the 3'-proximal third of a MHV DI RNA (Hsue & Masters, 1999), although the amount of sg mRNA produced from a core body TRS was lower than when the entire 18-nt mRNA7 TRS was introduced (Makino & Joo, 1993). Deletion or mutations within the body TRS influenced sg RNA synthesis dramatically (Makino *et al.*, 1991; Joo & Makino, 1992; van

der Most *et al.*, 1994). Additional factors or structural elements must contribute to attenuation, because the nidovirus genome contains many more TRS-like sequences that are not functional as body TRSs in the natural situation (den Boon *et al.*, 1991; Makino *et al.*, 1991; Joo & Makino, 1992; Pasternak *et al.*, 2000). This was emphasized by the finding, in several independent cases, that the body TRS is not even necessary for sg RNA synthesis. Leader-body fusions could occur either randomly, but with low efficiency, within a defined region on the genome, generating multiple species of sg mRNA (Zhang & Lai, 1994; Fischer *et al.*, 1997; de Vries *et al.*, 2001), or unique sequences were utilized that were different from the consensus TRS (Hofmann *et al.*, 1993; den Boon *et al.*, 1996; Zhang & Liu, 2000). Thus, the consensus body TRS is important, but it seems not to be an absolute prerequisite for leader-body fusion. The first evidence for a role of their flanking sequences came from studies in which a body TRS was inserted at different locations in a coronavirus DI RNA replicon. This resulted in different levels of sg RNA synthesis from the DI RNA genome, which were determined by the flanking sequences and not by the location of this TRS in the DI RNA (Jeong *et al.*, 1996). Since then, the repressing or enhancing effect of downstream or upstream sequences was demonstrated or suggested in several studies (van Marle *et al.*, 1995; An & Makino, 1998; Ozdarendeli *et al.*, 2001; Alonso *et al.*, 2002; Curtis *et al.*, 2004; Sola *et al.*, 2005).

In general, the smallest sg mRNAs of nidoviruses are more abundant than the larger sg mRNAs, which may be explained by the fact that smaller RNA molecules are produced at a faster rate. However, there are indications that this polar effect is determined prior to synthesis of the sg plus strands. For example, downstream body TRSs have a negative effect on transcription levels from upstream body TRSs (Joo & Makino, 1995; van Marle *et al.*, 1995; Hsue & Masters, 1999; Pasternak *et al.*, 2004). In the model for discontinuous minus strand extension, the number of transcription complexes reaching upstream body TRSs is determined by the number of attenuation events occurring at downstream body TRSs. Thus, what determines RdRp stalling? It has been argued that the accessibility of the body TRS itself determines or influences attenuation. However, the difference between a base-paired TRS and a TRS in an open conformation is not a general determinant of activity (Fig. 2-8A and B), and mutagenesis studies designed to open or close body TRS regions did not influence their activity (Pasternak *et al.*, 2000; Ozdarendeli *et al.*, 2001). Others have proposed a role of (upstream) attenuating structures that may stall the RdRp when it encounters a body TRS (Curtis *et al.*, 2004). As in *Tomato bushy stunt virus*, a hairpin network may terminate RNA synthesis during minus strand synthesis (Lin & White, 2004). There are some suggestions that sg RNA synthesis interferes with replication of a DI RNA. This interference was found to occur in *cis*, suggesting that a competition for replicative enzymes is an unlikely explanation (Jeong & Makino, 1992; Lin *et al.*, 1994).

How in the arterivirus and coronavirus system the attenuated nascent minus strand is transferred from the body TRS to the leader TRS is completely unknown. Circularization of the genome, or some kind of other higher order structural feature, may

bring leader TRS and body TRS in close proximity, an event that could be promoted by protein factors (Lai, 1998). Base-pairing between the leader TRS and a body TRS is an essential step during sg RNA synthesis and the type and position of TRS nucleotides influence the efficiency of this process (Joo & Makino, 1992; van Marle *et al.*, 1999a; Pasternak *et al.*, 2001). In general, the relative amount of sg mRNA correlates with the calculated stability of the corresponding leader TRS-body TRS duplex (Pasternak *et al.*, 2003; Zuniga *et al.*, 2004). In the past, experiments designed to unravel the function of the coronavirus leader in sg RNA synthesis were hampered by recombination between the helper virus RNA and the DI RNA replicons used for these studies. Nevertheless, deletion analysis of the leader region showed that at least the leader TRS and its flanking

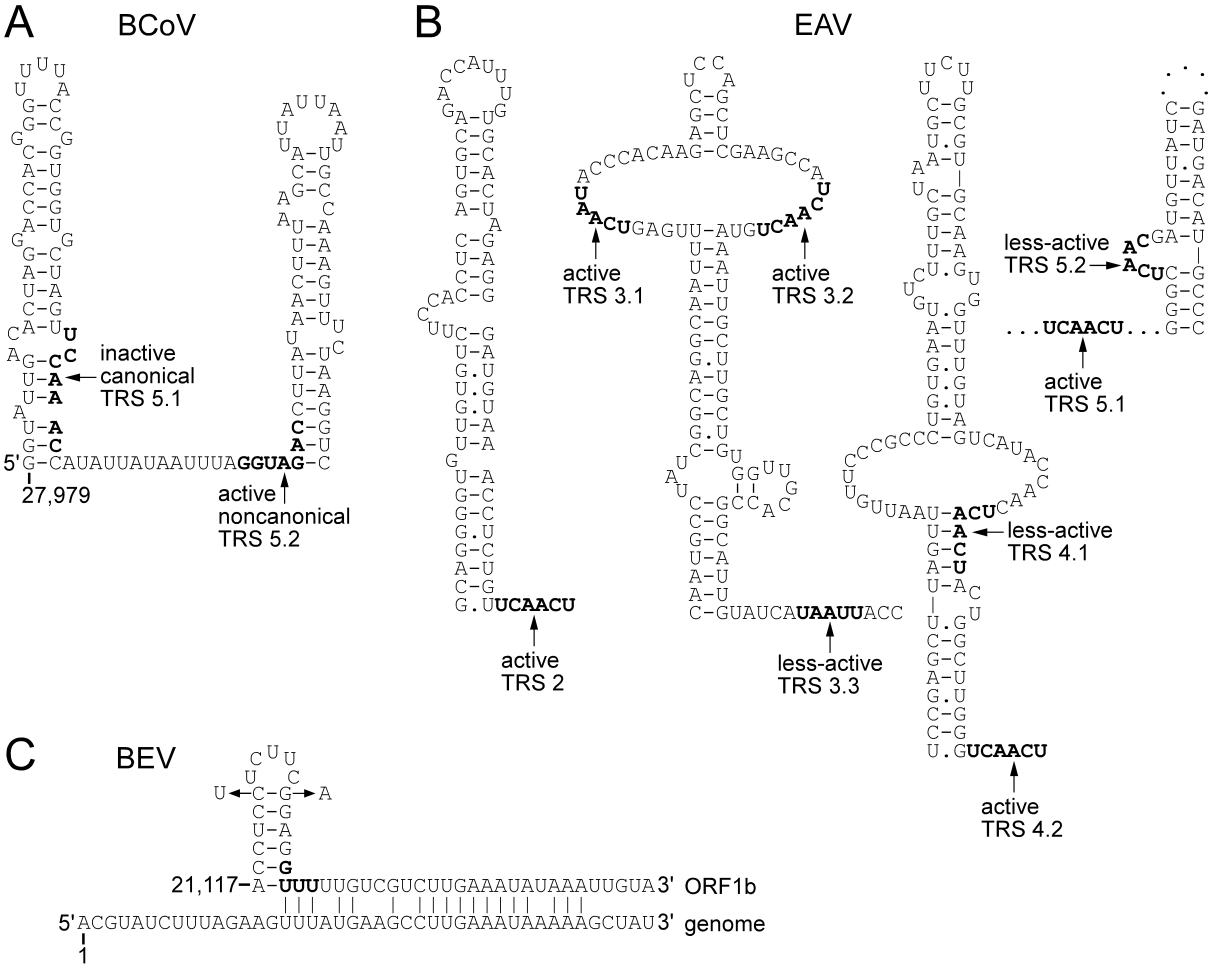


Figure 2-8. Predicted RNA secondary structure models of corona- and arterivirus body TRSs with their flanking sequences. *(A) Coronavirus BCoV (Mebus strain) body TRSs 5.1 and 5.2, taken from Ozdarendeli et al. (2001); (B) Arterivirus EAV (Bucyrus strain) body TRSs 2, 3, 4, and 5, taken from Pasternak et al. (2000), and Pasternak et al. (2003); (C) Torovirus BEV, taken from Van Vliet et al. (2002). Sequence variation with Porcine torovirus (PoTV) is indicated.*

sequences at both sides are required for the efficient generation of sg mRNAs (Lin *et al.*, 1994; Liao & Lai, 1994; Zhang *et al.*, 1994; Zhang & Lai, 1996; Wang & Zhang, 2000). Computer-based RNA secondary structure predictions of the 5'-proximal regions of corona- and arterivirus genomes, in which the leader TRS resides, identified a hairpin structure harboring the leader TRS (Fig. 2-1) (Shieh *et al.*, 1987; Chang *et al.*, 1994; van Marle *et al.*, 1999a). For one of these structures biochemical support was obtained (Chang *et al.*, 1996). In contrast, in another study, the TRS was predicted to be located in a single-stranded region, and not in a hairpin structure (Fig. 2-1) (Wang & Zhang, 2000; Stirrups *et al.*, 2000).

After leader-body TRS base-pairing, minus strand synthesis is resumed to add the leader complement and finalize the template for sg mRNA production. There are indications that sg-length minus strands are the primary templates of positive-strand sg RNA synthesis (Baric & Yount, 2000). Sg mRNAs themselves appear to be unable to function as replicons that are amplified to generate additional sg mRNAs, as was proposed previously (Sethna *et al.*, 1989). This was concluded from experiments demonstrating that a synthetic sg mRNA7 (with a natural 65 nt leader) that was transfected into BCoV-infected cells did not replicate, whereas the same construct with the 5'-terminal replication signals of 498 nt (Table 2-1) could be amplified (Chang *et al.*, 1994).

It was found that several deletions in the 3' UTR of a MHV DI RNA affected sg RNA production. Since in these experiments minus strands were still synthesized, this finding suggested that plus-strand RNA synthesis was affected. Thus, the 3'-proximal region may contain RNA elements that regulate sg RNA synthesis (Lin *et al.*, 1996). It has to be noted that the deletions may have altered the conformation of a bulged RNA hairpin or a pseudoknot that were predicted in this region (Goebel *et al.*, 2004a).

In addition to RNA elements, other factors may be involved in regulation of discontinuous sg RNA synthesis, such as proteins of cellular or viral origin. Two replicase proteins of EAV, nsp1 and nsp10, were shown to be involved in regulation of the abundance of sg mRNAs in cells (van Dinten *et al.*, 1997; van Marle *et al.*, 1999b; Tijms *et al.*, 2001). It is unlikely that structural proteins are involved in arteriviruses sg RNA synthesis, because by reverse genetics it was demonstrated that both genome replication and sg RNA synthesis are not affected by the absence of all structural proteins (Molenkamp *et al.*, 2000c). In contrast, there are strong indications that at least one coronavirus structural protein, the N protein, is involved in coronavirus replication (Compton *et al.*, 1987; Thiel *et al.*, 2001b; Almazan *et al.*, 2004; Schelle *et al.*, 2005). Also cellular factors may be involved in the generation of sg mRNAs. The consensus binding site of a cellular protein, hnRNP A1, is similar to the complement of the MHV TRS. A correlation was found between hnRNP A1 binding and sg RNA synthesis when body TRS7 was mutated (Furuya & Lai, 1993; Zhang & Lai, 1995; Li *et al.*, 1997). Another cellular protein, polypyrimidine tract-binding protein (PTB, also known as hnRNP I), was found to interact with the MHV leader TRS, and it was proposed that this interaction could regulate sg RNA synthesis (Furuya & Lai, 1993; Li *et al.*, 1999; Choi *et*

al., 2002a). There are some indications that PTB can influence sg RNA synthesis by inducing a conformational change of a structure located in the complementary strand of the genomic 3' UTR (Huang & Lai, 1999). However, the quality of the structure data presented by these authors is poor, and the authors neglected to evaluate the effect of the PTB interaction on DI RNA replication.

In contrast to arteri- and coronaviruses, viruses belonging to the genera *Ronivirus* and *Torovirus* produce sg mRNAs lacking a common leader sequence, with the exception of the largest sg mRNA (encoding the spike protein) of toroviruses (Snijder *et al.*, 1990b; Snijder *et al.*, 1991; van Vliet *et al.*, 2002; Cowley *et al.*, 2002). Conserved intergenic sequence (CIS) elements were identified preceding generally each ORF downstream of the roni- and torovirus replicase gene (Fig. 1-3). This sequence was also found at the 5' terminus of the torovirus genome, but was absent in the corresponding region of the ronivirus genome (Snijder *et al.*, 1990b; van Vliet *et al.*, 2002; Cowley *et al.*, 2002). This element does not precede the spike protein ORF, which is expressed from the single sg mRNA that appears to be produced via discontinuous synthesis of a minus strand sg template (van Vliet *et al.*, 2002). Recently, it has been shown that the CIS elements in the torovirus genome are essential and sufficient to drive sg RNA synthesis (Smits *et al.*, 2005). Most likely, the CIS elements play a dual role: the plus sense version acts as terminator of sg minus-strand RNA synthesis, whereas its complement at the 3' end of the sg minus-strand template is the promoter for sg plus strand synthesis (Fig. 7-1B). The available data suggests that the toro- and ronivirus CIS elements do not function as donor sites for similarity-assisted template switching, as was proposed for the arteri- and coronavirus body TRSs. Notable, the residues upstream of the fusion site of this largest torovirus sg mRNA can adopt a hairpin structure that may be conserved among toroviruses (Fig. 2-8C). It was suggested that the combination of this hairpin and sequence homology of sequences just downstream of the fusion site with the genomic 5'-proximal region may create a recombination hotspot during torovirus discontinuous sg RNA synthesis (van Vliet *et al.*, 2002).

Chapter 3

Sequence Requirements for RNA Strand Transfer During Nidovirus Discontinuous Subgenomic RNA Synthesis

Erwin van den Born, Alexander O. Pasternak, Willy J.M. Spaan, and Eric J. Snijder
(2001)
EMBO **20**: 7220-7228

Abstract

Nidovirus subgenomic mRNAs contain a leader sequence derived from the 5' end of the genome fused to different sequences ("bodies") derived from the 3' end. Their generation involves a unique mechanism of discontinuous subgenomic RNA synthesis that resembles copy-choice RNA recombination. During this process, the nascent RNA strand is transferred from one site in the template to another, either during plus or during minus strand synthesis, to yield subgenomic RNA molecules. Central to this process are transcription-regulating sequences (TRSs), which are present at both template sites and assure the fidelity of strand transfer. Here we present results of a comprehensive covariation mutagenesis study of *Equine arteritis virus* TRSs, demonstrating that discontinuous RNA synthesis depends not only on base-pairing between sense leader TRS and antisense body TRS, but also on the primary sequence of the body TRS. While the leader TRS merely plays a targeting role for strand transfer, the body TRS fulfils multiple functions. The sequences of mRNA leader-body junctions of TRS mutants strongly suggested that the discontinuous step occurs during minus strand synthesis.

Introduction

The genetic information of RNA viruses is organized very efficiently. Practically every nucleotide of their genome is utilized, either as protein-coding sequence or as cis-acting signal for translation, RNA synthesis, or RNA encapsidation. As part of their genome expression strategy, several groups of positive-strand RNA (+RNA) viruses produce subgenomic (sg) mRNAs (reviewed by (Miller & Koev, 2000)). The replication of their genomic RNA, which also is the mRNA for the viral replicase, is supplemented with the generation of sg transcripts to express structural and auxiliary proteins, which are encoded downstream of the replicase gene in the genome. Subgenomic mRNAs of +RNA viruses are always 3'-coterminal with the genomic RNA, but different mechanisms are used for their synthesis. Some viruses, like *Brome mosaic virus*, initiate sg mRNA synthesis internally on the full-length minus-strand RNA template (Miller *et al.*, 1985). Others, exemplified by *Red clover necrotic mosaic virus* (RCNMV), may rely on premature termination of minus strand synthesis from the genomic RNA template, followed by the synthesis of sg plus strands from the truncated minus-strand template (Sit *et al.*, 1998). Members of the order Nidovirales, which includes coronaviruses and arteriviruses, have evolved a third and unique mechanism, which employs discontinuous RNA synthesis for the generation of an extensive set of sg RNAs (reviewed by (Brian & Spaan, 1997; Lai & Cavanagh, 1997; Snijder & Meulenberg, 1998)). Nidovirus sg mRNAs fundamentally differ from other viral sg RNAs in that they are not only 3'-

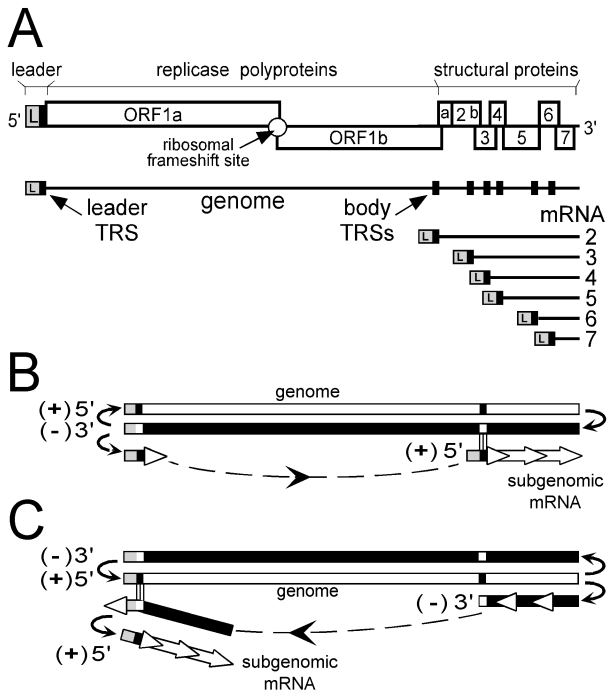


Figure 3-1. *(A) Schematic diagram of the genome organization and expression of EAV, the arterivirus prototype. The regions of the genome specifying the leader (L) sequence, the replicase gene (ORFs 1a and 1b), and the structural genes are indicated. The nested set of EAV mRNAs (genome and sg mRNAs 2-7) is depicted below. The black boxes in the genomic RNA indicate the position of leader and major body TRSs. (B and C) Alternative models for nidovirus discontinuous sg RNA synthesis. The discontinuous step may occur either during plus strand (B) or during minus strand (C) RNA synthesis. In the latter case, sg mRNAs would be synthesized from a sg minus strand template. For details see text.*

coterminal, but also 5'-coterminal with the genome (Fig. 3-1A). A 5' common leader sequence of 65 to 221 nucleotides, derived from the 5' end of the genomic RNA, is attached to the 3' part of each sg RNA (the "mRNA body").

Various models have been put forward to explain the cotranscriptional fusion of noncontiguous parts of the nidovirus genome during sg RNA synthesis (Fig. 3-1B and 3-1C). Central to each of these models are short transcription-regulating sequences (TRSs), which are present both at the 3' end of the leader and at the 5' end of the sg RNA body regions in the genomic RNA. The TRS is copied into the mRNA and connects its leader and body part (Spaan *et al.*, 1983a; Lai *et al.*, 1984). Synthesis of sg mRNAs was initially proposed to be primed by free leader transcripts, which would base-pair to the complementary TRS regions in the full-length minus strand, and would subsequently be extended to make sg plus strands (Fig. 3-1B; (Baric *et al.*, 1983; Baric *et al.*, 1985)). This model, however, was based on the report that sg minus strands were not present in coronavirus-infected cells (Lai *et al.*, 1982b). The subsequent discovery of such molecules (Sethna *et al.*, 1989) resulted in reconsideration of the initial "leader-primed transcription" model. (Sawicki & Sawicki, 1995) have proposed an alternative model (Fig. 3-1C), in which the discontinuous step occurs during minus instead of plus-strand RNA synthesis. In this model, minus strand synthesis would be attenuated after copying a body TRS from the plus-strand template. Next, the nascent minus strand, with the TRS complement at its 3' end, would be transferred to the leader TRS and attach by means of TRS-TRS base-pairing. RNA synthesis would be reinitiated to complete the sg minus strand by adding the complement of the genomic leader sequence. Subsequently, the sg minus strand would be used as template for sg mRNA synthesis and the presence of the leader complement at its 3' end might allow the use of the same RNA signals that direct genome synthesis from the full-length minus strand.

Using site-directed mutagenesis of TRSs of the arterivirus *Equine arteritis virus* (EAV), we have previously shown that base-pairing between the sense leader TRS and antisense body TRSs is crucial for sg mRNA synthesis (van Marle *et al.*, 1999a). However, base-pairing is only one step of the nascent strand transfer process and is essential in both models outlined in Fig. 3-1. The EAV genomic RNA contains several sequences that match the leader TRS precisely, but are nevertheless not used for sg RNA synthesis (den Boon *et al.*, 1996; Pasternak *et al.*, 2000). This suggests that leader-body TRS similarity alone is, though necessary, not sufficient for the strand transfer to occur.

To gain further insight into the cis-acting signals regulating sg RNA synthesis, we performed a comprehensive site-directed mutagenesis study of the EAV leader and body TRSs. Every nucleotide of the TRS (5' UCAACU 3') was substituted with each of the three alternative nucleotides. Our analysis revealed a number of striking similarities with the process of copy-choice RNA recombination, as it occurs in RNA viruses. Whereas the leader TRS plays only a targeting role in translocation of the nascent strand, body TRS nucleotides appear to fulfil diverse position-specific and base-specific functions. In

addition, the sequence of the leader-body junctions of the sg mRNAs produced by these mutants provided strong evidence for the discontinuous minus strand extension model.

Results and Discussion

EAV genome replication is not significantly affected by leader TRS and body TRS mutations

To dissect EAV RNA synthesis, we routinely use a full-length cDNA clone (van Dinten *et al.*, 1997), from which infectious EAV RNA is in vitro transcribed. Following transfection of the RNA into baby hamster kidney (BHK-21) cells, intracellular RNA is isolated and analysed by Northern blot hybridisation and RT-PCR (van Marle *et al.*, 1999a). Due to differences in transfection efficiency, the total amount of virus-specific RNA (genomic RNA and sg mRNA), isolated from transfected cell cultures, is somewhat variable. Thus, the accurate quantitation of sg mRNA synthesis by TRS mutants requires an internal standard for transfection efficiency. The amount of viral genomic RNA can be this standard, but only if its amplification is not dramatically affected by the TRS mutations. To prove that this is the case, we used the previously described mutants L4, B4, and LB4 (van Marle *et al.*, 1999a), in which five nucleotides of the TRS (5'-UCAAC-3') were replaced by the sequence 5'-AGUUG-3', either in the leader TRS (L4), RNA7 body TRS (B4), or both TRSs (LB4).

The three mutants were tested in three independent experiments. Intracellular RNA was isolated at 14 h post-transfection, early enough to prevent spread of the wt control virus to nontransfected cells (first cycle analysis). Transfection efficiencies were determined by immunofluorescence assays (see Materials and Methods) and varied between 10 and 23% (data not shown). Prior to RNA analysis, the amount of isolated intracellular RNA was corrected for the transfection efficiency of the sample, so that each lane in Fig. 3-2 represents EAV-specific RNA from an approximately equal number of EAV-positive cells. Phosphoimager quantitation revealed that genomic RNA replication of mutants L4, B4 and LB4 varied by not more than 30% (Table 1). These differences could reflect, for example, a slight influence of RNA secondary structure changes in the TRS regions on genomic RNA synthesis. Remarkably, however, the genomic RNA level of the leader/body TRS double mutant LB4 was not affected by more than 10%. In view of the results obtained with these pentanucleotide TRS mutants, we assumed that the amount of genomic RNA could indeed be used as an internal standard during the analysis of mutants containing only single nucleotide replacements in leader TRS and/or RNA7 body TRS.

Table 3-1. Effect of pentanucleotide TRS mutations on genomic RNA synthesis

Mutant	Genomic RNA synthesis (% of wild type)
L4	111 ± 12
B4	72 ± 14
LB4	89 ± 2

The RNA-RNA interaction between the leader and body TRSs is not the only factor that regulates EAV sg RNA synthesis

There are numerous examples of regulatory RNA-RNA interactions in both eukaryotic and prokaryotic cells, as well as in RNA viruses. Essential processes like translation, replication, and encapsidation of RNA virus genomes frequently depend on RNA-RNA interactions and higher order RNA structures. Regulation of sg RNA synthesis of +RNA viruses by RNA-RNA interactions is also not without precedent. In *Tomato bushy stunt virus*, an RNA element located 1000 nucleotide upstream of the sg RNA2 promoter base-pairs with the promoter and is necessary for sg RNA production (Zhang *et al.*, 1999). Similarly, base-pairing interactions between complementary sequences in the 5' end of the *Potato virus X* genomic RNA and sequences upstream of two major sg RNA promoters, are required for efficient sg RNA synthesis (Kim & Hemenway, 1999). In RCNMV, an intermolecular RNA-RNA interaction is required for sg RNA synthesis (Sit *et al.*, 1998).

Recently, we have established the pivotal role of an interaction between sense and antisense RNA sequences in the life cycle of EAV (van Marle *et al.*, 1999a). In that study, the role of TRS nucleotides C2 and C5 was tested by substituting them with G. It was concluded that base-pairing between the sense leader TRS and the antisense body TRS plays a crucial role in nidovirus sg RNA synthesis. We now took a more systematic approach and performed an extensive site-directed covariation mutagenesis study of the entire leader TRS and RNA7 body TRS, which directs the synthesis of the most abundant EAV sg RNA. Every nucleotide of the TRS (5'-UCAACU-3') was replaced with each of the other possible nucleotides. As in the study of (van Marle *et al.*, 1999a), every mutation

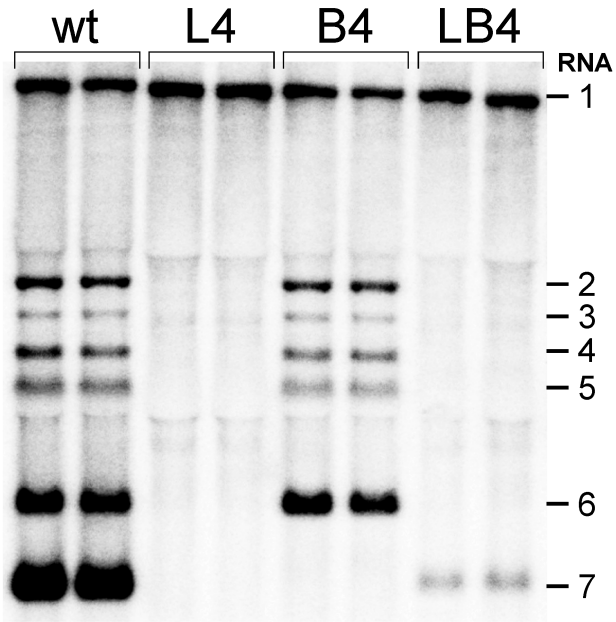


Figure 3-2. Northern analysis of EAV-specific RNA isolated from cells transfected with RNA transcribed either from the wt EAV infectious cDNA clone or from TRS pentanucleotide mutants (UCAAC to AGUUG). *The results of two independent experiments are shown.*

was introduced into leader TRS, RNA7 body TRS, and both TRSs, resulting in 54 mutant constructs. Each mutant was given a unique name: for example, BU1A refers to a mutant in which a U has been changed to A at position 1 of the body TRS, LU1A refers to the same substitution in the leader TRS, and DU1A means that these two substitutions were combined in one double mutant construct. The amount of sg RNA7 was quantitated by phosphoimager scanning of hybridised gels and was corrected for the amount of genomic RNA in the same lane (as outlined above). Fig. 3-3 shows the relative sg RNA7 level of the 54 mutants, compared to the RNA7 level of the wt control. For a selection of 11 interesting mutants (see below), the analysis was repeated three times (Fig. 3-4A and 3-4B), without observing significant variations in sg RNA synthesis.

The comprehensive analysis of the effects of TRS mutations considerably expanded our understanding of discontinuous sg RNA synthesis. Remarkably, the effects of single (leader or body) TRS mutations were mostly base-specific, i.e. different nucleotide substitutions at the same position affected sg RNA7 synthesis to different extents. For example, at position 1, BU1A mutant retained 44% of the wt RNA7 synthesis level, whereas both BU1C and BU1G mutants lost RNA7 synthesis almost completely. Conversely, when U1 of the leader TRS was changed to A or G, RNA7 synthesis was completely abolished, whereas 13% of the wt level were still maintained by LU1C. For position 2, only BC2U mutant retained 30% of the wt RNA7 synthesis level, while all the other position 2 single mutants have lost 90% or more of wt RNA7 synthesis. Another example is position 6: BU6C left only 5% of wt RNA7 synthesis, whereas BU6A produced much higher RNA7 levels. This implied that for some positions (1, 2, 6) certain mismatches in the duplex between plus leader TRS and minus body TRS, like U-U (BU1A, BU6A) or C-A (LU1C, BC2U), are allowed to a limited extent. In contrast, no mismatches were allowed for position 5, where all single nucleotide substitutions abolished RNA7 synthesis almost completely.

Surprisingly, both body TRS U to C substitutions at positions 1 and 6 (BU1C and BU6C) resulted in low levels of RNA7, despite the fact that these mutations allow the formation of a G-U base-pair between the plus leader TRS, providing the U nucleotide, and minus body TRS, providing the G. On the other hand, for positions 3 and 4, G-U base-pairing was shown to be functional, because mutants LA3G and LA4G, which can form G-U base-pairs between the G in the plus leader TRS and U in the minus body TRS, were the only position 3 and 4 single mutants that produced reasonable levels of RNA7. Taken together, these findings suggest that also other factors, besides leader-body base-pairing, play a role in sg RNA synthesis and that the primary sequence (or secondary structure) of TRSs may dictate strong base preferences at certain positions. Our analysis of the degree of complementation by the double mutants provided strong support for this assumption.

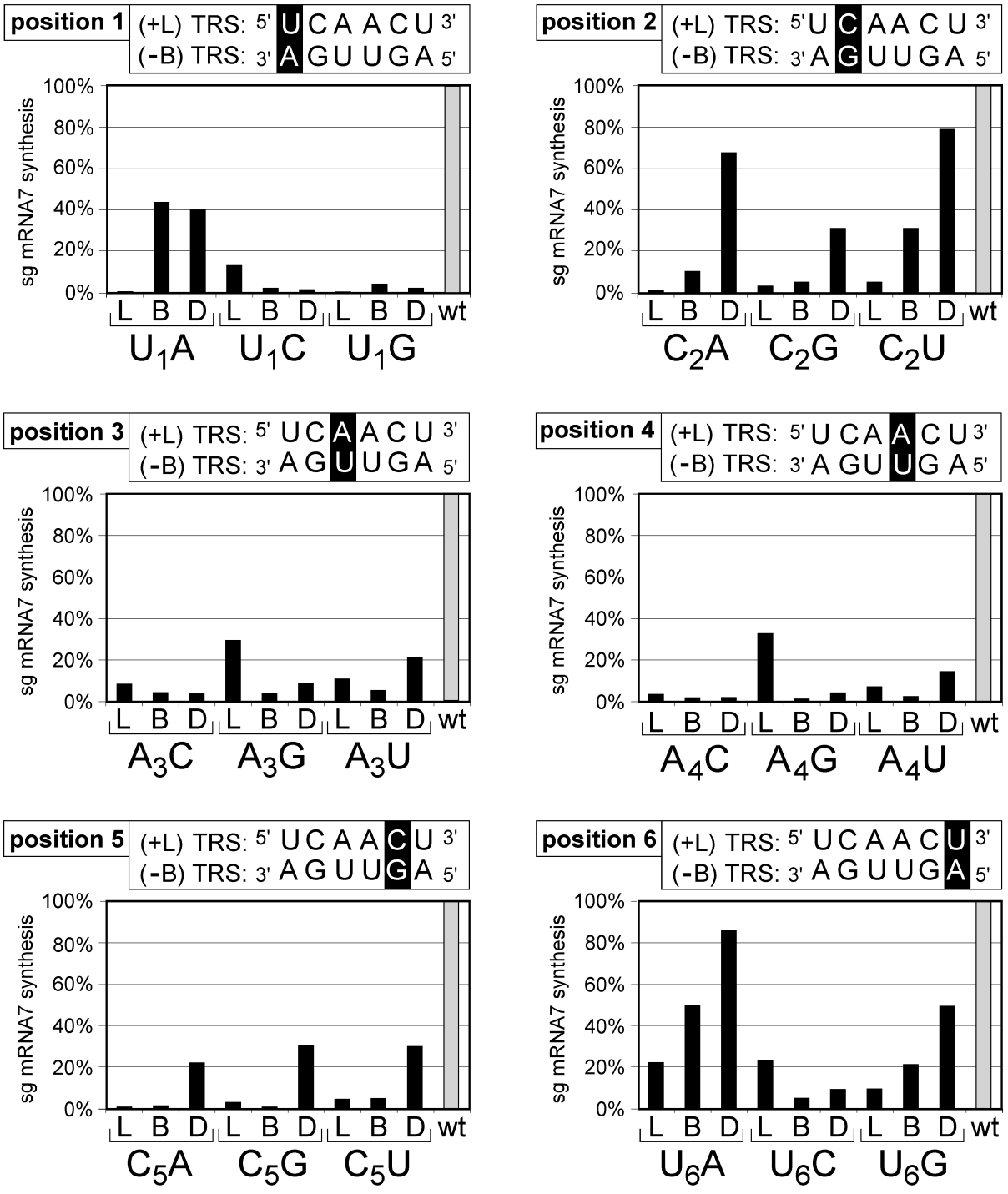


Figure 3-3. Relative amounts of sg RNA7 produced by EAV TRS mutants. *For each set of transfections (representing one experiment), the EAV wt construct pEAV030 (van Dinten et al., 1997) was taken along as a positive control. For every mutant, the level of sg RNA7 synthesis was calculated as ((sg/g)/(sg/g)wt)*100%: it was corrected for the level of genomic RNA (used as an internal standard; see text) and was subsequently related to the level of sg RNA7 produced by the wt control in the same experiment, which was also corrected for the corresponding genomic RNA level. The relative sg RNA7 level of the wt control was put at 100%.*

Differentiating between effects at the level of TRS sequence and the level of leader-body duplex formation

For some TRS nucleotides (2, 5, and 6, except in the case of DU6C), the RNA7 level of double mutants was clearly higher than that of the corresponding single mutants. This means that base-pairing between these leader and body TRS nucleotides is involved in sg RNA synthesis. However, none of these double mutants reached the wt sg RNA7 level. In the other double mutants (all position 1, 3, and 4 mutants, and DU6C), in clear contradiction to the predictions of the "base-pairing model", RNA7 synthesis was not significantly restored. Moreover, a comparison of the values for the B and D mutants in Fig. 3-3 showed that, for almost all of these mutants (e.g. the position 1 mutants), the amount of sg RNA7 produced by the double mutant appeared to be limited by the level allowed by the body TRS mutation. Sometimes the RNA7 level of the double mutant was even less than that of the leader mutant (DU1C, DA3G, DA4G, or DU6C). Clearly, for these substitutions, restoration of the possibilities for leader-body duplex formation did not restore sg RNA synthesis. Apparently this is because the effect of body TRS mutations at the level of primary sequence or secondary structure can be "dominant" over the duplex-restoring effects of the double mutations.

Body TRS mutants thus fell into two distinct types, determined by position and chemistry of the substitution. In mutants of the first type, sg RNA synthesis was impaired mainly because of the disruption of the leader-body TRS duplex. This effect could be compensated for by introduction of the corresponding mutation in the leader TRS, and in the double mutant sg RNA synthesis was restored compared to the corresponding single mutants. In mutants of the second type, sg RNA synthesis was down-regulated as a consequence of both TRS duplex disruption and disruption of the primary sequence (or secondary structure) of the body TRS. Obviously, the latter effect could not be compensated for by mutating the leader TRS, and the corresponding double mutants did not show restoration of sg RNA synthesis.

In contrast to our findings with the body TRS mutants, we did not obtain leader TRS mutations that appeared to determine the level of sg RNA7 synthesis of the corresponding double mutant (Fig. 3-3). Thus, effects of mutations in the leader TRS were not "dominant" over the duplex-restoring effects of the double mutations, suggesting that they only affected duplex formation. This indicated that the leader TRS probably does not have an additional, sequence-specific function in sg RNA synthesis in addition to its participation in TRS-TRS base-pairing. The fact that single leader TRS mutations at all six positions severely repressed RNA7 synthesis indicated that base-pairing of every TRS nucleotide contributes to sg RNA production. In this respect, it was significant that the two leader TRS mutants with the highest RNA7 levels, LA3G and LA4G, can form G-U base-pairs to maintain the duplex.

The observation that leader TRS mutations could be "rescued" by introducing complementary mutations in the body TRS, but that many body TRS mutations could not

be "rescued" by corresponding changes in the leader TRS, is clearly illustrated by the U1A mutants. Due to the restoration of TRS base-pairing possibilities, the RNA7 synthesis of double mutant DU1A was significantly increased compared to that of LU1A, but not above the level of BU1A. Thus, restoration of the leader-body duplex in DU1A exerted a clear effect on sg RNA7 production compared to LU1A, but had no effect on sg RNA synthesis compared to BU1A. This exemplified the dominant nature of a mutation in the primary sequence of a body TRS. In contrast, for instance, the BC2U mutation probably affected duplex formation only, because RNA7 synthesis was restored almost to wt levels in the DC2U double mutant.

These results indicate that there are strong base preference constraints for some body TRS positions. To accurately interpret these base preferences, one has to limit the analysis to the double mutants only, because in these mutants the down-regulation of sg RNA synthesis was only due to the sequence changes in the body TRS, and not due to the disruption of the leader-body TRS duplex. There were strict preferences for positions 1, 3, and 4 of the body TRS: at position 1, only the U to A substitution allowed for a significant RNA7 level (~40% of wt), and at positions 3 and 4 only the A to U mutants retained 15-20% of the wt level. For positions 2 and 5 the sequence constraints were less stringent (all substitutions allowed for more than 20% of wt level), but still only DC2A and DC2U reached more than 50%. At position 6 of the body TRS, only U to C was not allowed, whereas the other two double mutants still produced 50% or more of RNA7. In other words, the functional EAV RNA7 body TRS (based on the analysis of our single-

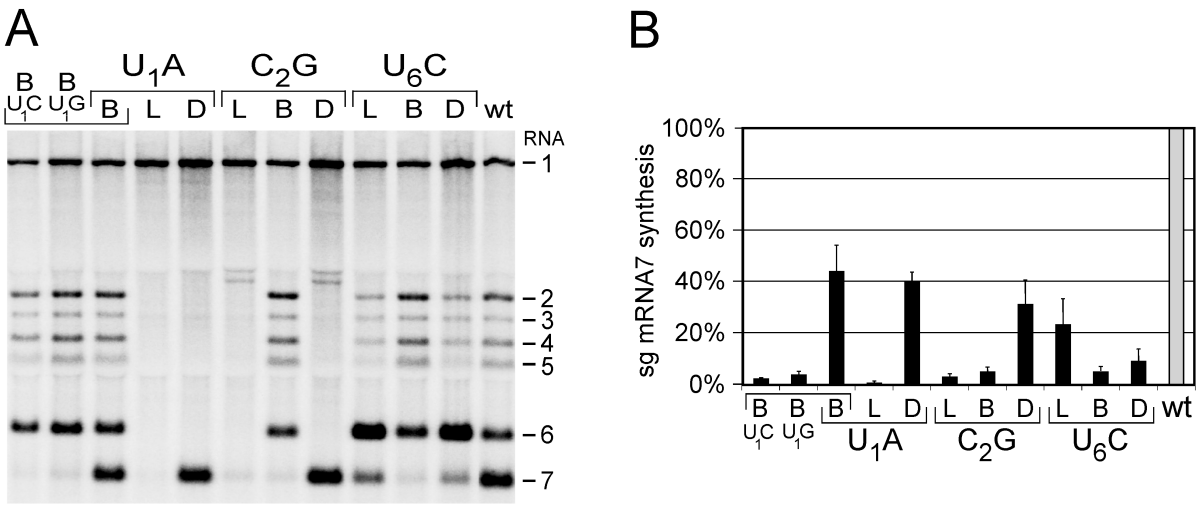


Figure 3-4. *(A) Northern analysis of EAV-specific RNA isolated from cells transfected with wt EAV RNA or 11 selected TRS mutants. (B) Relative amount of sg RNA7 produced by the selection of 11 TRS mutants. See the legend to Figure 3 for the calculation of sg RNA7 levels. For each construct in (B), experiments were repeated at least three times and average sg RNA levels are shown.*

nucleotide substitutions) can be described as U1(C/u/a)2A3A4C5(U/a/g)6, with wt nucleotides shown in upper case and nucleotides that allowed for at least 50% of wt RNA7 level shown in lower case. Remarkably, TRS nucleotides A3, A4, and C5 are conserved in the TRSs of all other arteriviruses (Snijder & Meulenberg, 1998). Also the fact that DC2U retained 80% of RNA7 synthesis corresponded nicely to the presence of a U at this position in other arteriviruses.

Until recently (Almazan *et al.*, 2000; Thiel *et al.*, 2001a), infectious cDNA clones were lacking for coronaviruses. Consequently, most studies on coronavirus sg RNA synthesis were carried out using defective interfering (DI) RNAs. These replicons carried body TRSs from which moderate levels of sg mRNAs could be produced in the presence of helper virus. Using this system, (Joo & Makino, 1992) and (van der Most *et al.*, 1994) performed body TRS mutagenesis studies for the *Murine coronavirus* (MHV). Joo and Makino systematically mutagenised the core of the MHV body TRS. In contrast to our results, they found that in only two of 21 body TRS mutants sg RNA synthesis from the DI RNA genome was abolished, whereas all others supported normal levels of sg RNA production. Thus, it is possible that the MHV TRS that was used in that study is more tolerant to single-nucleotide mismatches than the EAV sg RNA7 TRS.

In a similar study, (van der Most *et al.*, 1994) observed that U to C substitutions at positions 1 and 3 of the MHV body TRS, which maintained the duplex by changing a U-A base-pair into a U-G base-pair, reduced sg RNA levels more strongly than substitutions that disrupted the duplex (van der Most *et al.*, 1994). This implies that, as in the case of EAV, leader-body TRS duplex formation is not the only factor that determines coronavirus sg RNA synthesis. However, because of the limitations of the DI RNA system, the leader TRS could not be mutagenised in these studies, and body TRS-specific effects could not be distinguished from effects at the level of leader-body duplex formation.

The discontinuous step in nidovirus sg RNA synthesis occurs during minus-strand RNA synthesis

Due to recent studies of arterivirus and coronavirus sg RNA (van Marle *et al.*, 1999a; Baric & Yount, 2000; Sawicki *et al.*, 2001), the discontinuous minus strand extension model (Fig. 3-1C) has been gaining more and more ground. This model predicts that the TRS-derived sequence that forms the leader-body junction in the sg mRNA is a copy of the body TRS, and not of the leader TRS. The leader-primed transcription model predicts the opposite (Fig. 3-1B). Therefore, determining the origin of the leader-body junction of sg mRNAs would help to distinguish between the two models. However, in the wt situation, EAV leader and body TRSs are identical and consequently one cannot determine the origin of the sg mRNA leader-body junction. This problem could be overcome by tracing the mutations introduced in leader or RNA7 body TRS mutants, most

of which retained part of their ability to produce mRNA7. In a previous study (van Marle *et al.*, 1999a), we found that nucleotides 2 and 5 of the mRNA7 leader-body junction sequence were exclusively derived from the body TRS, and not from the leader TRS. This was shown by direct sequencing of RT-PCR products obtained from the residual mRNA7 produced by mutants BC2G, LC2G, BC5G, and LC5G (van Marle *et al.*, 1999a).

Using the same approach, we now analysed mRNA7 from mutants BC2A and BC2U, and also these transcripts contained the mutated nucleotide derived from the body TRS (data not shown). Assuming that only one crossover event occurs during leader-body joining, we could thus map this crossover between positions –1 and +2 of the sg RNA junction sequence. This left the intriguing question whether the crossover site could be mapped even more precisely. In other words, was nucleotide +1 of the junction sequence derived from the body TRS or the leader TRS?

Using the position 1 mutants described above, we could answer this question (Fig. 3-5). The most striking result was that mRNA7 of mutants BU1A, BU1G, and LU1C exclusively contained the body TRS-derived nucleotide at position +1. Thus, for these mutants, the crossover site could be precisely mapped between TRS nucleotide positions –1 and +1, meaning that the complete leader-body junction sequence in an EAV sg mRNA can be body TRS-derived. On the other hand, sg RNAs from mutants LU1A, BU1C, and LU1G contained mixed populations of leader TRS- and body TRS-derived nucleotides at position +1 (Fig. 3-5): A and U for LU1A, C and U for BU1C, and G and U for LU1G. Remarkably, this pattern correlated with the relative amounts of sg mRNA7 produced by these mutants (Fig. 3-3). Mutants that produced populations of sg RNAs that were mixed with respect to the origin of the nucleotide at position +1 of the leader-body junction had lost RNA7 synthesis almost completely. On the other hand, mutants that contained exclusively the body nucleotide at position +1 retained higher levels of RNA7 synthesis. This observation may be explained as follows: in the wt situation, the large majority of the crossovers probably occurs between positions –1 and +1, leading to a body TRS-derived nucleotide at position +1 in the sg RNA; however, a low number of crossovers take place between nucleotides +1 and +2, resulting in a leader TRS-derived nucleotide at position +1. Mutants in which almost all sg RNA synthesis is blocked by a substitution at position +1 may somehow be deficient in the crossover between –1 and +1, but may have retained the ability for crossovers between +1 and +2, which were detected by sequence analysis. Conversely, in position +1 mutants that retain reasonable sg RNA synthesis, most crossovers occur between positions -1 and +1, and they obscure the minority of crossovers between +1 and +2 in the sequencing electropherogram. Alternatively, position +1 TRS mutations that strongly interfere with sg RNA synthesis may force a shift of the crossover site in the remaining molecules.

We believe that our present findings strongly support the discontinuous minus strand extension model. Indeed, the fact that a complete body TRS can be copied into the sg RNA is very difficult to reconcile with the alternative model, in which sg RNA synthesis from the genomic minus-strand template is primed by free plus-strand leader

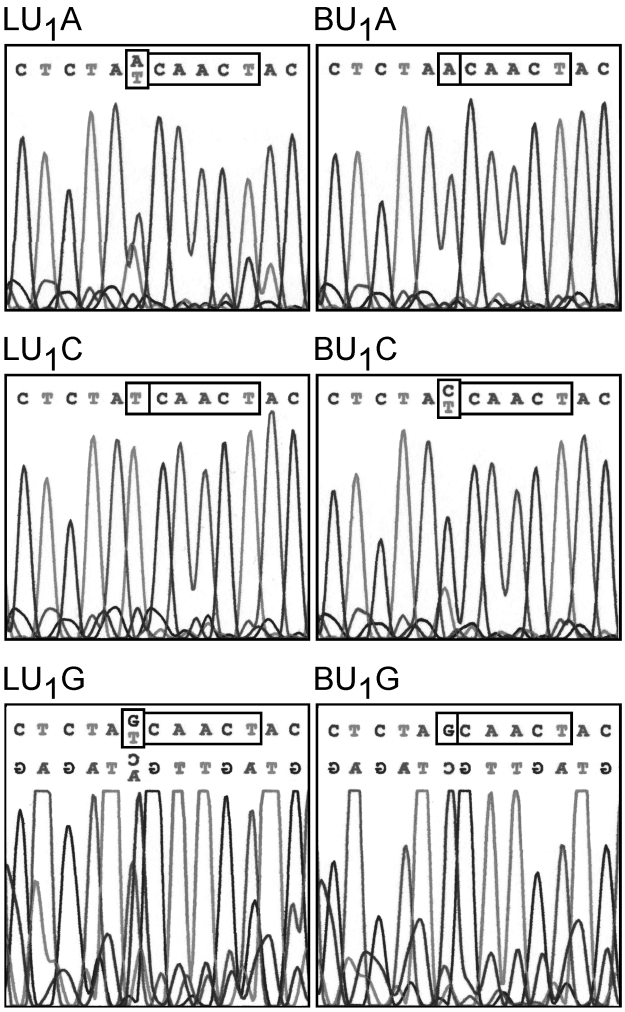


Figure 3-5. Sequence analysis of mRNA7 leader-body junctions from position 1 TRS mutants. *Sequences were determined directly from sg mRNA7-specific RT-PCR products. For the U1A and U1C mutants, the sequence shown corresponds to the plus strand of sg RNA7. For sequencing-related technical reasons, the minus strand sequence was determined for the U1G mutants; a mirror image of the electropherogram is shown with the corresponding plus strand sequence listed at the top of the panel. TRSs and the mRNA7 leader-body junctions detected by our sequence analysis are boxed.*

transcripts that contain the leader TRS at their 3' end (Fig. 3-1B). To explain the presence of a complete copy of the body TRS in the sg mRNA in this model, one would have to assume that a 3'-5' exonuclease activity trims back the free leader transcript prior to its extension into a sg mRNA (Baker & Lai, 1990). Note that there would not be a single base-pair left to hold these "trimmed" leader molecules on the template. Such an enzymatic activity, which is unprecedented in +RNA viruses, exists in yeast retrotransposon Ty5 (Ke *et al.*, 1999), in which reverse transcription is primed by an internal region in a tRNA. However, in this system, it is not a part of the duplex that is removed, but the single-stranded 3' tail of the tRNA, which cannot base-pair with the Ty5 RNA.

Removal of the TRS at the 3' end of the nidovirus leader, which has already base-paired with the template, would energetically be very unfavourable for the RdRp. Instead of starting elongation using the intact and properly positioned leader as a primer, it would have to disrupt the newly formed duplex, degrade part of the leader RNA, and then reinitiate polymerisation, without any base-pairing between primer and template. It has

been shown that *Influenza virus* transcription does not require a sequence match between the (cellular) RNA primer and the (viral) template (Plotch *et al.*, 1981). However, if also in the nidovirus system the "trimmed" leader RNA could be fixed on the template solely by RNA-protein interactions, the targeting of the nascent strand by TRS base-pairing would be extremely puzzling.

Sequence data of sg RNA leader-body junctions from other arteriviruses are also hard to reconcile with the leader-primed transcription model. For the porcine and simian arteriviruses (Meulenberg *et al.*, 1993b; Godeny *et al.*, 1998), the leader-body junctions of some sg RNAs mapped two nucleotides upstream of the body TRS, which again would not leave a single nucleotide to hold the putative free leader on the template after the hypothetical "back trimming". On the other hand, these findings and our data can be readily explained by the discontinuous minus strand extension model (Fig. 3-1C). The six-nucleotide duplex formed between the body TRS complement at the 3' end of the leaderless sg minus strand and the leader TRS in the genomic RNA template should suffice to properly position the nascent minus strand for subsequent elongation to add the complement of the leader sequence. In most cases, the nascent minus strand contains the entire body TRS complement at its 3' end at the moment of strand transfer, leading to a body TRS-derived leader-body junction sequence in the sg mRNA molecule. In a small number of transcripts, however, minus strand synthesis appears to be interrupted before nucleotide +1 of the body TRS is copied and, after strand transfer, resumes by incorporating the complement of the +1 nucleotide of the leader TRS. As stated above, we postulate that the detection of this phenomenon is determined by the level of crossovers between the –1 and +1 position that is allowed by the mutations introduced at the +1 position of body TRS or leader TRS. We cannot, however, formally exclude that a "back trimming" activity degrades the 3'-terminal nucleotide of the minus strand before or after strand transfer. However, note that in the discontinuous minus strand extension model (Fig. 3-1C), such an activity would not disturb the proper positioning of the nascent minus strand on the leader template, because the TRS-TRS duplex would be shortened with one nucleotide only.

Nidovirus discontinuous minus strand extension resembles similarity-assisted, copy-choice RNA recombination

Due to their discontinuous sg RNA synthesis, nidoviruses occupy a special "niche" in the +RNA virus world. Their mode of sg RNA production is clearly different from that of other +RNA viruses and resembles another well-documented +RNA virus feature: RNA recombination (see (Nagy & Simon, 1997; Aaziz & Tepfer, 1999; Worobey & Holmes, 1999), for recent reviews). Most of the experimental evidence supports an RdRp template switch (Kirkegaard & Baltimore, 1986) as the main mechanism of RNA recombination. Mechanistically, such a template-switch involves the transfer of a nascent

strand from one RNA template (donor) to the other (acceptor). Also nidovirus discontinuous sg RNA synthesis involves transfer of a nascent RNA strand, the sg RNA, but now from one site to another in the same template.

Based on the data currently available, we refer to the discontinuous minus strand extension model as our working model for nidovirus sg RNA synthesis. If one applies the "recombination terms" to this model (Chang *et al.*, 1996; Brian & Spaan, 1997; van Marle *et al.*, 1999a), the donor stand would be the body part of the genomic RNA template, the acceptor strand would be the leader part of the genomic RNA template, and the nascent strand would be the discontinuously synthesized minus strand. (Nagy & Simon, 1997) have defined three main classes of RNA recombination: similarity-essential, similarity-nonessential, and similarity-assisted recombination. The latter is defined as a mechanism in which strand transfer is determined by both sequence similarity between the parental RNAs and additional RNA determinants, present only in one of the parental RNAs.

The results of our present study strongly suggest that nidovirus discontinuous sg RNA synthesis can be considered a special case of high-frequency similarity-assisted RNA recombination. While the only obvious function of the leader TRS is to ensure the fidelity of the strand transfer by base-pairing with the 3' end of the nascent strand, the body TRS in the donor template indeed has additional, sequence-specific functions. One of these functions is apparently to pause (or terminate) nascent strand synthesis and thereby provide the opportunity for strand transfer. In addition, body TRS-derived nucleotides may play a role in the reinitiation of nascent strand synthesis on the acceptor template. Given the compact nature of the EAV TRS, it is quite possible that some nucleotides fulfil multiple tasks.

RNA secondary structure of the body TRS may regulate sg RNA synthesis

The sequence-specific function of the body TRS, revealed in this study, may be exerted either at the level of primary sequence or secondary structure. For a number of +RNA viruses, RNA secondary structure motifs located in the (proximity of) sg RNA promoters are vital for sg RNA synthesis. In *Alfalfa mosaic virus* (Haasnoot *et al.*, 2000), *Turnip crinkle virus* (TCV) (Wang *et al.*, 1999), and *Barley yellow dwarf virus* (Koev *et al.*, 1999), stem-loop structures in sg RNA promoter regions of the template strand are required for sg RNA synthesis. The sg RNA1 promoter of the latter virus is especially interesting, since it contains two stem-loop domains. For one of them, secondary structure, but not the primary sequence, is important for sg RNA synthesis, whereas the other domain acts through primary sequence, and not secondary structure (Koev *et al.*, 1999). Similarly, RNA secondary structure may play only a minor role in the sequence-specific recognition of the BMV sg RNA promoter by the RdRp (Adkins *et al.*, 1997; Siegel *et al.*, 1997).

We have suggested previously that RNA secondary structure of body TRS regions contributes to their attenuating potential and thereby determines the relative portion of the nascent minus strands that is transferred to the leader TRS in the template (Pasternak *et al.*, 2000). At present, it is unknown whether EAV body TRSs are part of an RNA structural motif that is essential for body TRS function, or whether they are recognized by a protein factor in a sequence-specific manner. However, the latter seems less likely than the former, since even LB4 (Fig. 3-2), in which five TRS nucleotides were substituted, still produced some sg RNA7, although ~30-fold less than the wt control. The fact that some sequences in the EAV genome match the leader TRS perfectly but are not used for sg mRNA synthesis, also argues against the recognition of a specific sequence (Pasternak *et al.*, 2000). More likely, mutagenesis of the RNA7 body TRS disturbed an RNA structure that is necessary for its function. This could, for example, explain the fact that the BU6C substitution reduced the amount of RNA7 by 20-fold (and could not be rescued by the same mutation in the leader TRS), whereas the wt RNA6 body TRS contains a C at the very same position. If a protein factor were involved in sequence-specific TRS recognition, then one would expect it to recognize all TRSs similarly. If RNA structure is important for recognition by such a protein, then the BU6C substitution probably disturbs a structural motif of the RNA7 TRS, which is not present in the RNA6 TRS. On the other hand, conservation of part of the TRS in other arteriviruses suggests a sequence-specific recognition. Further studies are required to distinguish between these possibilities.

In the TCV satellite RNA recombination system, the hairpin structure in the acceptor strand, as well as the donor-acceptor homology region, are necessary for the template switch (Nagy *et al.*, 1998). The hairpin has been postulated to bind the RdRp, whereas the homology region targets the nascent strand to the crossover site. The TCV RdRp likely recognizes the secondary and/or tertiary structure of the hairpin, while individual nucleotides play a less important role (Nagy & Simon, 1998). In EAV, the leader TRS in the acceptor template is predicted to reside in the loop of an extensive hairpin, and its base-pairing interaction with the body TRS complement at the 3' end of the nascent minus strand would resemble certain antisense RNA-regulated control mechanisms that are based on interactions between single-stranded tails and hairpin loops ((van Marle *et al.*, 1999a), and references therein). It is possible that also the EAV RdRp, or its accessory proteins, bind to the stem of the long hairpin that presents the leader TRS. In any case, the leader TRS itself does not seem to be recognized by a protein in a sequence-specific manner.

The body TRS is a better candidate to serve as a protein recognition site. This protein would then mediate the pausing of the nascent strand synthesis and/or nascent strand transfer. This would resemble the DNA-dependent RNA polymerase I termination system, in which specific DNA-binding terminator proteins bind to termination sequences (Reeder & Lang, 1997), or a function of the HIV nucleocapsid protein, which promotes the minus strand strong-stop DNA transfer (Guo *et al.*, 1997). The EAV replicase component nsp1, which was recently shown to possess a sg RNA synthesis-specific activity (Tijms *et al.*,

2001), may be a good candidate for such a regulatory role. Residues predicted to form a zinc finger structure in nsp1 were shown to be necessary for sg RNA synthesis. Interestingly, zinc finger structures in the HIV nucleocapsid protein facilitate strand transfer (Guo *et al.*, 2000). Finally, it should be noted that also the RNA structure of the nascent strand may influence pausing, strand transfer, or reinitiation, as illustrated by the fact that stable hairpin structures in the nascent strand promote termination of transcription by E.coli RNA polymerase (Wilson & von Hippel, 1995).

Materials and Methods

Site-directed mutagenesis, RNA transfections and immunofluorescence analysis

Site-directed mutagenesis of EAV leader and body TRSs was carried out as described by (van Marle *et al.*, 1999a) and all mutant constructs were sequenced. Following in vitro transcription from infectious cDNA clones, full-length EAV RNA was introduced into BHK-21 cells by electroporation as described by (van Dinten *et al.*, 1997). Immunofluorescence assays with EAV-specific antisera were done at 14 hours post-transfection as described by (van der Meer *et al.*, 1998). To visualize the nuclei for cell counting, nuclear DNA was stained with 5 µg/ml Hoechst B2883 (Sigma). Cells were counted using the Scion Image software (Scion corporation) and the percentage of transfected cells was calculated on the basis of the number of cells positive for the EAV replicase component nsp3 (Pedersen *et al.*, 1999).

RNA isolation and analysis

For RNA analyses, cells were lysed at 14 hours post-transfection. Intracellular RNA isolation was performed by using the acidic phenol method as described by (Pasternak *et al.*, 2000). Total intracellular RNA was resolved in denaturing agarose-formaldehyde gels. Hybridisation of dried gels with the radioactively labelled oligonucleotide probe E154, which is complementary to the 3' end of the EAV genome and recognizes all viral mRNA molecules (genomic and subgenomic), and phosphoimager quantitation of individual bands were performed as described by (Pasternak *et al.*, 2000). To determine the leader-body junction sequence of sg mRNA7, mRNA7-specific RT-PCR reactions were done as described by (van Marle *et al.*, 1999b) using an antisense (RT and PCR) primer from the RNA7 body region and a sense PCR primer matching a part of the leader sequence. RT-PCR products were sequenced directly as described by (Pasternak *et al.*, 2000) using the leader-derived primer, an ABI PRISM™ sequencing kit (Perkin Elmer), and an ABI PRISM™ 310 Genetic analyser (Perkin Elmer).

Acknowledgements

We are grateful to Stanley Sawicki, Dorothea Sawicki, Paul Masters, Alexander Gultyaev, Kees Pleij, Marieke Tijms, and Richard Molenkamp for helpful discussions and comments. We acknowledge Jessika Dobbe for technical assistance. A.O.P. was supported by grant 700-31-020 from the Council for Chemical Sciences of the Netherlands Organization for Scientific Research.

Chapter 4

Secondary Structure and Function of the 5'-Proximal Region of the Equine Arteritis Virus RNA genome

Erwin van den Born, Alexander P. Gultyaev, and Eric J. Snijder (2004)
RNA, **10**: 424-437

Reprinted with permission

Abstract

Nidoviruses produce an extensive 3' -coterminal nested set of subgenomic mRNAs, which are used to express their structural proteins. In addition, arterivirus and coronavirus mRNAs contain a common 5' leader sequence, derived from the genomic 5' end. The joining of this leader sequence to different segments (mRNA bodies) from the genomic 3'-proximal region presumably involves a unique mechanism of discontinuous minus-strand RNA synthesis. Key elements in this process are the so-called transcription-regulating sequences (TRSs), which determine a base-pairing interaction between sense and antisense viral RNA that is essential for leader-to-body joining. In order to identify RNA structures in the 5'-proximal region of the *Equine arteritis virus* genome that may be involved in subgenomic mRNA synthesis, a detailed secondary RNA structure model was established using bioinformatics, phylogenetic analysis, and RNA structure probing. According to this structure model, the leader TRS is located in the loop of a prominent hairpin (Leader TRS Hairpin; LTH). The importance of the LTH was supported by the results of a mutagenesis study using an EAV molecular clone. Besides evidence for a direct role of the LTH in subgenomic RNA synthesis, indications for a role of the LTH region in genome replication and/or translation were obtained. Similar LTH structures could be predicted for the 5'-proximal region of all arterivirus genomes and, interestingly, also for most coronaviruses. Thus, we postulate that the LTH is a key structural element in the discontinuous subgenomic RNA synthesis and is likely critical for leader TRS function.

Introduction

Like cellular mRNAs, genomes of positive-stranded RNA (+RNA) viruses possess a 5'-untranslated region (5'-UTR or leader) that precedes the first open reading frame. It is the primary function of this region of viral +RNA genomes to ensure their efficient use as a template for the host cell translation machinery. The reliable expression of the viral replicative enzymes is the crucial first step leading to viral reproduction. Viral genomes can contain a 5' cap structure (reviewed in (Gallie, 1998) or, as in the case of e.g. picornaviruses and some flaviviruses, an internal ribosomal entry site (IRES; reviewed in (Vagner *et al.*, 2001). Both structures promote translation initiation by recruiting initiation factors and ribosomes of the cellular translation machinery. An important secondary role of the 5' UTR of both cellular mRNAs and +RNA virus genomes is protection against degradation by exonucleases (Gallie, 1998). The 5' cap structure, but also hairpin motifs located close to the 5' end (Fuerst & Moss, 1989), can be important for the stabilization of the RNA. In addition to being the initiation site for translation of the viral genome into replicative (poly)proteins, RNA sequences in the 5'-proximal region of +RNA virus genomes can be involved in multiple additional functions, including genome replication, genome encapsidation and subgenomic (sg) mRNA synthesis.

Equine arteritis virus (EAV) is the prototype of the family *Arteriviridae* that belongs to the order *Nidovirales*, together with the *Coronaviridae* and *Roniviridae* families (Snijder & Meulenberg, 1998). The polycistronic +RNA genome of arteriviruses is 5' capped and 3' polyadenylated (Snijder & Meulenberg, 1998). The 5'-proximal three-quarters of the 12.7 kb EAV genome contain two large replicase open reading frames (ORFs), designated ORF1a and ORF1b, which encode polyproteins that are proteolytically cleaved into a set of 12 nonstructural proteins (reviewed in (Snijder & Meulenberg, 1998); (Ziebuhr *et al.*, 2000). The structural proteins of EAV and other nidoviruses, on the other hand, are not expressed from the genomic RNA, but from a 3'-coterminal nested set of sg mRNAs. In the case of arteriviruses and coronaviruses, these transcripts contain a common 5' sequence element that has been coined "leader sequence", even though in most mRNAs of these viruses the "translational leader" extends downstream of this common 5' sequence. In sg mRNAs, the leader sequence is linked to different, but overlapping sequences derived from the 3'-proximal quarter of the genome, which are referred to as "mRNA bodies" (Fig. 4-1A). Leader-to-body joining is thought to employ an RNA recombination-like strand transfer mechanism that probably operates during minus-strand RNA synthesis (see below), thus generating subgenomic minus-strand templates for subgenomic mRNA synthesis (Fig. 4-1B).

In most arteriviruses, the common leader sequence of sg mRNAs is identical to the 5' UTR of the genome (reviewed in (Snijder & Meulenberg, 1998). However, the 5' UTR of the EAV genome (nt 1-224) extends 13 nt beyond the 3' end of the common leader sequence (nt 1-211) of the mRNAs (de Vries et al. 1990; den Boon et al. 1991). Both the distance and the sequence between the 3' end of the common leader sequence and the

translation initiation codon of the most 5'-proximal ORF is different in each of the sg mRNAs. Thus, in terms of translation, the leaders of EAV mRNAs have different 3'-proximal segments, but all share the same 5'-terminal 211 nt. In this paper, for clarity, we will use the term "leader" exclusively in the context of sg mRNA synthesis, i.e. when referring to the 211-nt RNA element from the genomic 5' end that is linked to each of the sg mRNA bodies (Fig. 4-1C).

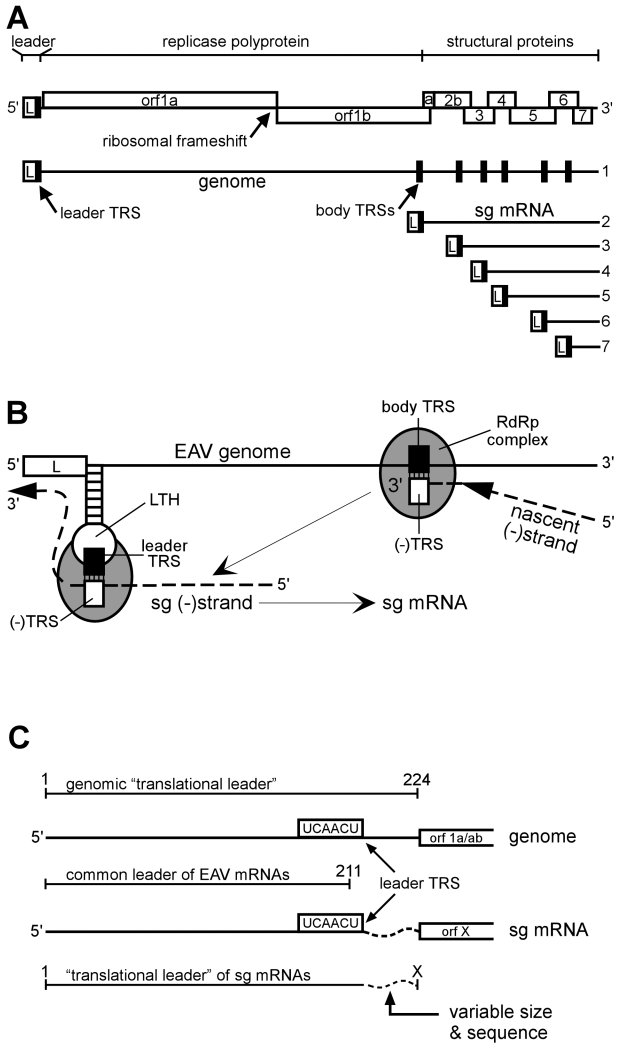


Figure 4-1. *(A) Schematic overview of the genome organization and expression of EAV. The replicase gene (ORFs 1a and 1b), the structural protein genes (2a, 2b, 3, 4, 5, 6 and 7) and the leader (L) are indicated. The black boxes in the genomic RNA indicate the position of the leader and major body TRSs. The nested set of sg mRNAs 2-7 is depicted below the genome. (B) The discontinuous minus strand extension model for nidovirus sg RNA synthesis (Sawicki and Sawicki 1995). After attenuation of the RdRp complex at a body TRS in the plus strand genomic template, the nascent strand is translocated to the leader TRS region. Following (-) body TRS to (+) leader TRS base-pairing, minus strand RNA synthesis resumes to add the complement of the leader sequence. Subsequently, the sg minus strand serves as a template for sg mRNA production. (C) Organization of the 5'-proximal region of EAV mRNAs. The EAV leader is defined as the 5' sequence that is common to the viral genome and all sg mRNA species and contains the leader TRS at its 3' end. The "translational leader" of the genome extends 13 nt beyond the 3' end of the common leader sequence, whereas in the "translational leader" of each of the sg mRNAs the segment between the common leader sequence and the first open reading frame differs both in size and sequence.*

Nidovirus sg mRNA synthesis is primarily governed by short conserved transcription-regulating sequences (TRSs), which have the consensus sequence 5'-UCAACU-3' in the case of EAV. TRSs are present both at the 3' end of the leader sequence (the leader TRS; nt 207-212 of the EAV genome) and at the 5' end of each of the body regions in the 3'-proximal region of the genome (body TRSs; Fig. 4-1A). Base-pairing between the leader TRS (in the plus strand) and the complement of the body TRSs

(in the minus strand) is essential for EAV sg mRNA synthesis (van Marle *et al.*, 1999a); Pasternak et al. 2001; (Pasternak *et al.*, 2003). Several models have been put forward to explain the production of the sg RNAs of arteriviruses and coronaviruses, which involves the fusion of sequences that are noncontiguous in the viral genome. In our opinion, the model of Sawicki and Sawicki (Sawicki & Sawicki, 1995), which proposes a discontinuous step during minus-strand RNA synthesis, is best supported by the available data from both biochemical and genetic studies with arteriviruses and coronaviruses (Schaad *et al.*, 1990; den Boon *et al.*, 1996; Baric *et al.*, 2001; Sawicki *et al.*, 2001; Pasternak *et al.*, 2001). According to this model (Fig. 4-1B), minus-strand RNA synthesis would be attenuated when the RdRp complex encounters a body TRS in the plus-strand genomic template. The nascent transcript, with the TRS complement at its 3' end, would then be transferred to the leader TRS and, guided by leader TRS-body TRS base-pairing, sg minus strand synthesis would be resumed to add the complement of the genomic leader sequence. Subsequently, the sg minus strand, now carrying the same 3'-terminal sequences as the full-length genomic minus strand, would serve as template for sg plus strand (mRNA) synthesis.

We have previously postulated (Pasternak *et al.*, 2001) that the primary role of the EAV leader TRS is to serve as a target for the base-pairing interaction with the nascent minus strand. However, as in the case of body TRSs, the functionality of the leader TRS is likely to be determined by additional factors than its primary structure, including higher order RNA structural context and protein factors that may interact with the leader TRS region of the genome. We have previously predicted the existence of a striking hairpin structure (leader TRS hairpin; LTH) that presents the leader TRS in its loop and may be a crucial element for EAV sg RNA synthesis (van Marle *et al.*, 1999a). However, as outlined above, RNA sequences in the 5'-proximal part of the EAV genome may have several additional functions in (i) genome translation, (ii) specifying the 3' end of the viral minus strand, which must contain essential recognition signals for genome synthesis and probably also for sg mRNA transcription from sg minus strands, (iii) genome encapsidation (the encapsidation signals of arteriviruses have not been mapped; (Molenkamp *et al.*, 2000b) and (iv) interactions between 5'- and 3'-terminal sequences in the viral genome. The latter have been found to occur in other viral systems and the resulting circularization of the genome may be a general feature in +RNA virus replication. For example, the 5' and 3' ends of the *Sindbis virus* genome RNA must interact to initiate replication and this is facilitated by translation initiation factors and PABP (Frolov *et al.*, 2001). Similarly, the cloverleaf structure at the 5' end of the poliovirus genome interacts with the polyA tail via a RNA-protein-protein-RNA bridge that involves (at least) PCBP and PABP (Herold & Andino, 2001).

To gain further insight into the structure and function of this multifunctional RNA element of the EAV genome, we have now analyzed in detail its secondary structure using a combination of computer and phylogenetic analyses, RNA structure probing, and site-directed mutagenesis. We present a well-supported structure model for the 5'-terminal 313

nt of the genome, a segment that is known to be sufficient for efficient genome replication and sg mRNA synthesis (Tijms *et al.*, 2001). Our combined data strongly support the existence in EAV of the LTH structure, which may also be conserved in other arteriviruses and in coronaviruses. We postulate that the LTH is a key structural element in EAV discontinuous sg RNA synthesis, but have also obtained indications that this region in the genomic 5' end is involved in other functions that are crucial for virus replication.

Results

The 5'-proximal region of the EAV genome contains two conserved domains and a variable region

As outlined in the Introduction, the 5'-proximal region of the EAV genome may harbor (potentially overlapping) RNA signals for translation, replication, encapsidation and subgenomic mRNA synthesis. In order to define conserved regions and sequence variation in this region of the genome, a comparative analysis of a number of EAV isolates was performed (Fig. 4-2). For this purpose, we selected eight EAV isolates (indicated with an asterisk in Fig. 4-2) that were diverged as much as possible (Stadejek *et al.*, 1999). These virus isolates were grown on BHK-21 cells, intracellular RNA was isolated and direct sequence analysis of RT-PCR products was used to determine the sequence of the 5'-terminal kilobase of the genome. In addition, we included the sequence of our Bucyrus "standard virus" (den Boon *et al.*, 1991) and 9 other EAV sequences that were retrieved from the database. The latter only ranged from the 5' end to the leader TRS (nt1-211). It should be noted that sequences downstream of the leader TRS of EAV and the coronavirus *Mouse hepatitis virus* (MHV) were previously found to be required for sg mRNA synthesis (Wang & Zhang, 2000; Tijms *et al.*, 2001) and are therefore an essential part of the alignment (Fig. 4-2).

Nucleotide variation in the alignment was less than 8 percent between any pair of isolates. Two regions in the alignment, nt 44-76 and 185-216 (Fig. 4-2, box I and III), were found to be highly conserved with the latter containing the leader TRS. In contrast, a pyrimidine-rich region between nt 137-163 was more variable, containing a relatively high number of nucleotide substitutions and also single-nucleotide insertions and deletions (Fig. 4-2, box II). Subsequently, the sequence conservation and variation observed in this sequence alignment were used as a tool in the prediction of a RNA secondary structure for the 5'-proximal region of the EAV genome (see below).

RNA secondary structure prediction for the 5'-proximal region of the EAV genome

Previously, we have reported that an EAV deletion mutant lacking nucleotides U^{297} to C^{1004} was able to replicate its genome efficiently, but did not produce any sg mRNAs (Tijms *et al.*, 2001). Subgenomic mRNA synthesis was restored when the translation product of the deleted region, nsp1, was expressed from an IRES-driven cassette inserted into the 3'-proximal region of the genome (Tijms *et al.*, 2001). Hence, it was concluded that the 5'-proximal RNA signals required for translation, replication and sg RNA

synthesis are contained in the first 296 nt of the EAV genome. The RNA secondary structure of this region, which includes the 5' UTR and a small part of ORF1a, was predicted by using the STAR program (Fig. 4-3A; (Gultyaev *et al.*, 1995). Our model contains a prominent hairpin that contains the leader TRS in its loop (hairpin G; Fig. 4-3A). This hairpin represents the top section of the extended LTH structure model that was described previously (van Marle et al. 1999; see also below).

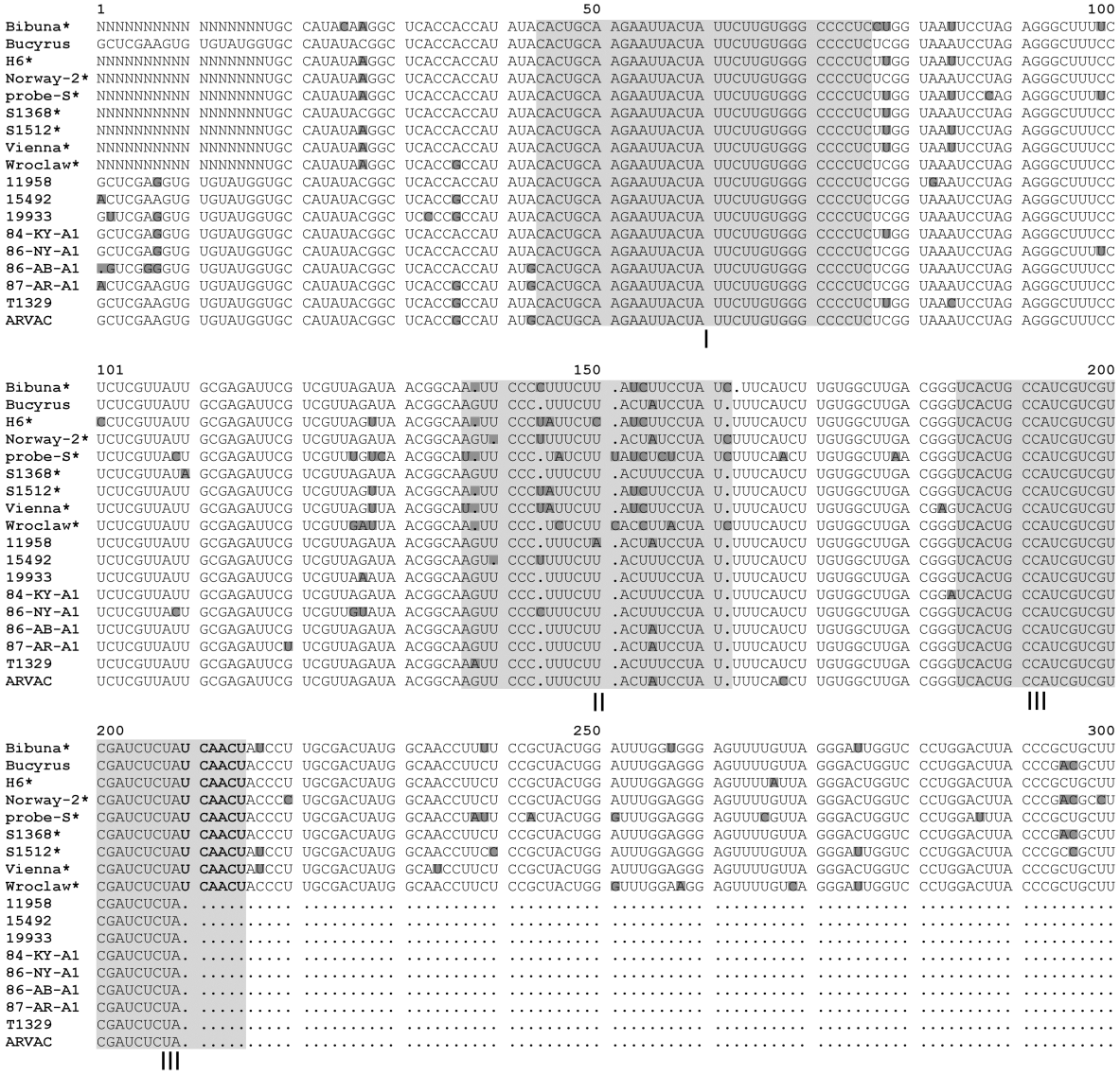


Figure 4-2. Alignment of the 5'-proximal 300 nucleotides of the genome of 18 EAV isolates. *cDNA sequences were retrieved from the database (see Material and Methods for accession numbers) or were experimentally determined (indicated with an asterisk). Nucleotides that differ from the consensus sequence are highlighted in dark gray. Two conserved regions were identified, which are highlighted with light gray boxes I and III. Box II represents a variable, pyrimidine-rich region.*

The RNA secondary structure model was optimized using the genetic variation observed in the sequence alignment (Fig. 4-2). Due to the low sequence variation among the EAV isolates, covariations supporting the predicted structure could not be found. Nevertheless, conserved nucleotides resided in the structured regions of our model, whereas e.g. the variable region (nt 137-163) was predicted to form a relatively long single-stranded stretch. In particular, the LTH stem and most of the nucleotides in its loop were found to be completely conserved, underlining the potential importance of this region. Furthermore, a similar stem-loop structure could be predicted in the corresponding region of other arterivirus genomes, despite rather high sequence variability (see also below).

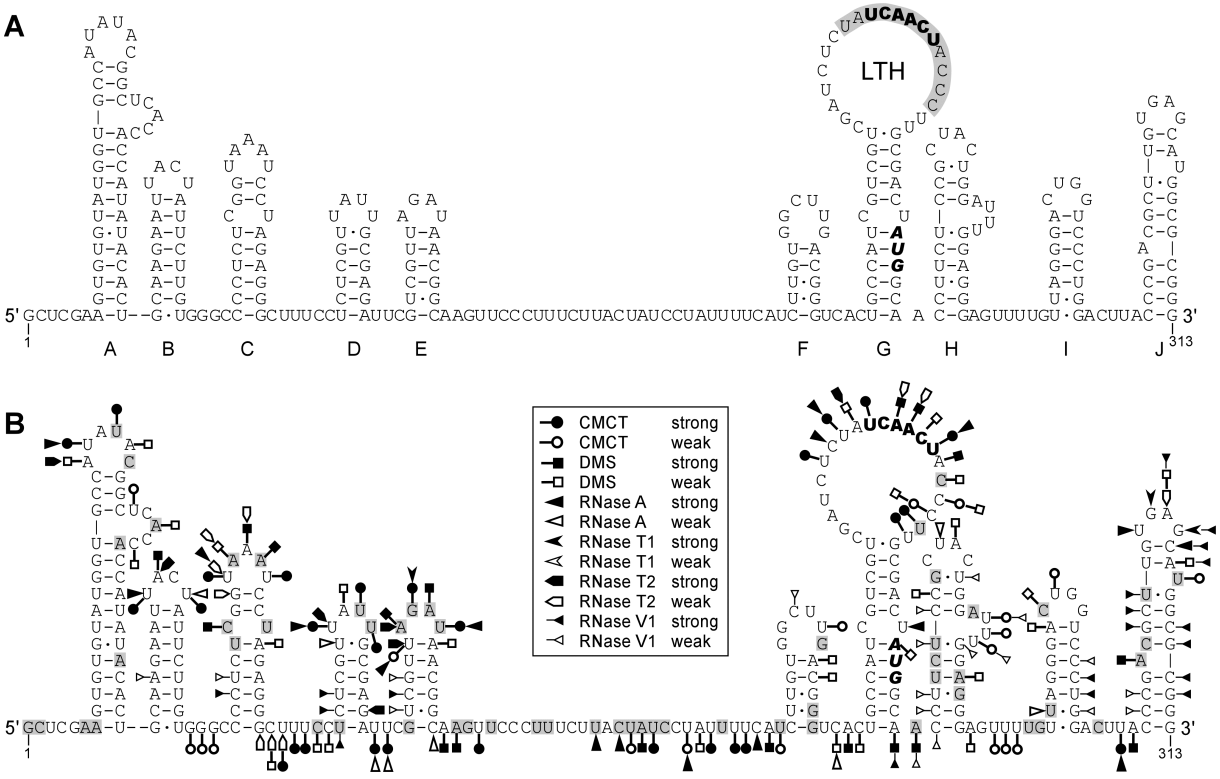


Figure 4-3. *(A) RNA secondary structure model of the 5'-proximal 313 nucleotides of the EAV (Bucyrus) genome. Hairpins are labeled A to J below the structure. LTH and nucleotide positions in the genome are indicated. The replicase ORF1a translation initiation codon is shown in bold italic font. In the leader TRS region absolutely conserved nucleotides are shown in bold font and the nucleotide stretch implicated in leader TRS-body TRS base-pairing (on the basis of sequence complementarity) is highlighted in gray (den Boon et al. 1996). (B) Summary of support for the structure model from phylogenetic analysis and structure probing. Nucleotides that showed sequence variation between EAV isolates (Fig. 2) are highlighted in gray. Sites susceptible to modification by CMCT, DMS, or RNases are indicated. See box in figure for details.*

Experimental support for the predicted secondary RNA structure

To gain experimental support for our structure model, we used chemical and enzymatic probing to supplement the phylogenetic analysis. A synthetic RNA transcript representing the 5'-terminal 589 nt of the EAV genome was produced and either treated with the methylating agents CMCT and DMS or digested with RNases A, T_1, T_2, and V_1 (reviewed in (Ehresmann *et al.*, 1987). CMCT alkylates N_1 of unpaired guanines or N_3 of unpaired uridines and DMS alkylates N_1 of unpaired adenines and N_3 of unpaired cytosines. RNase A cleaves internucleotide bonds 3' of unpaired uridines or cytosines, but also has some affinity for double-stranded regions (Christiansen *et al.*, 2003). RNase T_1 cleaves 3' of unpaired guanines and RNase T_2 cleaves 3' of unpaired adenines, but is less specific. Finally, RNase V_1 is specific for double-stranded regions present in helices and stacked nucleotides.

Following treatment of the synthetic transcript with the agents listed above, sites of RNA modification or cleavage were detected by using primer extension. Uncleaved and unmodified RNA was included as a control substrate, to detect nicks in the RNA or natural strong stops during reverse transcription. As an example of the results, a primer extension analysis of the LTH region is shown in Fig. 4-4. Noticeably, not all of the predicted double-stranded regions were confirmed by RNase V_1 digestion, but this was also reported previously by other investigators (Melchers *et al.*, 1997), (Damgaard *et al.*, 2002). The combined chemical and enzymatic probing results are summarized in Fig. 4-3B and were in good agreement with the structure model.

LTH mutations can severely affect EAV viability

To investigate the importance of the LTH structure, site-directed mutagenesis of an EAV infectious cDNA clone was used to change its primary and (predicted) secondary structure. The size of the large 21-nucleotide loop of the LTH was reduced to 13 nt and, in a second set of mutants, the upper part of the LTH stem was disrupted. In the first set of mutants, three nucleotides at the basis of the loop on the 3' or 5' side were changed to allow base-pairing with three nucleotides at the opposite side of the loop, thereby extending the stem with four base-pairs and creating mutants L-R and L-L, respectively. The normal size of the loop was restored in mutant L-LR by combining the mutations from mutants L-R and L-L (Fig. 4-5A).

To disrupt the upper part of the stem, we engineered mutant US-R in which nucleotides in the 3' side of the LTH stem were mutagenized to become identical to the nucleotides on the opposite side, with the exception of nucleotide A^{222} that occupies an important position in the context of the ORF1a translation initiation codon (Mathews, 2002). The same approach was used for the 5' side of the LTH stem in mutant US-L, while maintaining nucleotide U^{194} (to allow restoration of base-pairing between U^{194} and

A^{222} in the double mutant). By combining these mutations in double mutant US-LR, the base-pairs that form the LTH stem were restored. A schematic representation of these constructs is presented in Fig. 4-5A.

Full-length RNA was transcribed from all mutant EAV full-length clones and was transfected into BHK-21 cells by electroporation. Mutants were first analyzed by immunofluorescence double staining (data not shown) of transfected cells with antibodies specific for nonstructural protein 3 (nsp3), which is a marker for viral genome replication, and the nucleocapsid protein (N), which is translated from sg mRNA 7. Despite the fact

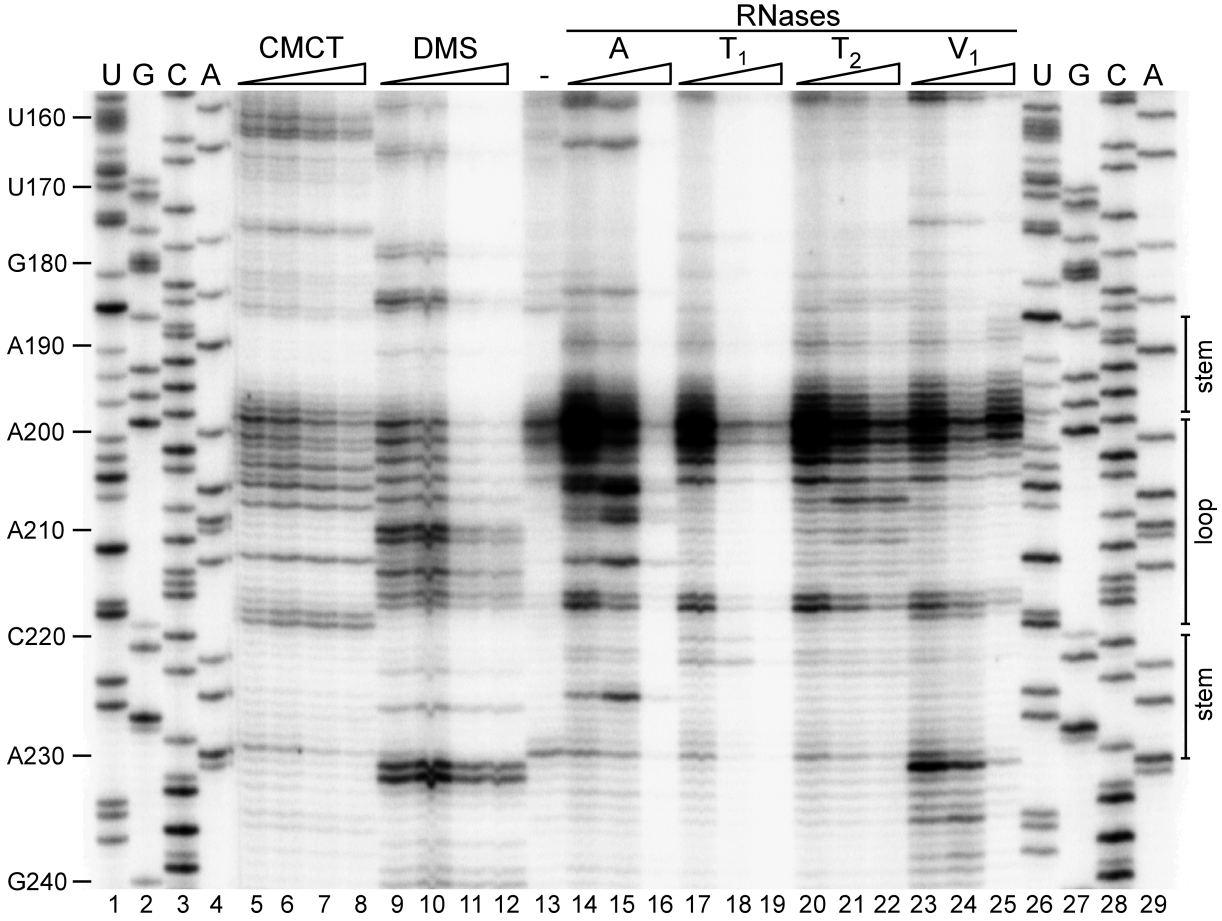


Figure 4-4. Chemical and enzymatic probing of the structure of the LTH region. *RNA was treated with alkylating reagents and digested with various RNases. Modified or cleaved sites were detected by primer extension analysis using a 5' end-labeled primer that was complementary to nt 285-310 of the EAV genome. Lanes 1-4 and 26-29 were used to run a dideoxy-sequencing reaction using the same primer. Lanes 5-8, RNA modified with 4.2, 8.4, 12.6 and 16.8 mg/ml of CMCT, respectively; lanes 9-12, RNA modified with 0.25, 0.5, 0.75 and 1.0% DMS, respectively; lane 13, untreated RNA; lanes 14-16, RNA digested with $5{\cdot}10^{-5}$, 10^{-4} or $5{\cdot}10^{-4}$ U of RNase A, respectively; lanes 17-19, RNA digested with 10^{-1}, $5{\cdot}10^{-1}$ or 10^{0} U of RNase T_1, respectively; lanes 20-22, RNA digested with 10^{-2}, $5{\cdot}10^{-2}$ or 10^{-1} U of RNase T_2, respectively; lanes 23-25, RNA digested with $5{\cdot}10^{-4}$, 10^{-3} or $5{\cdot}10^{-3}$ U of RNase V_1, respectively. Ten nucleotide intervals are indicated on the left side of the autoradiogram and the position of the LTH is indicated on the right side.*

that the *in vitro* transcription reactions yielded comparable amounts of full-length RNA for all mutants and the positive control, transfection of the RNA of the single mutants (L-R, L-L, US-R, and US-L) resulted in very low numbers of positive cells (< 0.01%), suggesting that the mutations severely affected genome replication. When the predicted structure of the LTH was restored in double mutants US-LR and L-LR, the number of positive cells was substantially higher. In the few cells that were positive for nsp3, only very low levels of N protein expression were detected for the mutants with a disrupted LTH stem, US-R and US-L, and the closed loop mutant L-R, whereas N protein expression could not be detected at all for closed loop mutant L-L. These results suggested severely reduced sg mRNA 7 synthesis for each of these mutants, although it is unclear to which extent this phenotype was due to the reduced overall fitness of these mutants, that was apparent from the low number of positive cells in these transfection experiments. Interestingly, however, both double mutants (US-LR and L-LR) showed restoration of N protein expression, although not to wild-type levels.

To investigate genome replication and sg mRNA synthesis in more detail, intracellular RNA was isolated from transfected cells and viral RNA was visualized by Northern blot hybridization with a probe specific for the 3' end of all viral mRNAs (Fig.

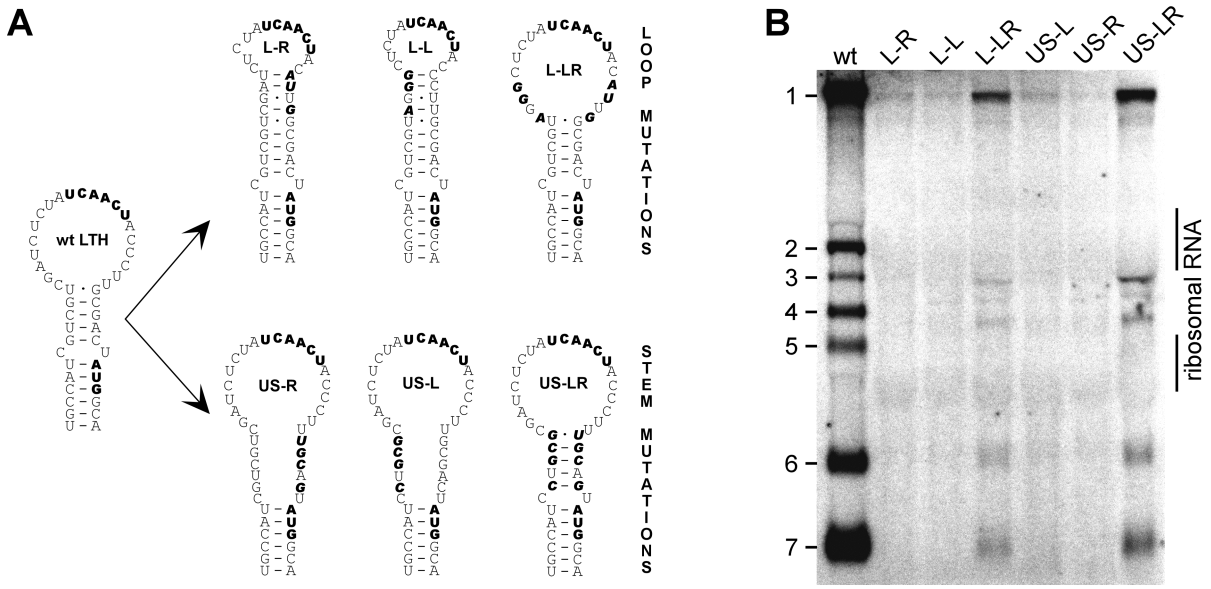


Figure 4-5. Mutagenesis of the EAV LTH. *(A) Schematic representation of the mutations introduced into the LTH structure. The partially closed loop mutants, L-R and L-L, and the reopened loop double mutant, L-LR, are depicted at the top of the panel. The open stem mutants, US-R and US-L, and the swapped stem double mutant, US-LR, are presented below. Leader TRS and ORF1a AUG codon are shown in bold font. Mutated nucleotides are indicated in bold italic font. (B) Northern blot analysis of virus-specific RNA isolated at 16 h post transfection from BHK-21 cells transfected with wt or mutant RNA. The position of the viral mRNAs is indicated: 1, genomic RNA; 2 to 7, sg mRNAs. The migration of sg mRNAs in the lanes of the mutants was somewhat aberrant due to the presence of an excess of rRNA. Due to the low transfection efficiencies obtained with these mutants, 20 times more total intracellular RNA was loaded compared to the wild-type control lane.*

4-5B). Despite the small number of positive cells, we were able to detect trace amounts of genome RNA for all single mutants, but sg mRNAs were not detectable. For both double mutants, the level of genome replication was clearly increased and small amounts of all sg mRNAs could be detected. Taken together, these results supported the importance of the LTH region of the EAV genome, although they mainly revealed that its structure may also be crucial for genome replication, in addition to its postulated role in sg mRNA synthesis.

The LTH structure is important for the synthesis of sg mRNAs

Because the lower part of the LTH stem harbors the replicase ORF1a translation initiation codon, it was not possible to mutate this region without affecting genome RNA translation. To circumvent this problem, we tried to uncouple translation from the LTH structure by introducing a novel translation initiation codon at nucleotide position 304-306, 79 nucleotides downstream of the native AUG, which was inactivated by mutating it to UAG. Previously, a similar approach was used to construct an EAV nsp1 knockout mutant, in which replicase translation started with nsp2 from an engineered AUG codon inserted at the 5' end of the nsp2-coding sequence (Tijms *et al.*, 2001). In our initial mutant (pLS-R; Fig. 4-6A), the mutations that inactivated the native ORF1a AUG codon also destroyed two U-A base-pairs in the lower part of the LTH stem and thereby expanded the bulge in the central part of the LTH stem. To restore nsp1 expression, the novel ORF1a translation initiation codon was followed by a sequence specifying the N-terminal 29 amino acids of nsp1. The codon usage in this sequence was altered to reduce the chance of RNA recombination between sequences up- and downstream of the novel AUG codon (mutant LS-R; Fig. 4-6A).

In contrast to the mutants described in the previous paragraph, the LS-R mutant virus replicated its genome with wild-type efficiency. However, it produced low amounts of sg mRNA and did not generate progeny virus, probably due to destabilization of the LTH (Fig. 4-6B). Subsequently, we restored base-pairing in the central part of the LTH stem by introducing compensatory mutations in the opposite side of the LTH stem. This new mutant, LS-LR, showed a surprisingly strong increase of sg mRNA synthesis, which reached levels similar to those of the wild-type control (Fig. 4-6B). This increase in sg mRNA synthesis sufficed to support the production of infectious progeny virus, which was further characterized in more detail. BHK-21 cells were transfected with an equal amount of full-length RNA transcript generated from the wild-type and LS-LR mutant plasmids. This resulted in an equal number of transfected cells and, in addition, immunofluorescence assays established efficient spread of both viruses within the same time frame after transfection (data not shown). Subsequently, after allowing infection of all cells, medium was harvested and plaque assays revealed similar virus titers and plaque morphology for LS-LR mutant and wild-type control (data not shown). After 4 passages, the 5'-proximal region of the genome of the LS-LR mutant was sequenced and found to be

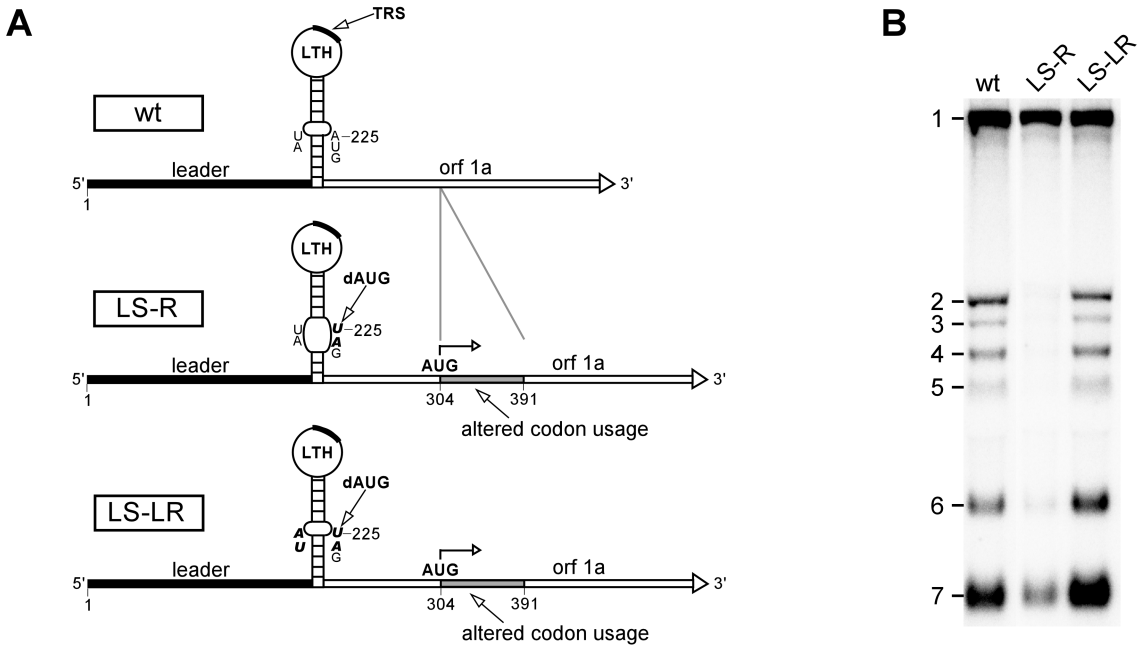


Figure 4-6. Mutagenesis of the lower part of the EAV LTH stem. *(A) Schematic representation of the mutants used. The 5'-proximal region of wild-type EAV genome is shown on top. LS-R contains two point mutations that partially open the LTH stem and simultaneously mutate the translation initiation codon of the replicase gene into UAG. In mutants LS-R and LS-LR, nt 225-304 were maintained as a noncoding RNA signal (Tijms et al. 2001) and replicase translation initiated at a new, engineered AUG codon at position 304-306. The complete EAV replicase gene was positioned downstream of this AUG codon, but the codon usage of the N-terminal 29 amino acids was altered to minimize the chance of recombination between homologous sequences up- and downstream of the novel AUG codon. By introducing two extra point mutations in the strand opposite the original AUG codon in the LTH stem, the base-pairing potential in the central region of the LTH was restored (mutant LS-LR). (B) Northern blot analysis of virus-specific RNA isolated at 16 h post transfection from BHK-21 cells transfected with wt, LS-R and LS-LR, respectively. The position of the viral mRNAs is indicated: 1, genomic RNA; 2 to 7, sg mRNAs.*

identical to that of the original mutant, suggesting that this mutant virus is genetically stable. The effects of LTH mutagenesis observed for the LS-R and LS-LR mutants, and the remarkable difference between these two viruses, strongly suggested that the LTH structure is important for EAV sg mRNA synthesis.

Prediction of an LTH structure in other arteriviruses and coronaviruses

One of hallmarks of nidoviruses is the production of a nested set of sg mRNAs in infected cells. Despite a number of similarities and the postulated common ancestry of the replicative enzymes involved, recent studies have uncovered differences between the mechanisms of subgenomic mRNA synthesis in different nidovirus subgroups. Ronivirus sg mRNAs were reported to lack a common 5' leader sequence (Cowley *et al.*, 2002) and this is also true for all but the largest of the sg mRNAs of the *Equine torovirus* (van Vliet

et al., 2002). Arterivirus and coronavirus sg mRNAs, on the other hand, both contain a common leader sequence, although its size differs substantially. To investigate whether the presence of an LTH is a common feature of nidoviruses that employ discontinuous RNA synthesis to attach a common leader sequence to each of their mRNAs, we extended our analysis to other arteriviruses and members of the genus *Coronavirus*.

An alignment of the leader regions of the four known arteriviruses (EAV, *Lactate dehydrogenase-elevating virus* (LDV), *Porcine reproductive and respiratory syndrome virus* (PRRSV), and *Simian hemorrhagic fever virus* (SHFV) did not reveal obvious sequence conservation (data not shown). Next, the structure of the 5'-proximal region of arterivirus genomes was predicted using the genetic algorithm implemented in the STAR program. For PRRSV we used six North American isolates and a single representative of the European genotype, which differs substantially from the North American isolates. Only one SHFV and two LDV sequences were available, which limited the possibilities to optimize the outcome of the analysis. The predicted structures for the other arteriviruses are depicted in Fig. 4-7. Interestingly, although a leader TRS-presenting hairpin was predicted for each virus, no other obvious similarities could be detected.

It was proposed that the differences in leader and TRS size between arteri- and coronaviruses might be the result of convergent instead of divergent evolution (van Vliet *et al.*, 2002). Evolving from a proto-nidovirus that lacked a common leader sequence, arteri-, corona- and toroviruses may have independently developed discontinuous transcription mechanisms to supply (some of) their mRNAs with a common leader sequence. Therefore, the presence of an LTH might be specific for arteriviruses only. On the other hand, a leader TRS presenting hairpin was previously predicted for *Bovine coronavirus* (BCoV), a prediction that was firmly supported by structure probing experiments (Chang *et al.*, 1996). Furthermore, a leader TRS-containing hairpin structure was previously predicted for MHV strain JHM (Shieh et al. 1987). To assess whether the presence of an LTH equivalent in the 5'-proximal region of the genome is a common feature of coronaviruses, we performed structure predictions for the leader TRS regions of seven different coronaviruses: *Infectious bronchitis virus* (IBV), *Transmissible gastroenteritis virus* (TGEV), *Human coronavirus* (HCoV) 229E, *Porcine epidemic diarrhea virus* (PEDV), *Sever acute respiratory syndrome-coronavirus* (SARS-CoV), BCoV, and MHV. For the latter virus, two strains (A59 and JHM) were included, which have two or three UCUAA repeats in their TRS region, respectively. Strikingly, for all of these the possibility of folding into similar hairpins was predicted, with the leader TRS region being located in the upper part of the hairpin (Fig. 4-8). For all coronaviruses, with the exception of PEDV, the structure presented in Fig. 4-8 was among the most stable structure alternatives according to the energy minimization program on the mfold web server.

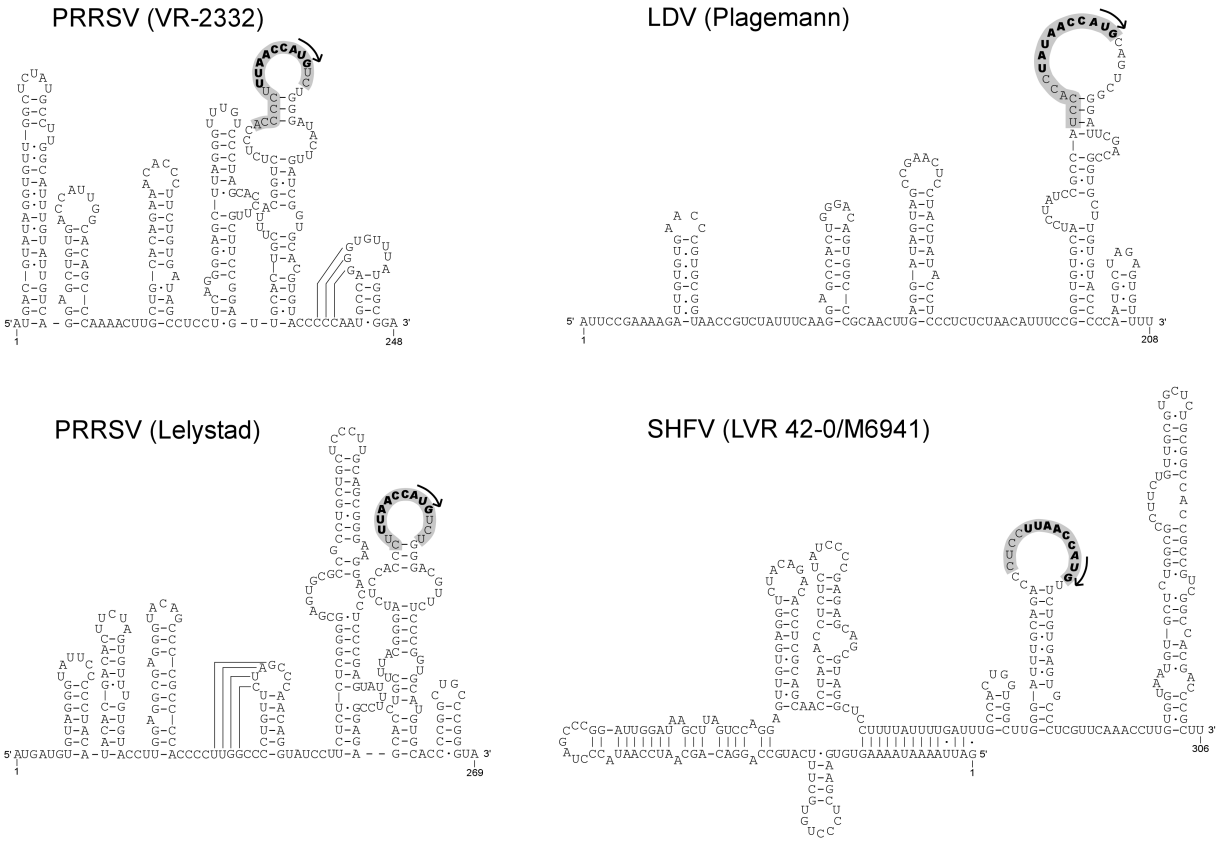


Figure 4-7. RNA secondary structure predictions of the 5'-proximal region of arterivirus genomes. *Presented are PRRSV, strains VR-2332 and Lelystad, LDV strain Plagemann and SHFV strain LVR 42-0/M6941. The replicase translation initiation codon is shown in bold italic font. The leader TRS region is indicated: conserved nucleotides are shown in bold font and nucleotides implicated in leader TRS-body TRS base-pairing are highlighted in gray (PRRSV-LV, Meulenberg et al. 1993; PRRSV-VR2332, Nelsen et al. 1999; SHFV, Godeny et al. 1998; LDV, Chen et al. 1993).*

Discussion

In this study we present an RNA secondary structure model of the 5'-terminal 313 nt of the EAV genome, a domain that was experimentally proven to contain all 5'-located signals required for genome translation, genome replication and sg mRNA synthesis (Tijms *et al.*, 2001). The model (Fig. 4-3A) contains a number of conserved 5'-proximal hairpins, a less conserved pyrimidine-rich region and a prominent leader TRS-containing hairpin (LTH; hairpin G). Previous structure predictions, based on computer modeling only, were limited to the region immediately flanking the leader TRS and revealed a similar LTH with an extended stem (van Marle *et al.*, 1999a). On the other hand, a previous model of (Kheyar *et al.*, 1998) did not include any sequences downstream of the leader TRS and therefore lacked the LTH structure since its 3' half is specified by sequences downstream of the common leader sequence that are unique to the genome (Fig. 4-1C). The latter observation may have important implications for EAV sg mRNA synthesis, since it means that the LTH can only be formed in the genomic plus strand and not in subgenomic plus strands. Assuming that the LTH is essential for leader-to-body fusion, this would imply that the 5' end of sg mRNAs cannot serve as a "target" for leader TRS-body TRS base-pairing. Consequently, in the model of Sawicki and Sawicki (Fig. 4-1B), the genome RNA would be the sole template for the discontinuous synthesis of minus-stranded subgenomic RNAs.

The structure presented for the EAV Bucyrus isolate in Fig. 4-3 was supported by phylogenetic data and structure probing experiments. Very comparable structures were predicted for other EAV strains, with most hairpins being identical to those of the reference isolate. However, it should be mentioned that computer analysis showed two conformations of the region surrounding the leader TRS. An extended LTH could be predicted at the expense of the hairpins F and H, which are directly adjacent to the LTH (Fig. 4-3). This structure (nt 164-249) corresponds exactly to the hairpin above the large bulge in the previously published large LTH structure (Van Marle et al. 1999). The phylogenetic and structure probing data appear to favor the short LTH version, but does not exclude the existence of the large hairpin. Although both conformations can be predicted for the other EAV strains, they adopt a somewhat different fold due to sequence variation. Therefore, the existence of alternative secondary structures for this region, which may also be influenced by the binding of protein factors, cannot be excluded. The question whether the short or the extended LTH (or both) are involved in sg mRNA synthesis is the subject of our ongoing studies.

We have also obtained experimental support for the importance of the LTH from mutagenesis experiments in which the LTH loop and stem were targeted. The loop was partly closed and re-opened in double mutant L-LR (Fig. 4-5). Likewise, the base-pairing in the upper and central parts of the LTH stem was disrupted and again restored by the introduction of compensatory mutations (Fig. 4-5 and 4-6). The results obtained with these mutants indicated that the LTH region probably also is important as an RNA

element in genome translation and/or replication, since the replication competence of many mutants was severely reduced (Fig. 4-5). In the double mutants L-LR and US-LR, restoration of the LTH loop or stem to its original conformation partially restored both genome replication and sg mRNA synthesis, underlining the importance of the structure as such, but not allowing us to discriminate between a specific function of the LTH in sg mRNA synthesis and other functions. This was different for the mutant presented in Fig. 4-6, in which disruption of the LTH did not affect genome replication while reducing sg mRNA synthesis to approximately 30 percent of the wild-type level. Following restoration of the base-pairing potential in the LTH stem in double mutant LS-LR, sg mRNA synthesis levels returned to normal and progeny virus could be recovered. In our opinion, the result obtained with the LS-R and LS-LR mutants provides evidence for a direct role of the LTH in subgenomic mRNA synthesis. Presumably, only an intact LTH structure is capable of properly presenting the leader TRS and making it accessible for the crucial body TRS-leader TRS base-pairing interaction (Fig. 4-1B).

It should be stressed that the 3'-proximal region of the EAV genomic minus strand is predicted to adopt a fold that is almost identical to that of its complementary sequence, the 5'-proximal domain of the plus strand. Although largely a mirror image of the plus strand model proposed in Fig. 4-3, the loop of the LTH complement may be partially closed by intraloop base-pairing (data not shown). The complement of the 5'-proximal hairpins present in the plus strand model (hairpins A to E) may serve as signals in the initiation of plus strand synthesis at the 3' end of the genomic and subgenomic minus strands. In this paper, we focused mainly on the putative role of the LTH in sg mRNA synthesis, but the results obtained with e.g. the mutants described in Fig. 4-5 make it likely that EAV RNA signals involved in different processes may indeed overlap in the 5'-proximal region of the EAV genome. The detailed characterization of these functions awaits the development of *in vivo* or *in vitro* assays that will allow the separation of processes like genome translation, replication, and sg mRNA synthesis.

All four arteriviruses are likely to employ the same mechanism for sg RNA synthesis and some preliminary evidence that the LTH is an important element in this mechanism is presented in this paper. Support for the relevance of the LTH comes from its identification in other arteriviruses (Fig. 4-7), despite the lack of convincing similarity at the level of primary structure. We previously proposed that the mechanism of nidovirus discontinuous sg RNA synthesis resembles similarity-assisted RNA recombination (Pasternak *et al.*, 2001). The 3'-proximal region of the genome, from which the nascent RNA strand is synthesized, can be considered as the donor strand, whereas the leader region is the acceptor RNA strand (Nagy & Simon, 1997). The large loop of the LTH in arteriviruses may be used to provide a single-stranded region and make the leader TRS accessible as an acceptor during template switching. Likewise, during RNA recombination of plant RNA tombusviruses junction sites in acceptor templates do not reside in double-stranded regions (White & Morris, 1995).The predicted coronavirus LTH-like structures contain a significantly smaller loop than their arterivirus counterparts, but on the other hand the

complementarity between coronavirus body TRSs and the corresponding leader TRS region is generally larger. Hence, the domain involved in base-pairing with the body TRS complement is predicted to extend into the stem region below the loop (depicted in gray in Fig. 4-8). At least part of the core of the coronavirus leader TRS resides in the loop and this may be sufficient to provide a single-stranded region for template switching. Future experiments will be aimed at the question whether the LTH acts as an independent RNA signal in subgenomic RNA synthesis by positioning the leader TRS in a conformation that is recognized by the replicase/transcriptase complex, a process in which protein factors are likely to play an important role as well.

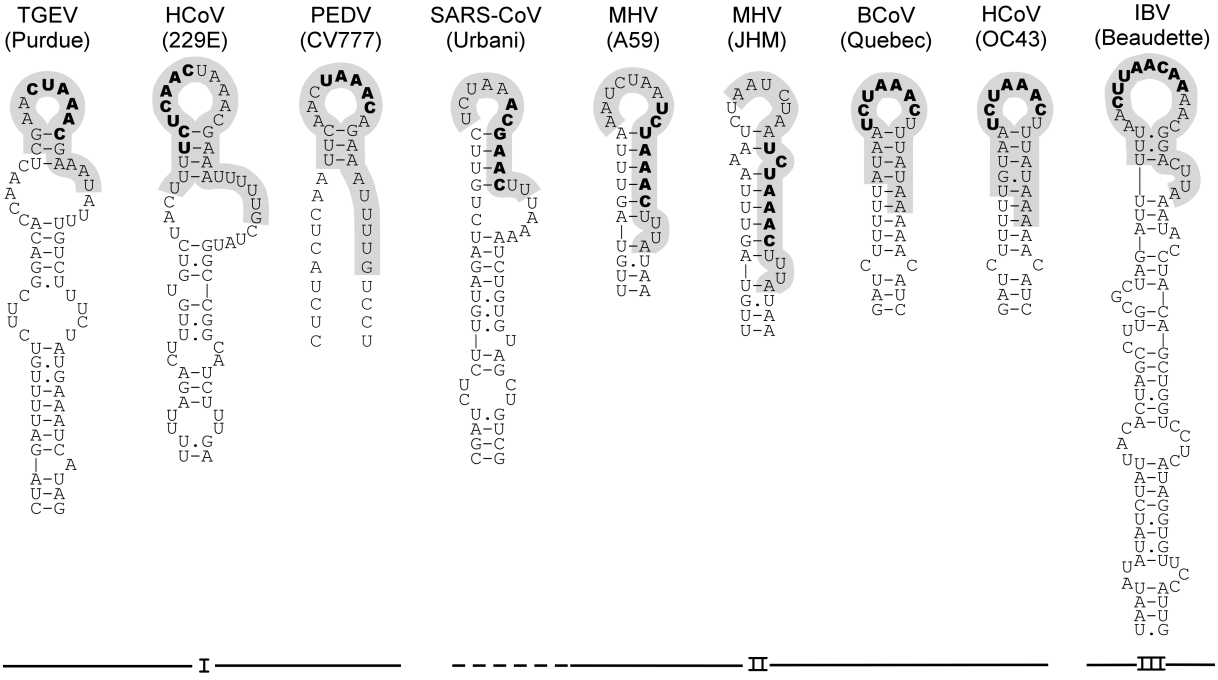


Figure 4-8. RNA secondary structure predictions of hairpins containing the leader TRS of a variety of coronaviruses. *Presented are TGEV strain Purdue, HCoV strain 229E, PEDV strain CV777, SARS-CoV strain Urbani, MHV strain A59, MHV strain JHM, BCoV strain Quebec and IBV strain Beaudette, respectively. The leader TRS region is indicated, with conserved nucleotides shown in bold font and the nucleotide stretch implicated in leader TRS-body TRS base-pairing is highlighted in gray (TGEV, Alonso et al. 2002; HCoV-229E, Herold et al. 1993; PEDV, Tobler and Ackermann 1995; SARS-CoV, Thiel et al. 2003; MHV-A59, Schaad and Baric 1993 (and references therein); MHV-JHM, Joo and Makino 1992; BCoV, Ozdarendeli et al. 2001; HCoV-OC43, database accession number AY391777; IBV, Stirrups et al. 2000). The structures are grouped according to the three coronavirus subgroups, which are indicated below.*

Materials and Methods

RNA transfection, infection and immunofluorescence analysis

Baby hamster kidney (BHK-21; ATCC CCL10) cells were maintained (van Berlo *et al.*, 1982) and used for RNA transfection experiments as described previously (van Dinten *et al.*, 1997). Immunofluorescence assays with EAV-specific antisera for nsp3 (98.E3) and nucleocapsid protein N (MAb 3E2) were done on transfected and infected cells according to (van der Meer *et al.*, 1998).

Virus strains and viral sequences

To obtain sequences of the 5'-proximal region of the EAV genome, including leader and downstream sequences, of EAV strains Bibuna, H6, Norway-2, Probe-S, S1368, S1512, Vienna and Wroclaw, virus was grown in BHK-21 cells. RNA was isolated 16 hours post infection according to the method of (Spaan *et al.*, 1981). Standard RT-PCR reactions were performed and products were directly sequenced (database accession numbers AJ488294 to AJ488301). Other arterivirus sequences were obtained from the database: EAV strain Bucyrus (NC002532); EAV strain 11958 (U65726); EAV strain 15492 (U65727); EAV strain 19933 (U65728); EAV strain 84-KY-A1 (U65724); EAV strain 86-NY-A1 (U46945); EAV strain 86-AB-A1 (U65723); EAV strain 87-AR-A1 (U46946); EAV strain T1329 (U65725); EAV strain ARVAC (AF001260); LDV strain Plagemann (U15146); LDV strain neurovirulent type C (NC002534); PRRSV strain VR-2332 (AY150564); PRRSV strain BJ-4 (AF331831); PRRSV strain P129 (AF494042); PRRSV strain CH-1a (AY032626); PRRSV strain HB-2(sh)/2002 (AY262352); PRRSV strain SP (AF184212); PRRSV strain Lelystad (M96262) and SHFV strain LVR 42-0/M6941 (AF180391). Coronavirus sequences: BCoV strain Quebec (AF220295), IBV strain Beaudette (NC001451), TGEV strain Purdue (NC002306), PEDV strain CV777 (NC003436), SARS-CoV strain Urbani (AY278741), HCoV strain 229E (NC002645), HCoV strain OC43 (AY391777), MHV strain A59 (NC001846), MHV strain JHM (M16615).

Sequence alignment and RNA structure prediction

Sequence alignments were created using the Clustal W algorithm incorporated in VectorNTI Suite v6.0. RNA secondary structures were predicted using the genetic

algorithm of STAR v4.4 (Gultyaev *et al.*, 1995) and the energy minimization program of mfold web server (Zuker, 2003).

Structure probing

Synthesis of a 589 nt-long RNA corresponding to the 5'-proximal region of the EAV genome was performed *in vitro* in 50 µl from *Mlu*I-linearized plasmid pU0026, which contains EAV nt 1-2611, using T7 polymerase (Promega). The RNA transcript was purified by acidic phenol (pH4.0) and acidic phenol/chloroform extraction in the presence of 0.2 M of NH_4OAc, twice precipitated with 2 volumes of isopropanol and 0.67 M NH_4OAc. Prior to chemical and enzymatic probing, essentially described by (Hsue *et al.*, 2000), RNA was denatured by incubation at 65°C for 10 minutes in the appropriate reaction buffer and renatured by slowly cooling down to 37°C.

For chemical modification of RNA with dimethyl sulfate (DMS; Fluka) or 1-cyclohexyl-3-[2-morpholinoethyl] carbodiimide metho-*p*-toluene sulfonate (CMCT; Aldrich), 1 µg of synthetic RNA was denatured in 12.5 µl of buffer containing 10 mM $MgCl_2$ and 80 mM sodium cacodylate pH7.2 or 50 mM sodium borate pH8.0, respectively. This total mix was added to the modification reactions. DMS modifications were performed in 50 µl of 80 mM sodium cacodylate, 5 mM $MgCl_2$, 100 mM KCl containing 0.25, 0.5, 0.75 and 1.0 % DMS. CMCT modifications were performed in 50 µl of 12.5 mM sodium borate pH8.0, 2.5 mM $MgCl_2$ and 12.5 mm KCl containing 4.2, 8.4, 12.6 and 16.8 mg/ml CMCT. After a 10-minute incubation at 37°C, the reactions were stopped by ethanol precipitation and the pellet was dissolved in 10 µl of water.

For enzymatic modification of RNA, 1 µg of synthetic RNA was denatured in 19 µl of 30 mM TrisHCl pH7.5, 20 mM $MgCl_2$ and 300 mM KCl. To this mix 1 µl of RNase was added. Incubation were performed at 37°C for 30 min with $5\cdot10^{-5}$, 10^{-4} or $5\cdot10^{-4}$ U of RNase A (Amersham Pharmacia Biotech); 10^{-1}, $5\cdot10^{-1}$ or 1 U of RNase T_1 (Invitrogen); 10^{-2}, $5\cdot10^{-2}$ or 10^{-1} U of RNase T_2 (Invitrogen); or $5\cdot10^{-4}$, 10^{-3} or $5\cdot10^{-3}$ U of RNase V_1 (Kemotex Bio Ltd.). The reactions were stopped by phenol/chloroform extraction and ethanol precipitation, after which the pellet was dissolved in 10 µl of water.

Primer extension

Three different antisense primers, EU078 (5'-AGTAAGAAAGGGAAC-3'), E125 (5'-CGCCATGCTCACACGCGTCGGGTAAG-3') and E564 (5'-GTCCCATGAACAAGCCAGG-3') were used to cover the complete 5'-terminal 313 nucleotide region of the EAV genome. Primers were ^{32}P end labeled and 0.5 pmol was incubated with 0.5 µl of modified or control RNA in a total volume of 10 µl for 10 min at

70°C and cooled down immediately on ice. Primer extension reactions were carried out in a total volume of 30 µl containing 6 µl of 5x first-strand buffer, 1 mM of deoxynucleoside triphosphates, 10 mM dithiothreitol, 20 U of RNase OUT and 100 U of Superscript II reverse transcriptase (all compounds from Invitrogen) for 45 min at 45°C. Reaction products were separated on standard 8 M urea-8% polyacrylamide gels. A sequence ladder was generated using a Ladderman Dideoxy Sequencing Kit (Takara) on DNA template pU0026 with the same end-labeled primer.

Mutagenesis of the EAV full-length cDNA clone

Mutations were introduced into shuttle plasmids by standard site-directed PCR mutagenesis. After sequence analysis, restriction fragments containing the desired mutations were transferred to the full-length clone pEAV211, a derivative of pEAV030 (van Dinten *et al.*, 1997) containing a few translationally silent, engineered restriction sites. Virus derived from clone pEAV211 has a wild-type phenotype, as determined by comparison with pEAV030-derived virus in growth curve experiments and plaque assays.

The LTH was mutagenized to alter its predicted secondary structure (Fig. 4-5A & 4-6A). To disrupt the upper part of the LTH stem (Fig. 4-5A), mutations were introduced into the 5' side of the stem (5'-GUCGU-3' → 5'-CUGCG-3') creating pUS-L, or 3' side of the stem (5'-GCGAC-3' → 5'-UGCAG-3'), creating pUS-R. The combination of these mutations in pUS-LR was predicted to restore base-pairing between both sides of the stem. The LTH loop was partly closed by introducing mutations in the 5' side of the loop (5'-CGAU-3' → 5'-AGGG-3'), creating pL-L, or the 3' side (5'-CCUU-3' → 5'-AUUG-3'), creating pL-R. The combination of these mutations in pL-LR was predicted to re-open the loop of the hairpin and restore it to its original size.

To partly disrupt the lower part of the LTH stem (LS-R; Fig. 4-6) the native start codon for translation (nt 225-227) was mutated to UAG and an alternative AUG was engineered at nt positions 304-306 to drive ORF1a/1ab translation. Downstream of the novel AUG codon, the nsp1-coding sequence was restored by a 87-nt insertion, but to reduce the chance of recombination between sequences up- and downstream of the novel AUG codon the codon usage for the N-terminal 29 amino acids of nsp1 was changed as much as possible (5'-AUG GCU ACG UUU AGU GCA ACA GGU UUC GGU GGC UCA UUC GUA CGC GAU UGG AGU UUA GAU CUU CCG GAU GCA UGC GAA CAC GGG GCC-3'; mutations are underlined). Subsequently, pLS-R was used to introduce two point mutations in the 5' side of the LTH stem to restore the base-pairing possibility in the lower part of the stem (construct pLS-LR; nucleotides $A^{190}U^{191}$ changed to UA).

RNA isolation and analysis

Intracellular RNA isolation was performed by using the acidic phenol method as described by (Pasternak *et al.*, 2000). RNA was separated in denaturing agarose-formaldehyde gels and detected with a ^{32}P-labeled probe E154 (5'-TTGGTTCCTGGGTGGCTAATAACTACTT-3'), which is complementary to the 3' end of the EAV genome and recognizes both genomic and subgenomic RNA.

Acknowledgements

We would like to thank Claudia Baule and Sandor Belák (The National Veterinary Institute, Uppsala, Sweden) for providing EAV isolates, Udeni Balasuriya and Jim Maclachlan (University of California, Davis, USA) for providing monoclonal antibodies and Kees Pleij (Leiden Institute of Chemistry, Leiden University, the Netherlands) for assistance with RNA structure probing. We are grateful to Sasha Pasternak and Richard Molenkamp for helpful discussions. E. van den Born was supported by grant CW 99-010 from the Council for Chemical Sciences of the Netherlands Organisation for Scientific Research.

Chapter 5

Discontinuous Subgenomic RNA Synthesis in Arteriviruses Is Guided by an RNA Hairpin Structure Located in the Genomic Leader Region

Erwin van den Born, Clara C. Posthuma, Alexander P. Gultyaev, and Eric J. Snijder
(2005)
Journal of Virology **79**: 6312-6324

Reprinted with permission

Abstract

Nidoviruses produce an extensive 3'-coterminal nested set of subgenomic (sg) mRNAs, which are used to express structural proteins and sometimes accessory proteins. In arteriviruses and coronaviruses, these mRNAs contain a common 5' leader sequence, derived from the genomic 5' end. The joining of the leader sequence to different segments derived from the 3'-proximal part of the genome (mRNA bodies) presumably involves a unique mechanism of discontinuous minus-strand RNA synthesis, in which base-pairing between sense and antisense transcription-regulating sequences (TRSs) plays an essential role. The leader TRS is present in the loop of a hairpin structure that functions in sg mRNA synthesis. In this study, the minimal sequences in the 5'-proximal region of the EAV genome that are required for sg RNA synthesis were delimitated through mutagenesis. A full-length cDNA clone was engineered in which this domain was duplicated, allowing us to make mutations and monitor their effect on sg RNA synthesis without seriously affecting genome replication and translation. The leader TRS present in the duplicated sequence was used and yielded novel sg mRNAs with significantly extended leaders. Our combined findings suggest that the LTH and its immediate flanking sequences are essential for efficient sg RNA synthesis and form an independent functional entity that could be moved 300 nucleotides downstream of its original position in the genome. We hypothesize that a conformational switch in the LTH region regulates the role of the 5'-proximal region of the arterivirus genome in subgenomic RNA synthesis.

Introduction

Many eukaryotic plus-strand RNA (+RNA) viruses generate subgenomic (sg) mRNAs as a strategy to regulate the expression of one or several viral proteins (Miller & Koev, 2000). In addition to internal initiation of translation, sg mRNA synthesis provides a mechanism to express genes positioned downstream of the 5'-proximal gene in polycistronic genomes of +RNA viruses of eukaryotes. Different mechanisms are employed by +RNA viruses to achieve the transcription of sg mRNAs (White, 2002). The generation of an extensive 3'-coterminal nested set of sg mRNAs to express structural and/or accessory proteins is a common property of viruses belonging to the order *Nidovirales* (*Coronaviridae*, *Arteriviridae*, and *Roniviridae*) (Spaan *et al.*, 1983b; de Vries *et al.*, 1997). In the case of arteriviruses and coronaviruses, all transcripts contain a common 5' sequence element, the so-called "leader", which is colinear with the 5'-proximal part of the genome and is fused to different regions derived from the 3'-proximal third of the genome, which are termed "mRNA bodies" (Fig. 5-1A). For two other nidovirus subgroups, roniviruses and toroviruses (with the exception of sg mRNA2 of the latter), sg mRNAs without a common 5' leader sequence have been described (Snijder *et al.*, 1990b; van Vliet *et al.*, 2002; Cowley *et al.*, 2002).

For sg RNAs of coronaviruses and arteriviruses, and also the largest sg RNA of toroviruses (van Vliet *et al.*, 2002), the fusion of the sg RNA body to the common leader sequence has been postulated to involve the discontinuous extension of minus-strand RNA synthesis, which presumably yields sg minus-strand templates that are used to produce the sg plus strands (Sawicki & Sawicki, 1995). After attenuation of RNA synthesis, the nascent minus strand is thought to be translocated to the 5'-proximal region of the genomic template, where RNA synthesis is resumed to add the complement of the genomic leader sequence, thus completing the minus-strand template for sg mRNA synthesis (Fig. 5-1B). Key elements in the joining of leader and body are short conserved transcription-regulating sequences (TRSs) that are present both at the 3' end of the leader sequence (leader TRS) and at the 5' end of each of the body regions in the 3'-proximal region of the genome (body TRSs; Fig. 5-1A). Base-pairing between the leader TRS (in the plus strand) and the complement of the body TRSs (in the nascent minus strand) has been shown to be essential for sg mRNA synthesis in arteriviruses and coronaviruses (Joo & Makino, 1992; van Marle *et al.*, 1999a; Pasternak *et al.*, 2001; Pasternak *et al.*, 2003; Zuniga *et al.*, 2004).

Many details of the mechanism that regulates the fusion of sequences that are non-contiguous in the genome during sg RNA synthesis of arteriviruses and coronaviruses remain to be elucidated. It was proposed that the process may mechanistically resemble similarity-assisted copy-choice RNA recombination (Chang *et al.*, 1996; Brian & Spaan, 1997; Pasternak *et al.*, 2001). As in the case of RNA recombination, the higher order structure of the template (and possibly also of the nascent strand) may play an important role, in particular in the regions where leader TRS and body TRSs reside. For *Equine*

arteritis virus (EAV), the arterivirus prototype, we previously predicted the presence of a hairpin structure in the 5'-proximal region of the genome that contains the leader TRS in its loop (van Marle *et al.*, 1999a). Subsequently, the presence and importance of this leader TRS hairpin (LTH) was supported by phylogenetic analysis, RNA structure probing experiments, and a limited mutagenesis study (Fig. 5-1C) (van den Born *et al.*, 2004). Although the latter analysis was hampered by side effects of the mutations at the level of genome replication, the proposed role of the LTH in sg mRNA synthesis was generally supported by the fact that altering its conformation decreased the synthesis of sg mRNAs, while the introduction of compensatory mutations restored sg mRNA production to a certain extent. The identification of potential LTHs in predicted structures of the 5'-proximal genome regions of other arteriviruses and coronaviruses supported the importance of the structure for the base-pairing interaction between the sense leader TRS and antisense body TRS (Chang *et al.*, 1996; van den Born *et al.*, 2004). Furthermore, the proposed common ancestry of the viral enzymes involved suggested that common principles may underlay the mechanisms for sg RNA synthesis in both virus groups (den Boon *et al.*, 1991).

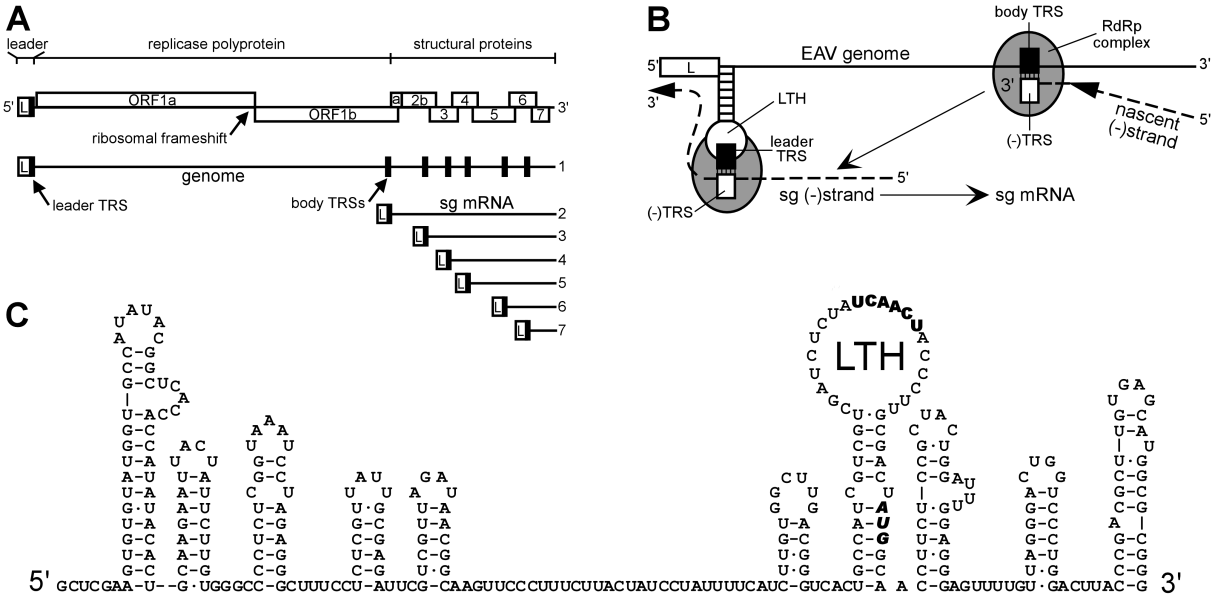


Figure 5-1. *(A) Schematic overview of the genome organization and expression of EAV. The replicase gene (ORFs 1a and 1b), the structural protein genes (2a, 2b, 3, 4, 5, 6 and 7) and the leader (L) are indicated. The black boxes in the genomic RNA indicate the position of the leader TRS and (major) body TRSs. The coterminal nested set of sg mRNAs 2-7 is depicted below the genome. (B) The discontinuous minus strand extension model for nidovirus sg RNA synthesis (Sawicki & Sawicki, 1995). After attenuation of the RdRp complex at a body TRS in the plus strand genomic template, the nascent strand is translocated to the leader TRS region. Following (-) body TRS to (+) leader TRS base-pairing, minus strand RNA synthesis resumes to add the complement of the leader sequence. Subsequently, the sg minus strand serves as a template for sg mRNA production. (C) RNA secondary structure model of the 5'-proximal 313 nt of the EAV RNA genome (van den Born et al., 2004). The leader TRS and the replicase translation initiation codon are depicted in bold font.*

Here, we have focused on the role of the 5'-proximal region of the EAV genome in sg RNA synthesis. This region, including the leader TRS and LTH, was successfully duplicated in the context of an EAV full-length cDNA clone, allowing the introduction of mutations in signals for sg RNA synthesis without notably interfering with genome replication and translation. Rigorous changes in the 5'-proximal domain of the genome did not result in apparent defects at the level of genome replication, implying that also genome translation was not significantly affected. Still, an analysis of the translation of RNAs carrying wt or mutant EAV 5’ genomic sequences suggested that the virus probably conforms to the conventional 'ribosome scanning' model of translational initiation. A LTH deletion mutagenesis study revealed that the LTH and its immediate flanking sequences are required for efficient sg RNA synthesis. This region can fold into an extended conformation of the previously described LTH, which may function as an independent entity that regulates the production of all sg mRNAs and is hypothesized to be involved in a conformational RNA switch in the 5'-proximal region of the genome.

Results

Leader-to-body fusion can occur at a different site in the EAV genome

In a recent mutagenesis study, the loop and stem regions of the EAV LTH were targeted to disrupt their primary and secondary structure. It was found that mutants in which the LTH conformation was predicted to be maintained showed the highest level of sg RNA synthesis. From this analysis, the EAV LTH was concluded to be generally important for transcription (van den Born *et al.*, 2004), but it should be noticed that the analysis of many of the mutants in this study was hampered by side effects at the level of genome replication, for which signals apparently (and not unexpectedly) overlap with those for sg RNA synthesis.

If the LTH would be an independent entity in sg RNA synthesis, functioning as a sort of acceptor "hot spot" for a copy-choice recombination-like process, it might be possible to move it to a different position in the genome while maintaining its function in leader-to-body fusion. Due to the compact organization of the EAV genome, locations where such an insertion would not interrupt one of the viral genes were not readily available. Thus, we chose to duplicate the 5'-proximal 301 nucleotides of the genome, which were previously found to suffice for efficient genome replication and sg RNA synthesis (Tijms *et al.*, 2001). The duplicated sequence ("Duplicated 5'-proximal Domain" or D5D) consisted of the 5' untranslated region and the first 29 codons of the replicase gene (nsp1-coding region) and was inserted at nucleotide position 306 of the EAV genome. In the construct with this 5'-proximal tandem repeat (D5Dwt; Fig. 5-2A), the native leader TRS and replicase translation initiation codon were inactivated (UCAACU to AGUUGU and AUG to UAG, respectively; see Materials and Methods). Thus, translation initiation should now occur from the D5D AUG codon (at nt 530 instead of 224 of the genome). For leader-to-body fusion, the D5D LTH was available and its use should result in addition of a 516- instead of 211-nt common leader sequence to all sg RNAs.

Construct D5Dwt was tested in a transfection experiment and immunofluorescence assays showed that transfection efficiencies were comparable to those of the wild-type (wt) control and that this construct was able to replicate its genome RNA. In addition, nucleocapsid protein was readily detectable, which is indicative for the synthesis of sg mRNA7 (data not shown). A hybridization analysis on total intracellular RNA isolated at 15 h p.t. indeed confirmed the synthesis of a complete set of sg RNAs that were each approximately 300 nt longer than those of the wt control (Fig. 5-2C; compare the two leftmost lanes). To exclude that input RNA contributed to the genomic RNA signal in hybridizations, the replication deficient SGA mutant (van Dinten *et al.*, 1999), carrying a fatal amino acid substitution in the RdRp domain, was included in the analysis. Even after

prolonged exposure, the transfected RNA of this mutant could not be detected (data not shown).

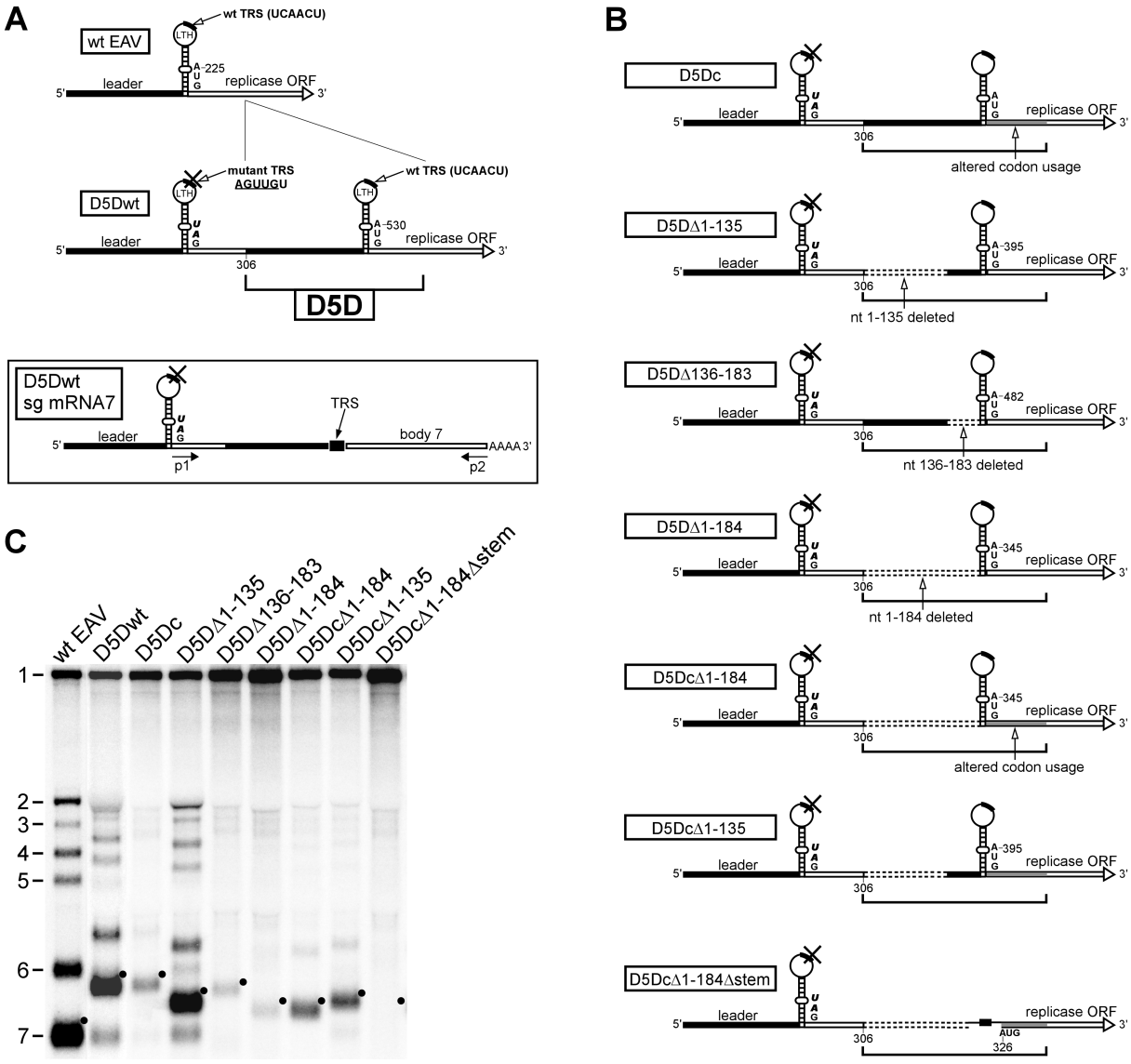


Figure 5-2. Duplication of 5'-proximal sequences in the EAV genome as a strategy to delimitate sequences required for sg RNA synthesis. *(A) Schematic representation of the 5'-proximal region of the genomes of wt virus and construct D5Dwt. The sg mRNA 7 transcript produced by D5Dwt is shown below, with the location of RT-PCR primers used to amplify the leader-to-body junction region of this transcript. The RNA region corresponding to the leader sequence is depicted as a black bar; replicase-coding sequences are depicted as a white bar. Replicase translation initiation codon, nucleotide numbers (counting from the 5' end of the genome), LTH, and the D5D domain are indicated. (B) Schematic overview of the 5'-proximal region of the genome of D5Dwt-derivatives carrying internal deletions in the D5D domain. When the codon usage of the replicase-coding region was altered, this is depicted as a gray bar. (C) Hybridization analysis of virus-specific RNA isolated at 15 h p.t. from BHK-21 cells transfected with wt RNA, D5Dwt RNA or RNA transcribed from the pD5Dwt-derivatives depicted in panel B. The position of the mRNAs of wt virus is indicated: 1, genomic RNA; 2 to 7, sg mRNAs. The sg mRNA7 bands are indicated with a black dot.*

Because the efficiency of D5Dwt and wt virus genome replication was very similar (see below), the ratio of subgenomic RNA versus genome RNA could be used to measure the transcriptional capability of the D5Dwt mutant. Quantitation of viral RNAs following hybridization analysis revealed that transcription levels for this mutant reached about 70% of those of the wt virus. The use of the LTH duplicate for leader-to-body fusion was verified by a specific RT-PCR analysis, which confirmed the presence of an extended sg mRNA7 that was not present in cells transfected with the wt control. Direct sequence analysis of the PCR product confirmed that in this novel sg mRNA the leader-to-body fusion indeed mapped to the leader TRS in the LTH duplicate (Fig. 5-2A; data not shown).

Not unexpectedly, the tandem repeat of two almost identical ~300-nt sequences in D5Dwt was prone to homologous recombination, rapidly leading to wt recombinant virus and explaining the background of wt-size sg mRNAs in the hybridization analysis (Fig. 5-2C). Nevertheless, our data confirmed that the 5'-proximal 301 nt of the EAV genome contains essential signals for efficient sg mRNA synthesis and that these signals can be moved 300 nt more distal from the genomic 5' end without a profound loss of function.

The flanking sequences of the LTH are required for sg RNA synthesis

The amount of D5Dwt genomic RNA detected in hybridization analysis (Fig. 5-2C) indicated that its replication, and therefore probably also its translation (see below), were not notably altered compared to the wt genome, a remarkable finding given the more than 2-fold extension of the 5’ UTR of the viral genome. Consequently, D5Dwt offered an excellent opportunity to analyze the minimal sequence requirements for EAV sg RNA synthesis. Construct pD5Dwt was used as a starting point to make deletions in the D5D sequence, although straightforward deletions downstream of the LTH could not be made, since the replicase-coding region overlaps with the 3’ side of the LTH stem. To circumvent this problem, the codons in the D5D that specified the N-terminal 29 amino acids of nsp1, the N-terminal subunit of the replicase, were altered in some of the mutants. This allowed us to investigate the presence of RNA signals in the sequence downstream of the LTH, without changing nsp1. The corresponding mutant D5Dc (Fig. 5-2B), which combined a maximum number of translationally silent mutations in the 5’ end of the replicase gene with an intact LTH structure (see Materials & Methods), produced only low amounts of sg mRNA (Fig. 5-2C), suggesting that RNA sequences downstream of the LTH are relevant for sg RNA synthesis.

To assess the function of the sequences upstream of the LTH, a set of three deletion mutants was engineered. The borders of the deletions were based on the previously established RNA secondary structure (van den Born *et al.*, 2004), which defined a 5'-proximal domain consisting of five hairpins followed by a single-stranded polypyrimidine-rich stretch, which in turn is followed by the hairpin that flanks the LTH

at its 5' side (Fig. 5-1C). Remarkably, upon deletion of the first domain (corresponding to nt 1-135 of the genome) from construct D5Dwt, sg RNA levels were somewhat increased compared to those of D5Dwt (mutant D5DΔ1-135; Fig. 5-2B). As expected, the sg mRNAs of D5DΔ1-135 were slightly smaller due to its smaller leader sequence compared to D5Dwt (Fig. 5-2C). This result indicated that nucleotides 1 to 135 of the D5D sequence were dispensable for transcription. On the contrary, the deletion of the domain comprising the single-stranded stretch and the hairpin upstream of the LTH blocked efficient sg RNA synthesis (mutant D5DΔ136-183; Fig. 5-2). A combination of both deletions (mutant D5DΔ1-184), which removed the entire D5D sequence upstream of the LTH, had a similar effect on transcription (Fig. 5-2C).

The above results suggested that flanking sequences on either side of the LTH duplicate are important for efficient sg RNA synthesis. Consequently, we expected that removal of sequences on both sides of the LTH would completely abolish transcription. Surprisingly, however, mutant D5DcΔ1-184 (Fig. 5-2B) produced moderate levels of sg mRNAs, exceeding the amounts produced by mutants D5DΔ1-184 and D5Dc, which lacked sequences upstream or downstream of the LTH duplicate, respectively (Fig. 5-2C). These paradoxical results may be explained by the fact that the RNA structures required for transcription are less disturbed in this mutant (see also Discussion). Alternatively, the presence of RNA structures, induced by the unnaturally rearranged sequences, may influence LTH structure or function. The important role of sequences downstream of the LTH was confirmed by altering the replicase codons in this region in mutant D5DΔ1-135, which showed the highest level of transcription in this deletion study so far (Fig. 5-2B; mutant D5DcΔ1-135). As in the case of the D5Dwt/D5Dc pair, mutations in this region were found to reduce sg RNA transcription (Fig. 5-2C).

To this point, our analysis indicated that the leader TRS in the context of the LTH alone could be utilized for leader-to-body fusion to a limited extent only, and that up- and downstream sequences clearly enhanced the efficiency of this process. To directly assess the role of the LTH structure itself, the stem region was removed from mutant D5DcΔ1-184 to obtain mutant D5DcΔ1-184Δstem (Fig. 5-2B). Although the entire 21-nt LTH loop sequence, including the leader TRS, was retained in this construct, hybridization analysis revealed that this deletion abolished sg RNA synthesis completely (Fig. 5-2C). In addition, by sensitive RT-PCR analysis only trace amounts of sg mRNA could be detected (data not shown). This finding clearly supported the importance of the LTH itself and its postulated role in sg RNA synthesis.

The minimal region required for efficient sg RNA synthesis may fold into an extended version of the LTH

During our recent RNA secondary structure analysis of the 5'-proximal 313 nucleotides of the EAV genome, the equivalent of the D5D sequence, we came across two

possible conformations of the LTH, differing in the length of their stem (Fig. 5-1C and 3A) (van den Born *et al.*, 2004). The formation of the extended conformation of the LTH (eLTH) involves the base-pairing of sequences that flank the LTH and, since the latter were found to be necessary for efficient sg RNA synthesis (see above), the eLTH conformation may in fact represent the structure that guides discontinuous RNA synthesis.

Experimental support for this hypothesis was obtained when the upstream flanking sequences of the LTH region in mutant D5DcΔ1-184 were extended in the 5' direction. This was done in a derivative of D5DcΔ1-184 (D5DcmΔ1-184) in which the 3' side of the LTH had been rendered noncoding by moving the replicase translation initiation codon 27 nt downstream (Fig. 5-3A). This change somewhat impaired sg RNA synthesis of D5DcΔ1-184 further (Fig. 5-3B), but allowed the free mutagenesis of LTH-flanking sequences (see below) without affecting the replicase ORF. Remarkably, the inclusion of additional flanking sequences at the 5' side of the LTH (corresponding to nt 164-184 of the genome; Fig. 5-3A) in construct D5DcmΔ1-163 dramatically increased the synthesis of sg mRNAs (compare the two rightmost lanes of Fig. 5-3B) to levels comparable to those of D5Dwt (Fig. 5-2C).

The data presented in the previous paragraph suggested that the formation of the eLTH structure might be relevant or, alternatively, that the sequence of nt 164-184 of the genome was important for sg RNA synthesis. These possibilities were further investigated by mutagenesis of sequences predicted to be involved in formation of the lower part of the eLTH stem (Fig. 5-3A). In view of its excellent transcriptional activity, construct D5DcmΔ1-163 was used as starting point for this analysis. Two regions in the predicted stem extension (A and B in Fig. 5-3A) were changed and in addition the entire predicted stem-extending region was targeted (C in Fig. 5-3A). All mutants were analyzed in three independent experiments, which revealed that genome replication was not affected by the introduction of mutations in this region of the RNA. Thus, we could use the ratio of sg mRNA7 (the most abundant sg transcript) versus genome RNA to quantify transcription levels (Fig. 5-3C).

The predicted 4-bp stem A, just below the original LTH, was disrupted by changing either its 3' or its 5’ side to the sequence present at the opposite side, resulting in mutants D5DcmΔ1-163\AR and D5DcmΔ1-163\AL, respectively (Fig. 5-3A). Both mutations decreased transcription levels 4 to 5 times compared to the parental constructs (Fig. 5-3C). By combining these mutations in double mutant D5DcmΔ1-163\ALR, the base-pairing potential was restored and this resulted in a 1.5 to 2-fold increase in sg RNA synthesis compared to that of the single mutants, although the original transcription levels were not reached (Fig. 5-3C). The large stem B region was subjected to a similar analysis (Fig. 5-3A and 3C), but transcription levels of single mutants D5DcmΔ1-163\BR and D5DcmΔ1-163\BL were only slightly affected compared to the level achieved by the parental construct. Transcription levels of double mutant D5DcmΔ1-163\BLR were in the range of that of the single mutants (Fig. 5-3C). Finally, the entire predicted eLTH stem

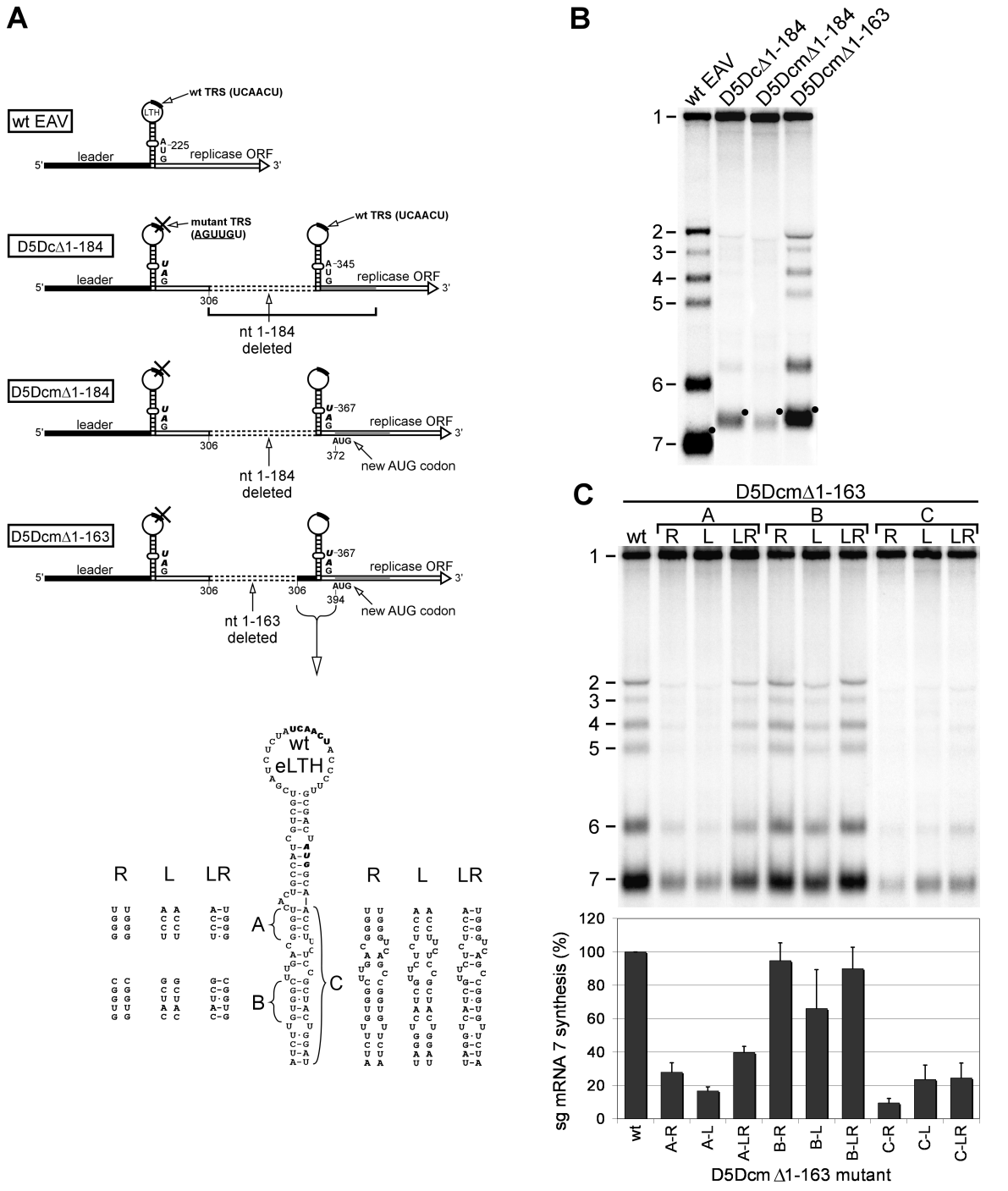


Figure 5-3. Analysis of the relevance of a LTH structure with extended stem for EAV sg RNA synthesis. *(A) Schematic representation of the constructs that revealed the importance of LTH-flanking sequences for optimal transcriptional activity. See also the legend to Fig. 2. Construct D5DcmΔ1-163 was used to test the importance of the eLTH by site-directed mutagenesis of the indicated region A, B and C (see text). (B) Hybridization analysis of virus-specific RNA isolated at 15 h p.t. from BHK-21 cells transfected with wt RNA or constructs D5DcΔ1-184, D5DcmΔ1-184, and D5DcmΔ1-163. Note the striking upregulation of sg RNA synthesis in the latter construct. (C) Hybridization analysis of virus-specific RNA isolated at 15 h p.t. from BHK-21 cells transfected with mutants in which regions A, B, and C of the eLTH stem were targeted (see panel A). For each region, the right side (R), left side (L), or both sides (LR) of the predicted stem were mutated, with potential restoration of base-pairing occurring in each of the LR mutants. The bar graph below represents the quantitation of the intensity of the sg mRNA 7 bands, which were normalized to the intensity of the genome band.*

(stem C; Fig. 5-3A) was altered, while keeping the bulges intact. Opening the stem by changing its complete 3' side (mutant D5DcmΔ1-163\CR), reduced transcription to approximately 10% of the level of the control construct, whereas changing the opposite side of stem C (D5DcmΔ1-163\CL) reduced it to approximately 20% (Fig. 5-3C). The combination of these mutations in mutant D5DcmΔ1-163\CLR resulted in a 2-fold increase of sg RNA synthesis compared to D5DcmΔ1-163\CR, but no significant upregulation was observed compared to D5DcmΔ1-163\CL.

Taken together, the data presented above demonstrated that the LTH-surrounding sequences clearly influence sg RNA synthesis and that in particular stem A (Fig. 5-3A) may contribute to the formation of an extended stem at the base of the originally proposed LTH structure.

Subgenomic mRNA synthesis requires an intact LTH

In a previous mutagenesis study (van den Born *et al.*, 2004), aimed at providing experimental support for the importance of the LTH structure, many of the mutations introduced into the wt genome severely affected genome replication, thus preventing a straightforward interpretation of the results with regard to sg RNA synthesis. The D5D construct and its derivatives, as described in the previous paragraphs, provided a solid alternative to study transcription-related RNA signals. Thus, previously engineered LTH stem and loop mutations were transferred to construct D5DΔ1-163 (Fig. 5-4A), in which nt 164-301 of the genome were duplicated.

As expected, all D5DΔ1-163-derived LTH mutants displayed efficient genome replication. Subgenomic RNA synthesis, on the other hand, was seriously affected by mutations that partly closed the LTH loop (Fig. 5-4; D5DΔ1-163\L-R and D5DΔ1-163\L-L) and was almost blocked when the upper part of the LTH stem was destabilized (Fig. 5-4; D5DΔ1-163\US-R and D5DΔ1-163\US-L). The combination of these sets of mutations in one construct, restoring either the loop size (mutant D5DΔ1-163\L-LR) or the base-pairing possibilities in the upper part of the stem (mutant D5DΔ1-163\US-LR), rescued sg RNA synthesis to levels close to that of the parental construct (Fig. 5-4B). In conclusion, the results confirmed and extended our previous mutagenesis study performed in the context of a wt full-length clone, which suggested that an intact LTH has an important function in sg RNA synthesis.

Replication competence and stability of mutants with duplicated 5'-proximal sequences

Potentially, all mutants described in the previous paragraphs were genetically unstable due to the presence of direct repeats in the 5'-proximal region of their genome. This was immediately apparent when long sequence stretches had been duplicated, like the entire 5'-proximal 301 nucleotides of the genome in the case of D5Dwt (Fig. 5-2C). Via a single crossover, the genome of this kind of mutants could be restored to wt sequence. The fact that wt-size sg mRNAs could be observed in a first cycle analysis strongly suggested the independent occurrence of many of such recombination events.

The contribution of recombination-derived wt virus to the genomic RNA detected in hybridizations was assessed in three experiments by using primer extension analysis on intracellular RNA isolated from cells transfected with one of a selected set of constructs: the wild-type control, D5Dwt, D5Dc and D5DcmΔ1-163. In RNA samples from cells transfected with D5Dwt or D5Dc, both containing the largest duplication used in this study, approximately 10% of the genome RNA originated from recombinant wt virus (data not shown). In contrast, wt-size genome RNA was not detected for mutant

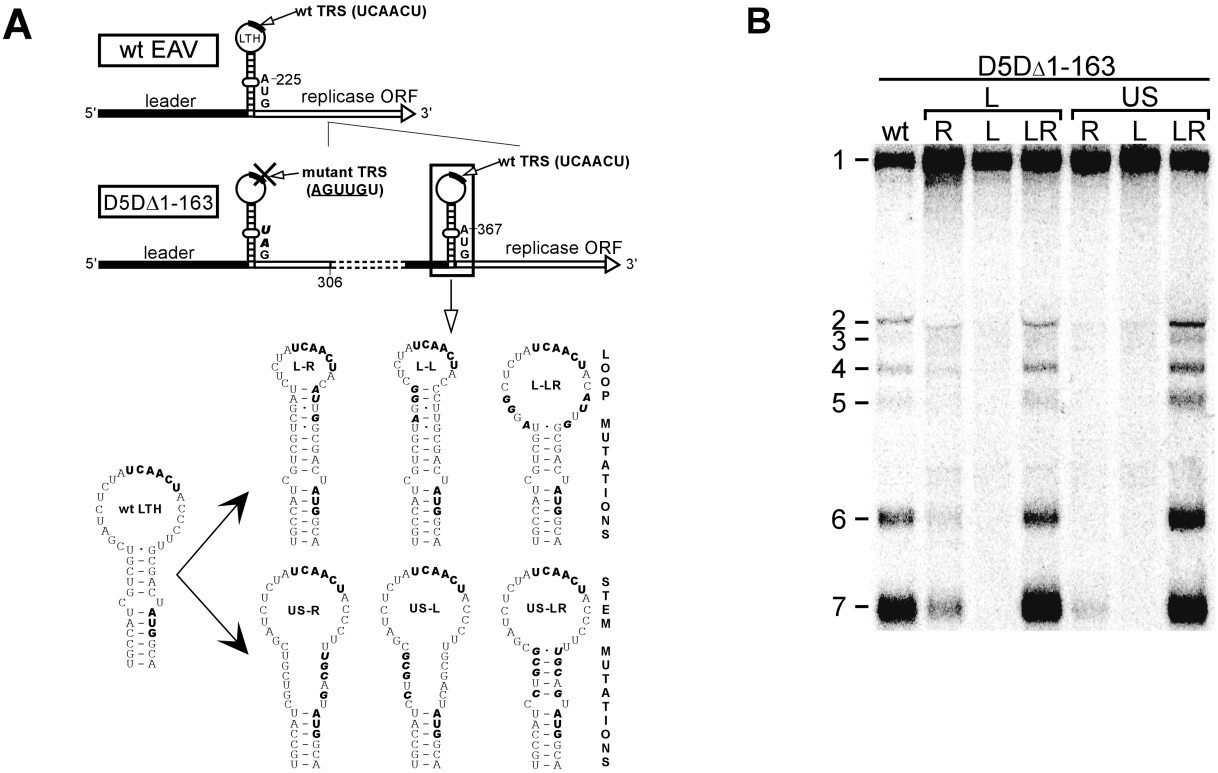


Figure 5-4. Site-directed mutagenesis underlines the importance of the EAV LTH structure for sg RNA synthesis. *(A) Schematic representation of the mutations introduced into the LTH structure of the parental construct pD5DΔ1-163. See also the legend to Fig. 2. (B) Hybridization analysis of virus-specific RNA isolated at 15 h p.t. from BHK-21 cells transfected with D5DΔ1-163 RNA or RNA from the mutants depicted in panel A.*

D5DcmΔ1-163, suggesting that recombination events were rare (data not shown). Subsequently, the same RNA was subjected to hybridization analysis and the amount of genomic RNA was quantified and corrected for transfection efficiencies. The three mutants were found to produce an equal or even larger amount of genomic RNA than the wt virus, indicating that genome replication was not affected (data not shown).

The properties of the most stable mutant, D5DcmΔ1-163 (Fig. 5-3), were evaluated in five subsequent passages and its growth characteristics were compared to those of the wt virus. This mutant harbored the smallest duplication that still allowed the same level of sg mRNA synthesis as in the original D5Dwt construct (Fig. 5-2). BHK-21 cells were transfected with *in vitro* transcribed D5DcmΔ1-163 RNA and the cell culture medium was harvested at 24 h p.t. and used for serial undiluted passaging on fresh BHK-21 cells for four subsequent rounds. Hybridization analysis of intracellular RNA isolated from these cell cultures showed that mutant D5DcmΔ1-163 was relatively stable and that after 5 passages the majority of the sg mRNAs were still derived from mutant genomes (Fig. 5-5A). In addition, the replication kinetics of mutant D5DcmΔ1-163 was compared to that of the wt virus and it showed no apparent delay in virus production, although the overall virus yield after the first infection cycle (~15 h p.i. at 37°C) was about 10- to 100-fold reduced (Fig. 5-5B).

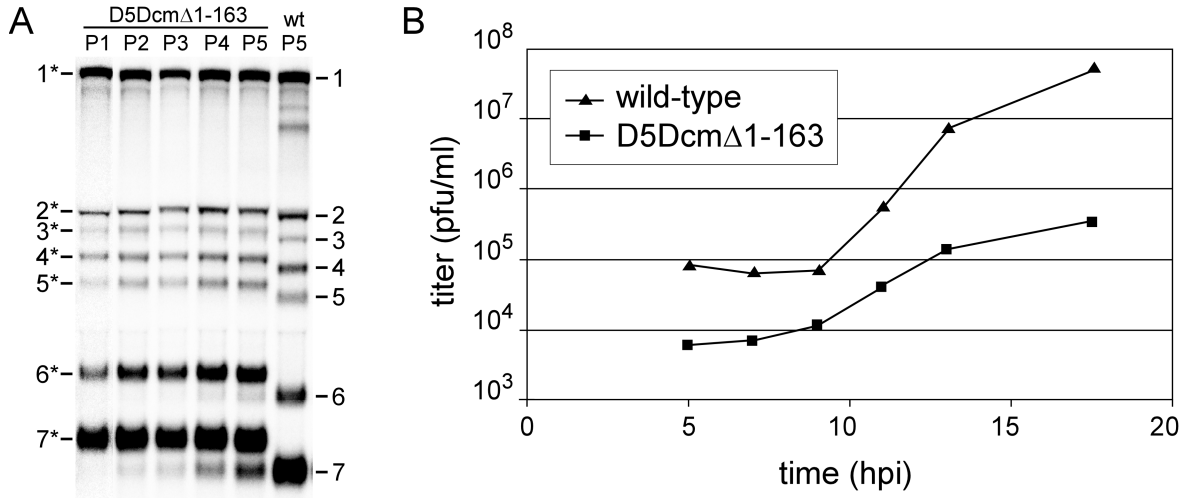


Figure 5-5. Characterization of the stability and growth of D5DcmΔ1-163 virus, a viable virus mutant with an extended common leader sequence on all its mRNAs. *(A) Gel hybridization analysis of D5DcmΔ1-163-specific RNA isolated at 15 h p.t. (P1) or post infection (P2-P5) from BHK-21 cells. After transfection of BHK-21 cells, medium was harvested at 24 h p.t. and subsequently passaged four times in undiluted form. RNA from a wt control is shown in the rightmost lane. The positions of the wt and D5DcmΔ1-163 (indicated with an asterisk) mRNAs are indicated. (B) Growth curve of plaque purified wt and D5DcmΔ1-163 virus. BHK-21 cells were infected with an m.o.i. of 5, medium was harvested at the indicated time points, and virus titers were subsequently determined by plaque assays.*

Efficient synthesis of a double nested set of mRNAs

In all constructs described thus far, the authentic leader TRS has been inactivated to avoid production of wt sg mRNAs. In a construct carrying a double wt leader TRS, each presented in the context of an LTH structure, we anticipated competition for leader-to-body fusion events. Such a construct would thus provide an indication of how the functionality of the LTH duplicate is related to that of the wt structure.

Construct D5DcmΔ1-163 was used to address this question, because the sizes of the (predicted) sets of sg mRNAs were suitable to achieve good separation upon agarose gel electrophoresis. Construct D5DcmΔ1-163\2xwt was created by restoring the authentic leader TRS to its wt 5'-UCAACU-3' sequence (Fig. 5-6A). The presence of two wt leader TRSs (the authentic one and the copy in the D5D sequence) indeed resulted in the synthesis of a perfect double set of 6 sg mRNAs each (Fig. 5-6B). Surprisingly, compared to construct D5DcmΔ1-163, the transcription level of the set of "larger" mRNAs was hardly affected by the activation of the production of the set of "smaller" mRNAs. The production of the latter set, in turn, was quite efficient when compared to the synthesis of the same mRNAs by the wt virus (Fig. 5-6B). Despite a small reduction in genome replication (to about 80% of wt levels), the total amount of sg mRNA produced by D5DcmΔ1-163\2xwt was similar to that of the wt virus and both leader TRSs yielded approximately equal amounts of sg mRNAs. This underlined once again that the LTH in the D5D sequence element indeed is an independent and efficient alternative signal for leader-to-body fusion during EAV sg RNA synthesis.

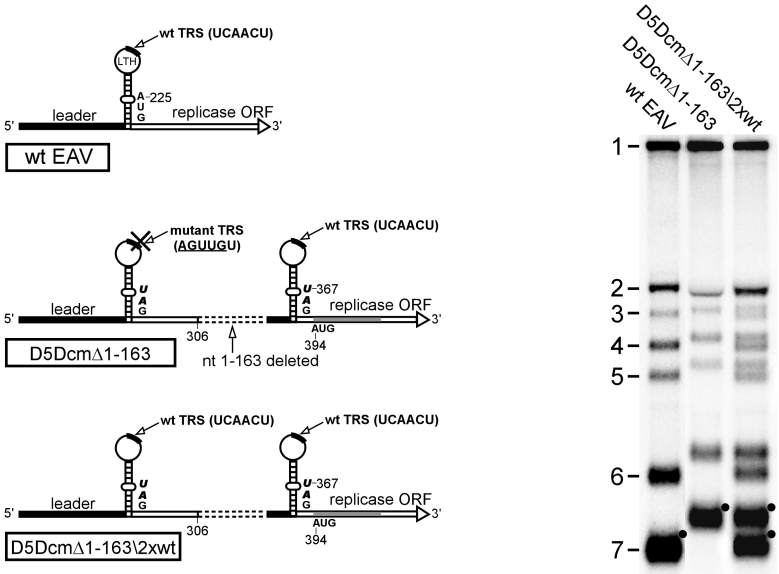


Figure 5-6. A D5D derivative with two wt LTH domains, including two wt leader TRSs, produces a perfect double nested set of sg mRNAs. *A schematic representation of the wt genome and constructs D5DcmΔ1-163 and D5DcmΔ1-163\2xwt is shown on the left. See also the legend to Fig. 5-2. On the right a hybridization analysis of intracellular RNA at 15 h p.t. is shown.*

Initiation of EAV genome translation

At 223 nt, the 5’ untranslated region (5’ UTR) of the wt EAV genome is relatively long. Besides, it contains a short “intraleader” ORF (L-ORF) of 37 codons, starting with an AUG codon at nt 14-16. In view of its proximity to the 5’ end of the genome, it is unclear whether this ORF is actually translated. Its product - if any - was previously found to be dispensable for replication in cell culture (Molenkamp *et al.*, 2000c). A remarkable finding during the analysis of the mutants described in the previous paragraphs was the fact that the 5’ UTR of the EAV genome could apparently be extended to 529 nt without a notable effect on replicase expression, a conclusion that was based on Western blot analysis of nsp1 expression in transfected cells and immunofluorescence assays (data not shown). In addition, some of these mutant 5' UTRs now contained two copies of the L-ORF upstream of the replicase translation initiation codon, with the second one no longer being positioned close to the 5’ end of the genome.

The presence of a cap structure has been reported for another arterivirus (Sagripanti *et al.*, 1986) and cap analogue is known to be an essential component of *in vitro* transcription reactions to make infectious full-length RNA (van Dinten *et al.*, 1997; Meulenberg *et al.*, 1998). Although this suggests that translation is initiated via a cap-dependent mechanism, the observations summarized above prompted us to investigate the mechanism of EAV genome translation in more detail. The possibility of translation initiation via an IRES element, a well-known mechanism proposed to be used by several virus groups (Vagner *et al.*, 2001) and a growing list of cellular mRNAs (Hellen & Sarnow, 2001), was evaluated in comparison to classical cap-dependent initiation followed by linear ribosome scanning, with initiation occurring at the most 5'-proximal AUG codon (Kozak, 1989; Kozak, 2002).

IRES activity of the EAV 5’ UTR (nt 1-224) was tested using a dual-luciferase reporter construct in which this element was inserted between the firefly and *Renilla* luciferase genes (pDualLuc-IRES1; Fig. 5-7A). To exclude the possibility that coding sequences downstream of the EAV 5' UTR are an essential part of a possible IRES element, the 5’ end of ORF1a (nt 225-287) was added to the 5' UTR and fused in-frame to the *Renilla* gene, thus creating an N-terminal extension of 21 amino acids (pDualLuc-IRES2). The ratio of R_{luc} over F_{luc} expression was taken as a measure for IRES activity. Since in both cases no detectable IRES activity was found, compared to the well-established activity of the EMCV IRES element (pDualLuc-IRES3 and pDualLuc-IRES4 respectively), the results obtained with the above constructs strongly argued against the presence of an IRES element in the 5’-proximal region of the EAV genome (Fig. 5-7A).

Using the EMCV IRES-mediated R_{luc} expression of construct pDualLuc-IRES3 as an internal standard, we subsequently analyzed whether the 5’ UTR of the EAV genome is scanned by ribosomes, presumably after cap-dependent translation initiation. When the EAV 5' UTR preceded the F_{luc} gene (pDualLuc-scan1), strong F_{luc} expression (F_{luc}/R_{luc})

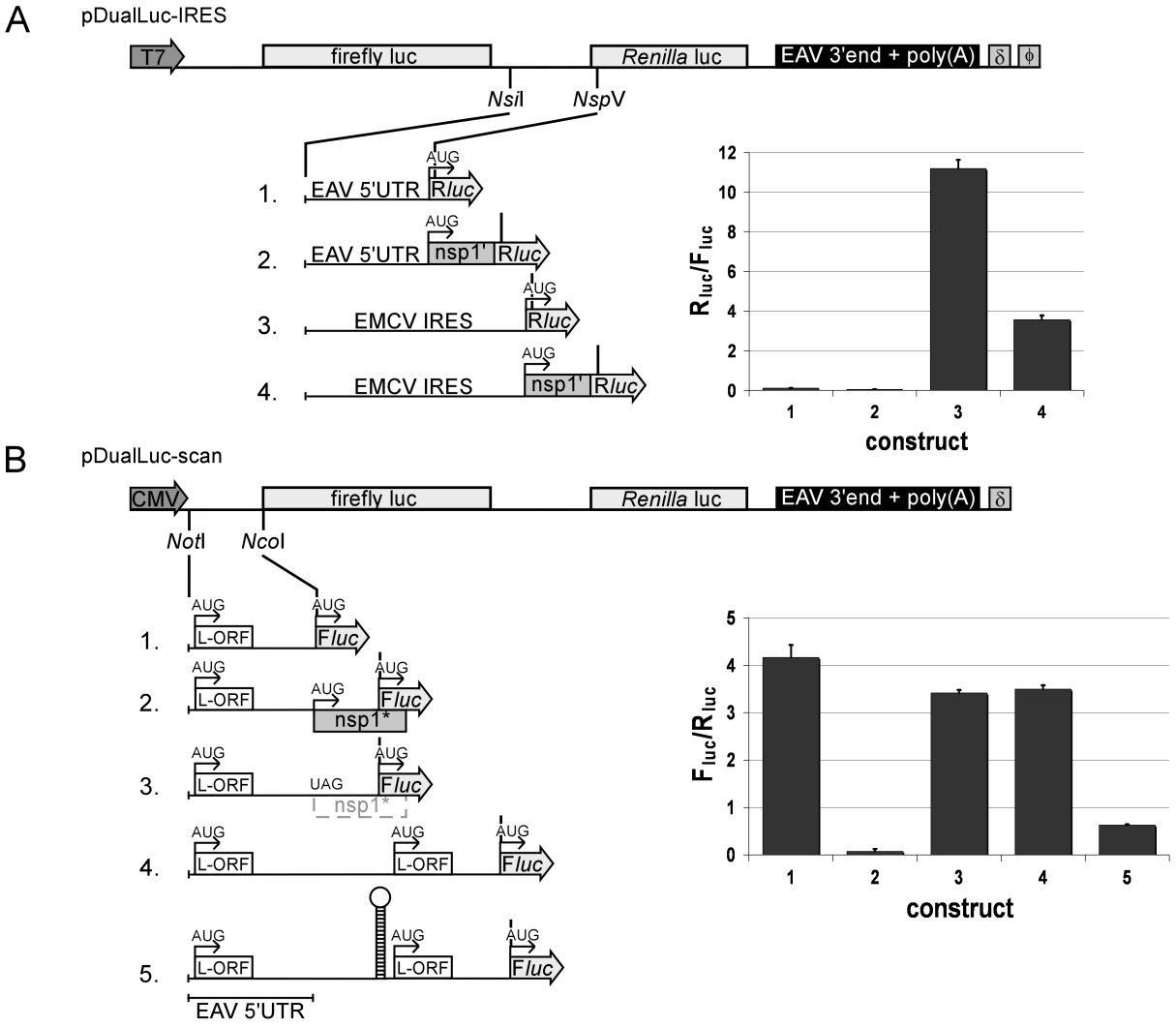


Figure 5-7. Analysis of translation initiation at the 5' end of the EAV genome by using bicistronic reporter constructs and dual luciferase translation assays. *(A) To assess the possibility of the presence of an IRES element, EAV 5'-proximal sequences were cloned in pDualLuc-IRES between the F_{luc} and R_{luc} reporter genes. The EMCV IRES element was used as positive control. Nsp1' indicates the N-terminal in-frame extension of the Renilla luciferase with the N-terminal 21 amino acids of nsp1. The T7 promoter (arrowhead) and terminator (Φ), the 3'-terminal region of the EAV genome starting from nucleotide 11,737 and including a poly(A) tail, and the hepatitis δ ribozyme (δ) are indicated. The numbers before each construct in the left panel correspond to the bar in the graph showing the relative expression of R_{luc} versus F_{luc}. (B) Analysis of ribosome scanning after cap-dependent translation initiation. Sequence elements were cloned into pDualLuc-scan upstream of the F_{luc} reporter gene. The intraleader (L-ORF; see text) and the position of the EAV 5' UTR are indicated. Nsp1* corresponds to an upstream ORF (out-of-frame with F_{luc}) that encodes the N-terminal 53 amino acids of nsp1. The CMV promoter (arrowhead) is indicated. The numbers before each construct in the left panel correspond to the bar in the graph showing the relative expression of F_{luc} versus R_{luc}.*

was observed (Fig. 5-7B). Subsequently, a small (53-codon) ORF was introduced upstream of the F_{luc} gene by extending the EAV 5' UTR with the 5' end of ORF1a (nt 227 to 303). Scanning ribosomes would be expected to translate this upstream ORF and as a

result less ribosomes would be available for the translation of the downstream F_{luc} gene (Kozak, 1991). Indeed, F_{luc} expression by this mutant was decreased dramatically (pDualLuc-scan2; Fig. 5-7B). In line with the scanning hypothesis, F_{luc} expression was restored to normal levels when the upstream AUG was converted to UAG (pDualLuc-scan3; Fig. 5-7B). The influence of the short L-ORF duplicate in the 529-nt long 5' UTR of mutant D5D (Fig. 5-2A) was evaluated by inserting this 5' UTR upstream of the F_{luc} gene (pDualLuc-scan4). Remarkably, F_{luc} expression was hardly affected by the increased size of this 5' UTR, nor was it affected by the presence of two copies of the L-ORF (Fig. 5-7B). This result indicated that, at least *in vitro*, the large 5' UTR of D5D-derived constructs did not negatively influence translation of the replicase. Finally, we evaluated the effect of the insertion of a stable hairpin structure upstream of the F_{luc} ORF, a change that should reduce translation efficiency because such structures have to be unfolded before scanning ribosomes can proceed (Pelletier & Sonenberg, 1985; Kozak, 1989). At nt 300 of the 529-nt D5D 5' UTR, we inserted a stable hairpin (ΔG = -60 kcal/mol) that was previously shown to inhibit CAT expression by 92 to 98%, depending on the position of this hairpin in the insulin-like growth factor II leader 1 (van der Velden *et al.*, 2002). In our hands, the presence of this hairpin in pDualLuc-scan5 reduced F_{luc} expression with 80%, suggesting that indeed this hairpin interfered with ribosome scanning (Fig. 5-7B). In conclusion, the experiments performed with the pDualLuc vectors strongly suggested that EAV genome translation is not IRES mediated and that translational initiation conforms to the conventional 'ribosome scanning' model (Kozak, 1989; Kozak, 2002).

Discussion

In addition to providing experimental support for an important role of the previously defined leader TRS-presenting hairpin in sg RNA synthesis, the constructs used in this study have revealed a remarkable functional flexibility of the 5'-proximal region of the EAV genome. Genome replication was not significantly affected in D5Dwt and its derivatives with duplicated 5'-proximal sequences. Even more striking, and logically the basis of efficient genome replication, was the fact that increasing the size of the genomic 5' UTR by more than twofold did apparently not cripple genome translation, despite the abundant presence of RNA structures (Fig. 5-1C) and in some cases an additional L-ORF. However, this short L-ORF did not affect expression of the F_{luc} reporter (Fig. 5-7B). The simplest explanation is that conventional scanning ribosomes do not recognize the L-ORF AUG codon, because it is located in a weak sequence context for translation initiation. The collective data from our translation experiments argued against IRES-mediated translation initiation (Fig. 5-7A). Besides, the region flanking the novel replicase translation initiation codon in the D5D was altered extensively during our mutagenesis studies and would likely have affected RNA structures with IRES activity. The translational flexibility of the system permitted the expansion of the 5' UTR of the genome and all sg mRNAs in D5Dwt and its derivatives. Some of these (e.g. D5DcmΔ1-163; Fig. 5-3 and 5-5) were even stable for several passages and replicated to titers 10- to 100-fold lower than those reached by the wt virus. The successful uncoupling of functions normally relying on overlapping RNA signals and the type of mutants developed in this study will be useful for further dissection of this multifunctional domain of the EAV genome.

In previous studies, we have analyzed the functions of the EAV leader TRS as well as body TRSs in sg RNA synthesis. Whereas multiple functions were attributed to the body TRSs, it was proposed that the primary role of the leader TRS is to serve as a "base-pairing target" in the recombination-like fusion process that results in the joining of leader and body sequences (Fig. 5-1B) (Pasternak *et al.*, 2001). Upon mutagenesis of the leader TRS, the presence of sufficient base-pairing potential appeared to be the only prerequisite for sg RNA synthesis and a direct correlation between the stability of the leader TRS-body TRS duplex and transcription levels was observed, both for arteriviruses and coronaviruses (Pasternak *et al.*, 2003; Zuniga *et al.*, 2004). Although TRS-like sequences are present at other positions in the EAV genome, we have never obtained evidence for productive base-pairing of a body TRS to an alternative acceptor TRS. Clearly, its 5'-proximal position makes the leader TRS unique, possibly also because 5'-3' interactions in the genome may bring body and leader TRSs in each other's vicinity (Lai, 1998). However, our previous studies of the RNA secondary structure of this region of the genome suggested additional specific features that might influence leader TRS function and preliminary evidence supporting the importance of the structural context provided by the LTH was obtained (van den Born *et al.*, 2004). The present study confirms and

extends these observations and demonstrates that the LTH is an independent structural entity that can be moved to direct leader-to-body fusion at another position in the 5'-proximal region of the EAV genome. Mutagenesis of the LTH in the context of the D5D constructs strongly supported the importance of the LTH as a transcription-guiding element (Fig. 5-4). We also delimitated the LTH-flanking sequences required for efficient EAV transcription by deletion mutagenesis of the D5D sequence (Fig. 5-2). The activity of the LTH structure was clearly influenced by the presence of both upstream and downstream flanking sequences (Fig. 5-2). These sequences were predicted to fold into a base-paired region that may extend the LTH stem (Fig. 5-3). Mutations in this region clearly affected transcription, although the attempts to obtain direct support for base-pairing interactions of these sequences were not very convincing (Fig. 5-3). In this regard, the best results were obtained with domain A (Fig. 5-3), the top part of the LTH stem extension, for which a certain level of restoration of sg RNA synthesis was observed in the LR double mutant. An additional role for the RNA primary structure or the formation of misfolded alternatives of the eLTH may explain the fact that the results with domain B and C mutants were less convincing (Fig. 5-3). Alternatively, the stability of the LTH stem closest to the leader TRS may be very critical for its function (Fig. 5-4), and the contribution of lower stem regions may be small (Fig. 5-3). Subtler single-nucleotide changes may be used in future studies to probe the function of the LTH stem regions in more detail.

Other support for the importance of an eLTH fold comes from the theoretical analysis of RNA structures in some of our mutants and in other arteriviruses. For example, compared to mutant D5DΔ1-184 (Fig. 5-2), sg mRNA synthesis by construct D5DcΔ1-184 (Fig. 5-3) was remarkably increased. In the latter construct, coding sequences downstream of the LTH were mutated. The entirely artificial combination of this set of translationally silent mutations downstream of the LTH and the Δ1-184 deletion upstream was found to give rise to an alternative potential LTH extension, mimicking the wt structure and with similar stability (Fig. 5-8A). During our analysis of other arterivirus genomes, the LTH was predicted to contain either a long or a short stem region (Fig. 5-8B) (van den Born *et al.*, 2004). In the structure prediction for the short version of the LTH, alternative hairpins were predicted at the expense of the formation of the lower part of the eLTH stem. The folding of the eLTH, in turn, is supported by a covariation observed at the bottom of the PRRSV structure (Fig. 5-8B). The two alternative predicted LTH conformations did tolerate the sequence variations found in different strains for EAV, LDV, and PRRSV (only one SHFV sequence was available). Additionally, the phylogenetic and probing data obtained during our analysis of the 5'-proximal region of the EAV genome did not exclude alternative conformations, although the data favored the short LTH to a certain extent (van den Born *et al.*, 2004).

The above supports our hypothesis that the eLTH structure may be important for sg RNA synthesis and also suggests that the type and stability of the stem may be a crucial factor. Nevertheless, biochemical evidence for the presence of these structures remains to

be obtained. In addition, protein factors may recognize the different LTH conformations and thereby e.g. influence sg RNA synthesis. Whether the alternative structures predicted for arteriviruses represent a conformational switch in this multifunctional region of the genome remains to be investigated. Examples of such conformational switches that are based on alternative RNA secondary structures, either self-induced (Nagel & Pleij, 2002) or modulated by other molecules (Mandal & Breaker, 2004), are widespread in RNA virology. Conformational transitions in viral RNAs were suggested to regulate RNA function at different stages of the viral replication cycle (Olsthoorn *et al.*, 1999; Diez *et al.*, 2000; van Meerten *et al.*, 2001; Dirac *et al.*, 2002). In the case of arteriviruses, a conformational switch in the genomic leader region could regulate e.g. functions involved in translation, replication and/or sg mRNA synthesis.

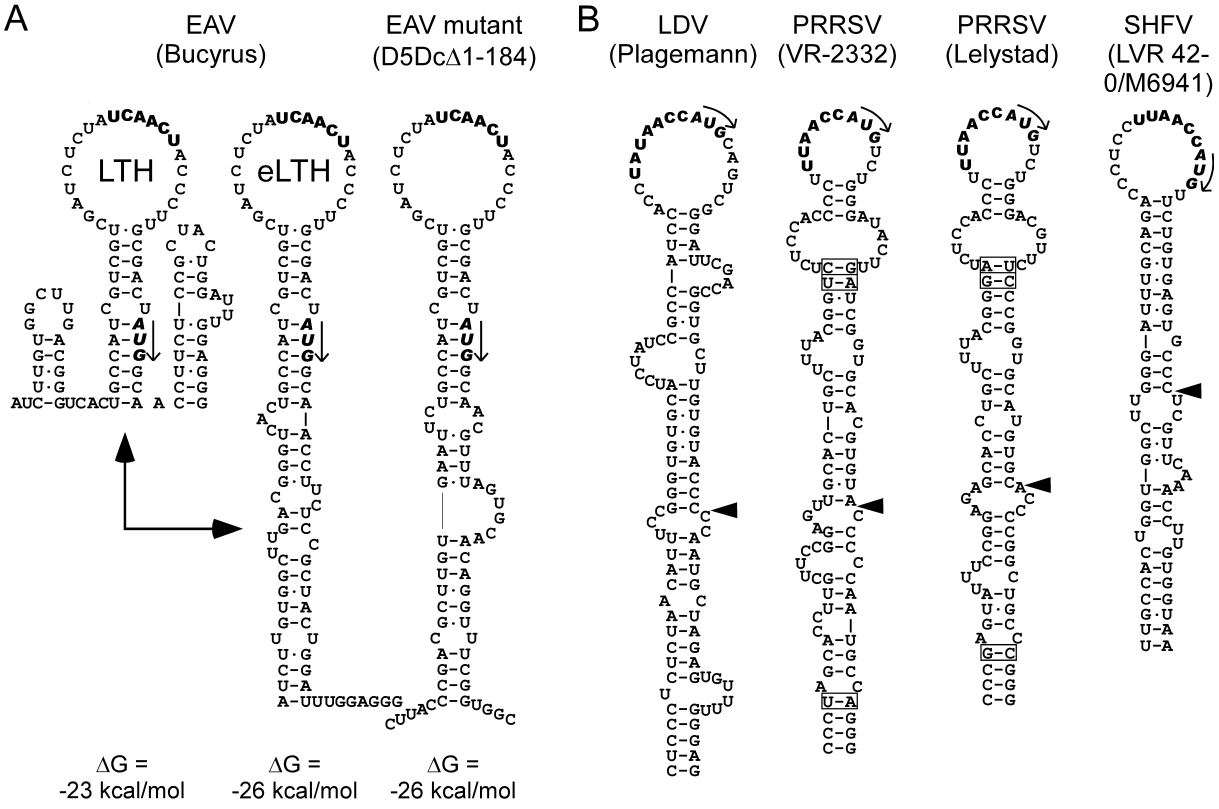


Figure 5-8. *(A) RNA secondary structure prediction of the LTH and flanking sequences of EAV strain Bucyrus and construct D5DcΔ1-184. The arrow signifies a possible switch between the two possible conformations of the EAV LTH region. Structure stability is indicated with a free energy value. (B) RNA secondary structures prediction of the LTH region of different arteriviruses, with possibilities for the formation of an extended LTH stem in each of them. Depicted are LDV strain Plagemann, PRRSV strains VR-2332 and Lelystad, and SHFV strain LVR 42-0/M6941. A box indicates base-pair covariations between two PRRSV strains. The position where the predicted stem extension of the LTH starts, is indicated with an arrowhead. The replicase translation initiation codon is shown in bold italic font and the leader TRS is indicated in bold font.*

Materials and Methods

RNA transfection, infection, and protein analysis

Baby hamster kidney cells (BHK-21; ATCC CCL10) were maintained (van Berlo *et al.*, 1982) and used for propagation of EAV and RNA transfection experiments as described previously (van Dinten *et al.*, 1997). Virus titration using plaque assays was done as described previously (Molenkamp *et al.*, 2000a). Immunofluorescence assays with EAV-specific antisera for nsp3 (98.E3) and nucleocapsid protein N (MAb 3E2) were done on EAV RNA-transfected and EAV-infected cells (van der Meer *et al.*, 1998). Prior to Western blot analysis of nsp1 expression using the anti-nsp1 monoclonal antibody 12A4 (Wagner *et al.*, 2003), cells were lysed at 15 h post transfection (p.t.) according to the method described previously (de Vries *et al.*, 1992). Proteins were separated by 15% SDS-PAGE and transferred to a Hybond-P PVDF transfer membrane (Amersham Biosciences), essentially as described before (Snijder *et al.*, 1994). The nsp1 band was visualized using the ECL plus Western Blotting Detection System (Amersham Biosciences) according to the manufacturer's instructions.

Mutagenesis of an EAV full-length cDNA clone

Mutations were introduced into shuttle plasmids by standard site-directed PCR mutagenesis. After sequence analysis, restriction fragments containing the desired mutations were transferred to full-length clone pEAV211, a derivative of pEAV030 (van Dinten *et al.*, 1997), containing a few translationally silent, engineered restriction sites. Virus derived from clone pEAV211 has a wt phenotype, as determined by comparison with pEAV030-derived virus in growth curve experiments and plaque assays (data not shown).

Construct pD5Dwt (Fig. 5-2A) contained a tandem repeat of the 5'-proximal 301 nucleotides (nt) of the EAV genome, which was generated by inserting a PCR cassette at nt 306 using an engineered *Eco*RI site and a *Mlu*I site (nt 589) of a pEAV211-derived clone, in which the original ORF1a translation initiation codon had been changed to UAG (Tijms *et al.*, 2001). Because this latter mutation introduced a two base-pair mismatch in the LTH stem, the base-pairing potential was restored by making two compensating mutations in the opposite side of the stem (van den Born *et al.*, 2004). Replicase translation in D5Dwt was initiated at the start codon present in the duplicated sequence. In addition, the original leader TRS in the genomic 5'-proximal sequence was inactivated (5'-UCAACU-3' → 5'-AGUUGU-3') (van Marle *et al.*, 1999a). In some pD5Dwt variants, the replicase codons present within the duplicate (encoding the N-terminal 29 amino acids)

were changed as much as possible without changing the encoded protein sequence (constructs containing this set of mutations have a lower-case "c" in their name, e.g. pD5Dc; mutated nucleotides are underlined: 5'-AUG GCU ACG UUU AGU GCA ACA GGU UUC GGU GGC UCA UUC GUA CGC GAU UGG AGU UUA GAU CUU CCG GAU GCA UGC GAA CAC GGG GCC-3').

Several deletions were made in the duplicate of the EAV 5' sequence in constructs pD5Dwt and pD5Dc. The region corresponding to nt 1 to 135 of the genome was deleted, yielding pD5DΔ1-135 and pD5DcΔ1-135, respectively. The region directly upstream of the LTH (corresponding to nt 136 to 183 of the genome) was removed from pD5Dwt to create pD5DΔ136-183. In addition, nt 1 to 184 were deleted from the duplicated sequence, generating pD5DΔ1-184 and pD5DcΔ1-184, respectively. From the latter construct, the LTH stem was removed, generating pD5DcΔ1-184Δstem. Three constructs were generated in an adapted version of pD5Dc, pD5Dcm, in which the replicase ORF initiated 27 nt downstream of the normal AUG codon present in the duplicated sequence. The latter start codon was changed into UAG and the base-pairing in the LTH stem was restored as described above. The region encompassing nt 1-163 or 1-184 was deleted from the duplicate, yielding pD5DcmΔ1-163 or pD5DcmΔ1-184, respectively. Furthermore, a derivative of pD5DcmΔ1-163 was generated in which the original wt leader TRS was not inactivated by mutagenesis and which therefore contained two wt leader TRSs (pD5DcmΔ1-163\2xwt).

To investigate the function of the LTH by mutagenesis, construct pD5DΔ1-163 was used with an intact (pD5DΔ1-163\wt) or mutant LTH in the duplicate. To disrupt the upper part of the LTH stem, the 5' side of the stem was mutated (5'-GUCGU-3' → 5'-CUGCG-3') creating pD5DΔ1-163\US-L, or the 3' side of the stem was mutated (5'-GCGAC-3' → 5'-UGCAG-3') to create pD5DΔ1-163\US-R. Combining these mutations resulted in pD5DΔ1-163\US-LR. The LTH loop was partly closed by introducing mutations in the 5' side of the loop (5'-CGAU-3' → 5'-AGGG-3'), creating pD5DΔ1-163\L-L, or the 3' side of the loop (5'-CCUU-3' → 5'-AUUG-3'), creating pD5DΔ1-163\L-R. Combining these mutations resulted in pD5DΔ1-163\L-LR.

To investigate the presence and role of the predicted extended version of the LTH stem, construct pD5DcmΔ1-163 was used as a basis. The entire 5' side of the stem region was mutated (5'-AUCUUGUGGCUUGACGGGU-3' → 5'-UAGGUCAUCGUUCUCUCCA-3'), creating pD5DcmΔ1-163\CL, or the entire 3' side of the stem was mutated (5'-ACCUUCUCCGCUACUGGAU-3' → 5'-UGGGUCAGCCGGUGUUCUA-3'), creating pD5DcmΔ1-163\CR. Combining these mutations resulted in pD5DcmΔ1-163\CLR. Also a smaller base-paired region directly below the LTH was probed by mutating the 5' side of the stem (5'-AUCUUGUGGCUUGACGGGU-3' → 5'-AUCUUGUGGCUUGACUCCA-3'), creating pD5DcmΔ1-163\AL, or by mutating the 3' side of the stem (5'-ACCUUCUCCGCUACUGGAU-3' → 5'-UGGGUCUCCGCUACUGGAU-3') to create pD5DcmΔ1-163\AR. Combining these mutations resulted in pD5DcmΔ1-163\ALR.

Finally, a second smaller base-paired region was changed by mutating the 5' side of the stem (5'-AUCUUGUGGCUUGACGGGU-3' → 5'-AUCUUCAUCGUUGACGGGU-3'), created pD5DcmΔ1-163\BL, or the 3' side of the stem (5'-ACCUUCUCCGCUACUGGAU-3' → 5'-ACCUUCUCCCGGUGUGGAU-3'), creating pD5DcmΔ1-163\BR. Combining these mutations resulted in pD5DcmΔ1-163\BLR.

RNA isolation and analysis.

Intracellular RNA from BHK-21 cells was isolated at 15 h p.t., separated in denaturing agarose-formaldehyde gels and detected by hybridization to a ^{32}P-labeled probe (5'-TTGGTTCCTGGGTGGCTAATAACTACTT-3'), which is complementary to the 3' end of the EAV genome and recognizes both the genome and all sg mRNAs (van Marle *et al.*, 1999a). Dried gels were exposed to phosphorimager screens, which were scanned with a Personal Molecular Imager FX (Bio-Rad) after exposure. Band intensities were quantitated with Quantity One v4.2.2 (Bio-Rad). The same RNA was also used as template for primer extension analysis, essentially as described previously (van den Born *et al.*, 2004), and for reverse transcription reactions and subsequent PCRs. PCR products were sequenced using standard protocols.

RNA secondary structure prediction

RNA secondary structures were predicted using the genetic algorithm of STAR v4.4 (Gultyaev *et al.*, 1995) and the energy minimization program of Mfold Web server (Zuker, 2003).

Translation assays

To test for internal ribosomal entry site (IRES) activity of specific EAV sequences, cDNA cassettes were placed in between the *Photinus pyralis* (Firefly) and *Renilla reniformis* (*Renilla*) luciferase genes of a pDualLuc reporter plasmid (Reusken *et al.*, 2003). Bicistronic mRNA transcripts were initiated from a T7 RNA polymerase promoter and contained the firefly (F_{luc}) and *Renilla* (R_{luc}) luciferase genes, the 3'-terminal region of the EAV genome starting from nucleotide 11,737 and including a poly (A) tail, and the hepatitis δ ribozyme (Fig. 5-7A). In construct pDualLuc-IRES1, the EAV 5' UTR (nt 1-224) was tested with the replicase translation initiation codon being the starting point for R_{luc} expression. The 3' end of the EAV 5' UTR was extended with sequences encoding the N-terminal part of nsp1 (nt 225-287) in construct pDualLuc-IRES2. For both plasmids

a positive control was generated, in which the EAV 5' UTR was replaced with the *Encephalomyocarditis virus* (EMCV) IRES element (Jang *et al.*, 1989; Jackson & Kaminski, 1995), yielding pDualLuc-IRES3 and pDualLuc-IRES4, respectively. Rabbit kidney (RK-13) cells (de Vries *et al.*, 1992) were infected with *Vaccinia virus* recombinant vTF7-3 (Fuerst *et al.*, 1986), which expresses the T7 RNA polymerase, at an m.o.i. of 10 and transfected with 1 µg of plasmid using lipofectamine PLUS (Invitrogen) at 2 h p.i. Cells were lysed at 8 h p.i. and assayed for F_{luc} and R_{luc} activities. Enzymatic activities were measured using a luminometer (Turner Designs TD-20/20) and performed with the Dual-luciferase Reporter assay according to the manufacturer's protocol (Promega).

To assess whether cap-dependent translation initiation of the EAV genome is followed by classical linear ribosome scanning, different cassettes were cloned upstream of the F_{luc} gene in pDualLuc-IRES3. Because the capping efficiency of the vTF7-3-driven expression system is relatively poor, the T7 promoter was replaced with a *Cytomegalovirus* (CMV) immediate early promoter (Fig. 5-7B). The EAV 5' UTR (nt 1-224) was placed upstream of the F_{luc} gene, yielding pDualLuc-scan1. In construct pDualLuc-scan2, the EAV 5' UTR was extended until nt 303 (including the 5’ end of the replicase gene) containing an extra AUG codon at its 3' end to drive F_{luc} expression. Subsequently, the replicase AUG was mutated into UAG (pDualLuc-scan3). Finally, the 5' UTR of D1-301 (see above) was inserted (pDualLuc-scan4) and a stable hairpin was introduced into this region at nt 306 to block scanning ribosomes, generating pDualLuc-scan5 (van der Velden *et al.*, 2002). RK-13 cells were transfected with 1 µg of plasmid and were lysed at 20 h p.t. and assayed for F_{luc} and R_{luc} activities.

Acknowledgements

We thank Udeni Balasuriya and Jim Maclachlan (University of California, Davis, USA) for providing monoclonal antibodies. Chantal Reusken kindly provided some of the pDualLuc reporter plasmids. We appreciated helpful discussions with Willy Spaan and Gijs Versteeg. E. van den Born was supported by Council for Chemical Sciences grant CW 99-010 from the Netherlands Organisation for Scientific Research.

Chapter 6

Antiviral Activity of Morpholino Oligomers Designed to Block Different Aspects of Equine Arteritis Virus Reproduction in Cell Culture

Erwin van den Born, David A. Stein, Patrick L. Iversen, and Eric J. Snijder (2005)
Journal of General Virology **86**: 3081-3090

Reprinted with permission

Abstract

The antiviral efficacy of ten antisense phosphorodiamidate morpholino oligomers (PMOs) directed against *Equine arteritis virus* (EAV), a nidovirus belonging to the Arteriviridae family, was evaluated in mammalian (Vero-E6) cells. Peptide-conjugated PMOs (P-PMOs) supplied in cell-culture medium at micromolar concentrations were efficiently taken up by Vero-E6 cells and were minimally cytotoxic. The P-PMOs were designed to base pair to RNA sequences involved in different aspects of EAV amplification: genome replication, subgenomic mRNA synthesis, and translation of genome and subgenomic mRNAs. A novel recombinant EAV, expressing green fluorescent protein as part of its replicase polyproteins, was used to facilitate drug screening. A moderate reduction of EAV amplification was observed with relatively high concentrations of P-PMOs designed to anneal to the 3'-terminal regions of the viral genome or antigenome. To determine if the synthesis of subgenomic mRNAs could be specifically reduced, transcription-regulating sequences essential for their production, but not for the production of genomic RNA, were targeted, but these P-PMOs were found to be ineffective at transcription interference. In contrast, all four P-PMOs designed to base pair with targets in the genomic 5' untranslated region markedly reduced virus amplification in a sequence-specific and dose-responsive manner. At concentrations in the low micromolar range, some of the P-PMOs tested completely inhibited virus amplification. In vitro translation assays showed that these P-PMOs were potent inhibitors of translation. Our data suggest that these compounds could be useful as reagents for exploring the molecular mechanics of nidovirus translation, and have anti-EAV potential at relatively low concentrations.

Introduction

In recent years, phosphorodiamidate morpholino oligomers (PMOs) have been used as a loss-of-function tool to inhibit gene expression and have provided a useful method to study gene function, especially in the area of developmental biology (Heasman, 2002; Liu *et al.*, 2003; Boguslavsky *et al.*, 2003; Muto *et al.*, 2004). PMOs are single-stranded DNA-mimics, typically synthesized to 20-25 bases in length, and display high biological stability and low toxicity. Most of the PMOs reported to effectively reduce gene expression targeted the 5' untranslated region (5' UTR) of mRNA or the first 25 nt 3' of the translation initiation codon (Summerton, 1999; Heasman, 2002). In instances of translation inhibition, it is presumed that the oligomers form a complex with their mRNA target that physically blocks critical events in pre-initiation, ribosomal scanning, or initiation at the start site region. The documented success at translation inhibition has raised the prospect of evaluating the potential of PMOs to block the amplification of viruses, in particular positive-stranded RNA viruses whose replication cycle depends on translation of incoming genomic RNA. For example, PMOs targeted to the *Hepatitis C virus* (HCV) internal ribosome entry site (IRES) were shown to reduce translation of HCV RNA both *in vitro* and *in vivo* (Jubin et al., 2000; McCaffrey et al., 2003). Highly efficient uptake of these uncharged oligomers by cultured cells can be achieved by conjugating PMO to any of a number of arginine-rich peptides (Moulton *et al.*, 2003; Moulton *et al.*, 2004). Peptide-conjugated morpholino oligomers (P-PMOs) have been demonstrated to have sequence-specific inhibitory effects on the amplification of a coronavirus (Neuman *et al.*, 2004) and several flaviviruses (Kinney et al., 2005; Deas et al., 2005) in cell culture.

In this report, we describe the evaluation of the efficacy of P-PMOs designed to inhibit the amplification of *Equine arteritis virus* (EAV), the prototype of the *Arteriviridae*, which is united with the *Coronaviridae* and *Roniviridae* in the order *Nidovirales* (Snijder & Meulenberg, 2001; González *et al.*, 2003). In addition to the full-length genome RNA (Fig. 6-1A), EAV and other nidoviruses produce a set of nested subgenomic (sg) mRNAs in infected cells. In the case of arteriviruses and coronaviruses, these sg mRNAs contain a common leader sequence, derived from the 5'-terminal region of the genome, that is fused to different but overlapping "body sequences", which are derived from the 3'-proximal region of the genome. Thus, sg mRNAs are both 5'- and 3'-coterminal with the genomic RNA. Conserved transcription-regulating sequences (TRSs) precede every open reading frame in the 3'-proximal quarter of the genome (body TRSs) and an additional copy is found at the 3' end of the leader (leader TRS). Base pairing between the leader TRS (in the plus strand) and body TRS complements in the minus strand was found to play a critical role in joining the leader and body segments of sg RNAs in arteriviruses and coronaviruses (Pasternak *et al.*, 2001; Pasternak *et al.*, 2003; Zuniga *et al.*, 2004). The fusion of the sg RNA body to the leader sequence has been postulated to involve discontinuous extension during minus-strand RNA synthesis (Fig. 6-

1B) (Sawicki & Sawicki, 1995). After attenuation of RNA synthesis at the body TRS, the nascent minus strand is thought to translocate to the 5'-proximal region of the genomic template. Following TRS-guided base pairing with the template, RNA synthesis is

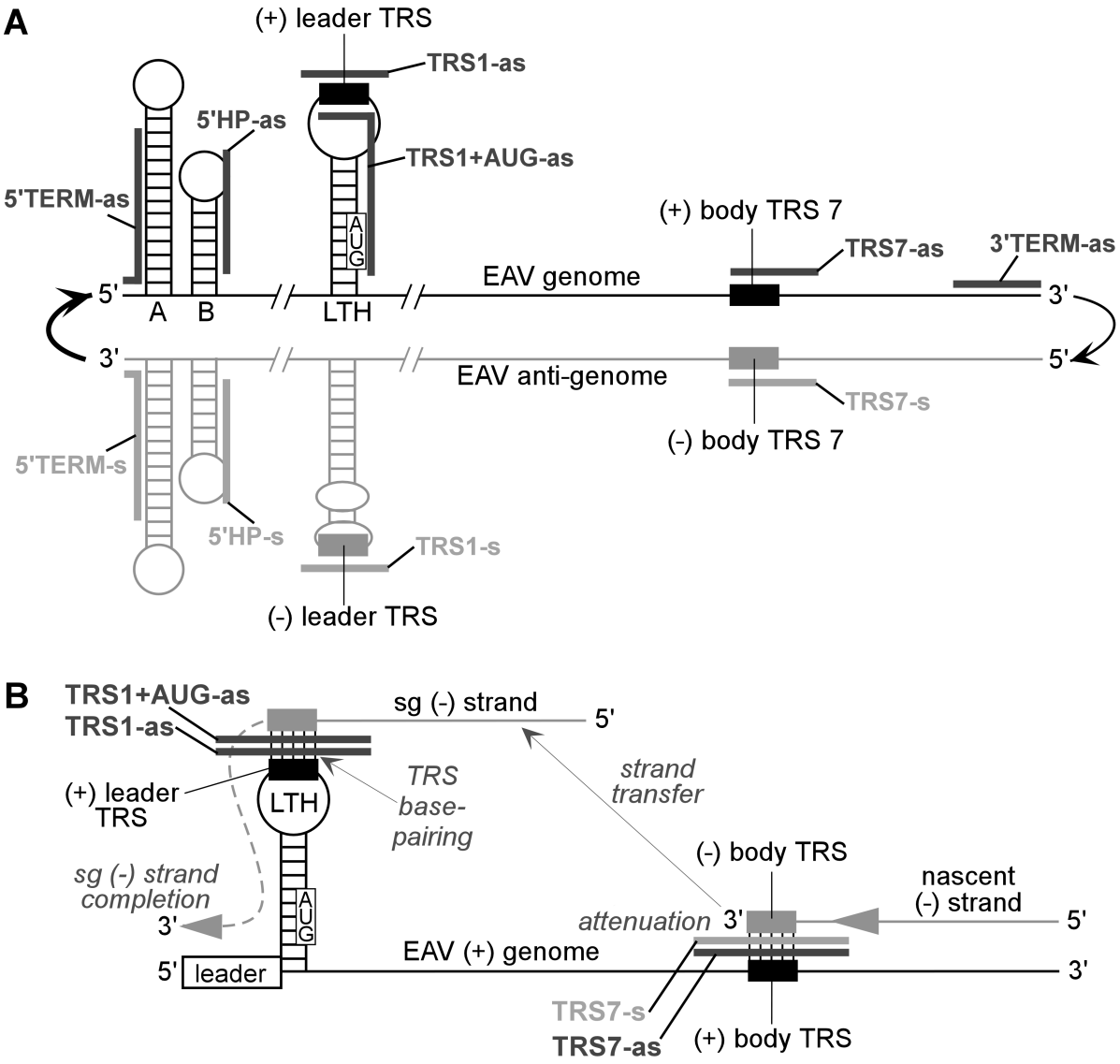


Figure 6-1. Overview of EAV P-PMO target locations. *(A) Schematic overview of the position of the P-PMO targets in the EAV genome and antigenome (antisense and sense P-PMOs are depicted in red and orange, respectively). Elements from the previously described structure model for the 5'-proximal region of the genome are depicted (van den Born et al., 2004). The RNA structures drawn in the minus strand (grey) are hypothetical and have not been demonstrated via structure probing. Depicted are the 5'-terminal hairpin A, the conserved hairpin B, and the leader TRS hairpin (LTH). The black boxes indicate the position of the leader TRS and RNA7 body TRS in the genome; the grey boxes indicate the corresponding complementary sequences in the antigenome. (B) P-PMOs designed to interfere with discontinuous minus strand synthesis (see text and (Sawicki & Sawicki, 1995)). P-PMOs TRS1-as, TRS7-s and TRS1+AUG-as were designed to interfere with the leader TRS-to-body TRS base-pairing step.*

resumed, adding the complement of the genomic leader sequence, thus completing the sg minus strand. Positive-strand sg mRNAs are then transcribed from corresponding minus-strand sg RNA templates (Sawicki & Sawicki, 1995).

The arterivirus genome, and presumably all sg mRNAs, are capped and polyadenylated. It is most likely, translation initiation of the EAV mRNAs is cap-dependent and the ribosomal pre-initiation complex scans the 5' UTR in a linear manner to identify the translation initiation codon of the replicase gene (van den Born *et al.*, 2005). The viral RNA elements involved in EAV genome replication are poorly defined, but data from our previous studies suggest that at least the 5'-proximal 290 and the 3'-proximal 354 nt are required for efficient replication (Molenkamp *et al.*, 2000b; Molenkamp *et al.*, 2000c; Tijms *et al.*, 2001; van den Born *et al.*, 2005).

Ten P-PMOs were designed to specifically bind to RNA sequences involved in different aspects of EAV amplification: genome replication, sg mRNA synthesis, and translation of genome and sg mRNAs. The 3'-terminal regions of the genome and antigenome were chosen as P-PMO target sites in an attempt to block replication, but only a moderate reduction of virus amplification was observed at relatively high oligomer concentrations. To assess whether the synthesis of exclusively sg mRNAs could be blocked, TRS motifs were targeted, but these P-PMOs were found to be ineffective at interfering with transcription. In contrast, all four of the antisense P-PMOs designed to base pair with sequences in the positive-strand genomic 5' UTR efficiently reduced virus amplification in a dose-responsive manner. At concentrations in the low micromolar range, some of these compounds completely, and others near-completely, inhibited virus amplification. *In vitro* translation assays indicated that these P-PMOs likely inhibited genome translation. Our data suggest that several of these compounds have antiviral potential at relatively low concentrations and also represent useful tools to study the molecular biology of arteriviruses.

Results

Design of P-PMOs targeting EAV RNA sequences involved in replication, sg mRNA synthesis, and translation

It has been established that the efficient replication of many RNA viruses requires essential sequence and structural elements in terminal regions of genome and antigenome, which are recognized by the viral RNA-dependent RNA polymerase (RdRp) complex (Panavas *et al.*, 2002; Tortorici *et al.*, 2003). Specifically, the RdRp complex interacts with the 3' end of genome and antigenome to initiate the synthesis of minus and plus strand, respectively (Fig. 6-1A). In an attempt to interfere with such interactions, P-PMOs complementary to the 3' end of genome (3'TERM-as) and antigenome (5'TERM-s) were designed. A previous analysis of the secondary structure of the 5'-proximal 300 nt of the EAV genome revealed that the 5' UTR contains a RNA hairpin that is conserved among different EAV isolates and is predicted to occur in both the plus and the minus strand conformations (hairpin B; Fig. 6-1A) (van den Born *et al.*, 2004). Because of its sequence and structural conservation and its close proximity to the 3' end of the antigenome, it likely plays a role in the initiation of plus strand synthesis. Two P-PMOs were designed to disrupt this RNA structure either in the plus or in the minus strand (5'HP-as and 5'HP-s, respectively).

To assess whether using P-PMOs could reduce the synthesis of sg mRNAs, several TRS motifs were targeted (Fig. 6-1B). PMO-TRS interactions were expected to interfere with body TRS-to-leader TRS base pairing, which is a crucial step in the production of sg RNAs (van Marle *et al.*, 1999; Pasternak *et al.*, 2001). The leader TRS of EAV is predicted to be located in the 20-nt loop of the so-called leader TRS hairpin (LTH; Fig. 6-1). One P-PMO was complementary to the entire LTH loop (TRS1-as), whereas another P-PMO spanned both the leader TRS and the downstream replicase translation initiation codon (TRS1+AUG-as). A P-PMO complementary to TRS1-as (TRS1-s) was an appropriate control in these transcription-inhibition studies as it was designed to target the complement of the leader TRS, which is not involved in TRS base pairing. In addition, the RNA7 body TRS was targeted with two P-PMOs designed to interfere specifically with sg RNA7 synthesis. One P-PMO (TRS7-as) was complementary to the RNA7 body TRS in the genomic template and was designed to interfere with the attenuation step that has been postulated to precede the discontinuous step during sg minus-strand RNA synthesis. The other (TRS7-s) was designed to anneal to the RNA7 body TRS complement at the 3' end of the nascent minus strand that, following attenuation, is presumed to be translocated to and base pair with the leader TRS.

Four of the P-PMOs were designed against targets located in the genomic 5’ UTR. In addition to the 5’HP-as, TRS1-as and TRS1+AUG-as compounds, described above, a

P-PMO with a sequence antisense to the 5'-terminal 20 nucleotides (5'TERM-as) was designed to interfere with pre-initiation of translation of all sg mRNAs and genomic mRNA. Although P-PMO TRS1+AUG-as (Fig. 6-1A) was designed to interfere with genome translation (as well as sg mRNA synthesis) it is the only one of the four with little likelihood of interfering with sg mRNA translation, as all sg mRNAs lack most of the 16 contiguous nucleotides immediately downstream of the leader TRS that are part of the TRS1+AUG-as target sequence (den Boon *et al.*, 1996). In contrast, translation of all sg mRNAs and genome RNA could be affected by 5'TERM-as and 5'HP-as as they target sequences in the common leader. A 'nonsense' P-PMO (SCR) was also synthesized as a control for off-target effects of the P-PMO chemistry. P-PMO design details are summarized in Table 6-1.

Table 6-1. P-PMO details

P-PMO	target in EAV	P-PMO polarity	P-PMO sequence (5'→ 3')	position*	$IC_{50(GFP)}$ (μM)†
5'TERM-as	genomic 5' terminus	antisense	GCACCATACACACTTCGAGC	1 - 20	1.5
5'TERM-s	antigenomic 3' terminus	sense	GCTCGAAGTGTGTATGGTGC	1 - 20	6.5
5'HP-as	conserved hairpin in 5' UTR	antisense	GGCCCACAAGAATAGTAATT	53 - 72	3
5'HP-s	conserved hairpin in 5' UTR complement	sense	AATTACTATTCTTGTGGGCC	53 - 72	10
TRS1-as	(+) leader TRS	antisense	AAGGGTAGTTGATAGAGATC	199 - 218	3
TRS1-s	(-) leader TRS	sense	GATCTCTATCAACTACCCTT	199 - 218	>20
TRS1+AUG-as	(+) leader TRS & replicase initiator AUG	antisense	CCATAGTCGCAAGGGTAGTTGA	207 - 228	2
TRS7-as	(+) RNA7 body TRS	antisense	ACCACTACCTGAGTAGTTGA	12,252 - 12,271	7
TRS7-s	(-) RNA7 body TRS	sense	TCAACTACTCAGGTAGTGGT	12,252 - 12,271	20
3'TERM-as	genomic 3' terminus	antisense	GGTTCCTGGGTGGCTAATAACT	12,683 - 12,704	10
SCR	nonsense (control)	-	AGTCTCGACTTGCTACCTCA	-	>20

** Location of the target sequence in the EAV genome (database accession number NC_002532).*
† Approximate P-PMO concentration generating 50% reduction in the number of eGFP-positive cells; derived from Fig. 6-3B.

P-PMOs display low cytotoxicity and have EAV-specific antiviral activity

A cytotoxicity assay was performed in which each P-PMO was tested on Vero-E6 cells at concentrations of 0, 5, 10, 20, 40 or 80 μM in the culture medium. Cells were cultured with P-PMOs for 6 h in the absence of serum to allow P-PMO uptake, after which P-PMO-free medium containing 8% FCS was given. Cell proliferation was monitored by microscopy after three days of incubation. At the highest concentration (80 μM) some of the P-PMOs clearly affected cell growth, as was evident from the lower cell density compared to untreated cells. At 40 μM no morphologic growth anomalies were

apparent (data not shown), but as a precaution the P-PMO concentration never exceeded 20 µM in any subsequent experiment.

To assess P-PMO specificity and cytotoxicity, the effect of each compound on EAV amplification was compared to its effect on the amplification of the (unrelated) togavirus *Sindbis virus* (SinV). Cells were post-treated with 20 µM of each P-PMO following inoculation with EAV or SinV under identical cell culture conditions, and virus yields were measured at 36 or 24 h p.i., respectively. SinV titers in the untreated control reached approximately 10^6 PFU/ml and were not significantly altered by P-PMO treatment (Fig. 6-2). Remarkably, several P-PMOs reduced the EAV titer by several orders of magnitude or even to undetectable levels (Fig. 6-2). At 20 µM, control P-PMO SCR induced an approximately threefold reduction in EAV titer, an effect that was not observed with SinV. However, the much stronger anti-EAV effect observed with most EAV-targeted P-PMOs was clearly specific and not due to cytotoxicity or a generic barrier to cellular entry or egress of virus, as such non-specific effects would likely have caused an inhibition of SinV amplification as well.

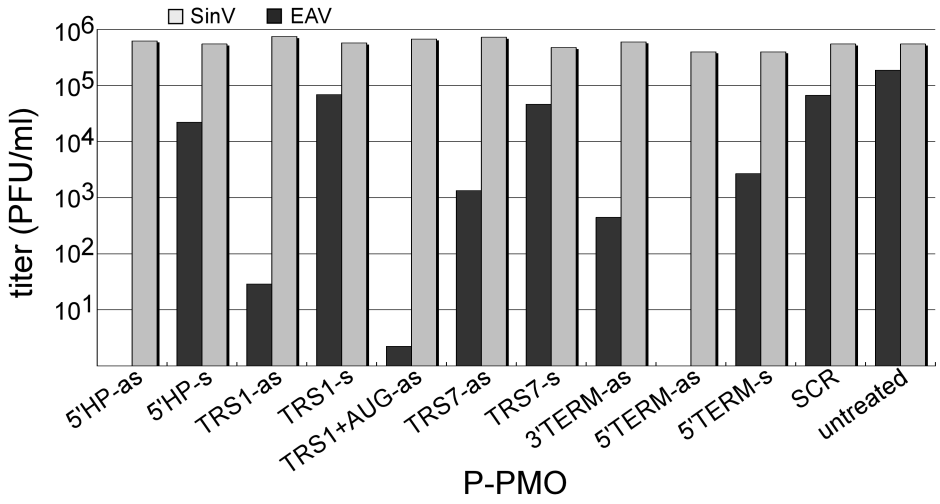


Figure 6-2. Evaluation of the efficacy and specificity of EAV-targeted P-PMOs. *Cells were infected with EAV or SinV, then treated with P-PMO at a final concentration of 20 µM in the media. Virus titers (PFU/ml) were determined by plaque assays.*

Rapid evaluation of P-PMO efficacy using a GFP-expressing recombinant EAV

To facilitate the screening for P-PMO inhibitory effects, a novel recombinant EAV (EAV-GFP) was used. A detailed description of EAV-GFP will be published elsewhere (E. van den Born, H. van den Berg and E.J. Snijder, unpublished data). Briefly, EAV-GFP expresses eGFP as an integral part of the replicase polyproteins through an in-frame

insertion at the position in ORF1a specifying the nsp1/nsp2 cleavage site of a cassette (eGFP/2A) consisting of the eGFP-coding sequence and the *Foot-and-mouth disease virus* 2A autoprotease. The nsp1-eGFP/2A junction is cleaved by the nsp1 autoprotease, whereas the 2A sequence can release the N-terminus of nsp2. Replication of EAV-GFP was not significantly affected by the eGFP insertion and its growth characteristics and virus titers were comparable to that of wild-type EAV (data not shown). EAV-GFP infected cells were readily detected by microscopy through eGFP autofluorescence (Fig. 6-3A), which allowed us to quickly screen a wide concentration range of P-PMO. Most of the eGFP expressed by EAV-GFP co-localized with other nonstructural proteins in the perinuclear region of the cell (data not shown), whereas a small portion was distributed throughout the cytoplasm (Fig. 6-3A). A number of comparative studies using wild-type EAV and EAV-GFP showed that the titer of both viruses was similarly affected by a given P-PMO concentration (data not shown), thus validating the use of EAV-GFP for the inhibition assays in this study. Consequently, the recombinant virus was used throughout the experiments described in this report.

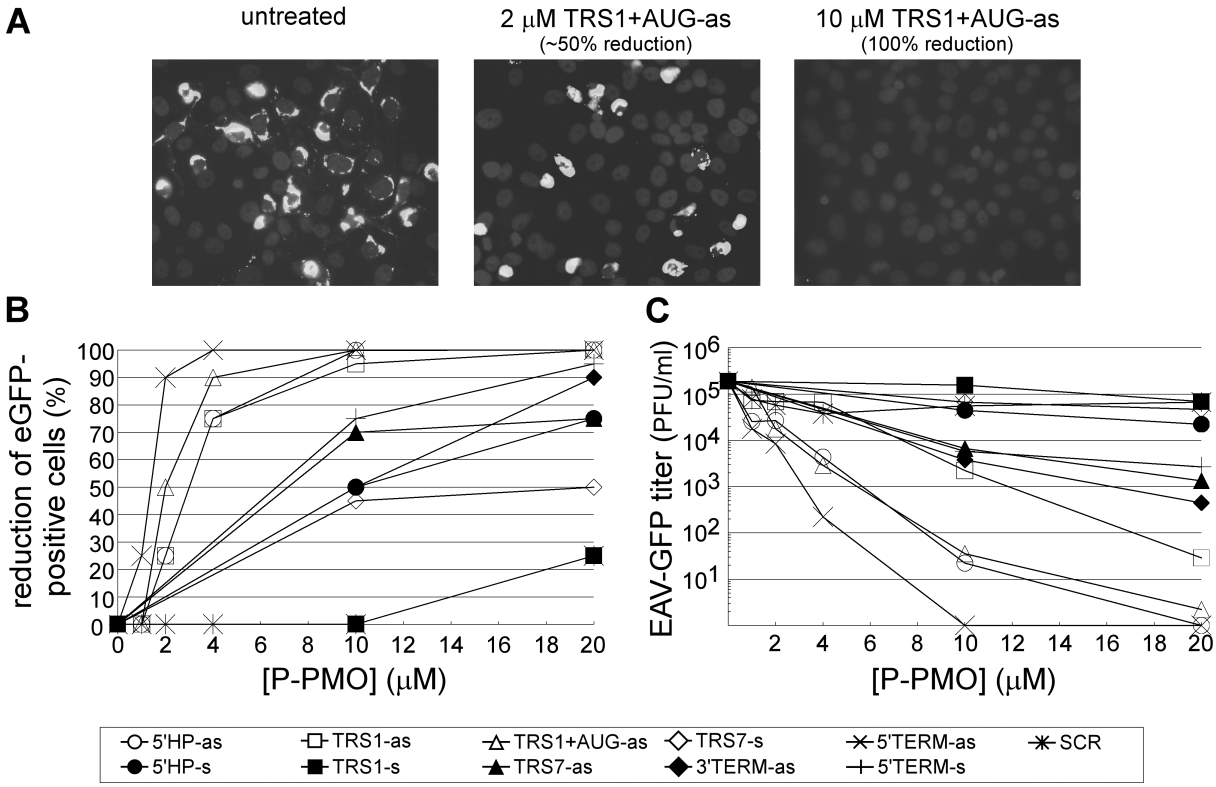


Figure 6-3. Dose-response evaluation of the antiviral activity of P-PMOs. *(A) Example of reduction in the number of EAV-GFP positive cells due to P-PMO (TRS1+AUG-as) treatment as monitored by GFP-autofluorescence microscopy. Nuclei were labeled by Hoechst staining. (B) The reduction of the number of eGFP-positive cells observed microscopically, which is indicative of the inhibition of EAV-GFP amplification, plotted against the P-PMO treatment concentration used. (C) Cell culture medium from the experiment in (B) was harvested at 36 h p.i. and virus titers (PFU/ml) determined by plaque assay.*

To further assess the efficacy of the various P-PMOs, cells were given medium with a final concentration of 0, 1, 2, 4, 10 or 20 μM of one of the compounds for 6 h prior to (pre-treatment) or following (post-treatment) EAV-GFP infection at a M.O.I. of 0.5. Intracellular eGFP expression was visually monitored at 36 h p.i. by microscopy, which revealed that several P-PMOs affected the percentage of virus-infected cells, but that individual cells that were eGFP-positive did not exhibit much variation in eGFP expression level (Fig. 6-3A). Intermediate eGFP expression was observed in one case only, when cells were treated with 3'TERM-as, although it should be noted that visual observation via microscopy is not truly a quantitative method. P-PMO pre-treatment resulted in an approximate twofold stronger inhibition than post-treatment. However, only post-treatment was used in subsequent experiments, for reasons explained in the Discussion. Remarkably, a 25-fold M.O.I. increase (to M.O.I. 12.5) did not affect the number of eGFP-positive cells present after a P-PMO treatment of any given concentration (data not shown). This finding indicated that the protective effect of a P-PMO was not overcome when cells were challenged with higher virus concentrations.

The P-PMOs could be divided into three groups, based on their ability to reduce the number of eGFP-positive cells. The first group consisted of TRS1-s and SCR, both of which had almost no effect on EAV-GFP amplification when compared to the untreated control cells. At the highest P-PMO concentration used (20 μM), a relatively small reduction of about 25% in the number of eGFP-positive cells was observed (Fig. 6-3B). A second P-PMO group, exhibiting intermediate efficacy, consisted of 5'HP-s, TRS7-as, TRS7-s, 3'TERM-as, and 5'TERM-s, which, at relatively high compound concentrations, protected the majority (50-95%) of the cells (Fig. 6-3B). The third group was comprised of the four most active P-PMOs. Three of these compounds (5'HP-as, TRS1+AUG-as, and 5'TERM-as) when present at 10 μM, and the fourth (TRS1-as) at 20 μM, were able to apparently protect all cells from virus infection (Fig. 6-3B). In summary, most of the P-PMOs generated a substantial degree of inhibition of virus amplification when administered to cells immediately after virus infection.

Several P-PMOs inhibit virus amplification in a dose-dependent manner

To quantify the relationship between dose and efficacy of the various P-PMOs, cells were post-treated at varying P-PMO concentrations for 6 h following EAV-GFP infection. Virus titers in cell culture medium harvested at 36 h p.i. were determined by plaque assays. Not unexpectedly, the relative titer reductions observed with the different P-PMO dose levels (Fig. 6-3C) were in agreement with the relative number of infected cells determined by eGFP autofluorescence (Fig. 6-3B). The group of four P-PMOs with the highest inhibition efficiency (5'TERM-as, 5'HP-as, TRS1+AUG-as, and to a lesser extent TRS1-as) generated EAV-GFP titer reductions of four logs or greater, from a peak titer of $2.0 \cdot 10^5$ PFU/ml in untreated cells, to 30 PFU/ml or below in treated cells. These

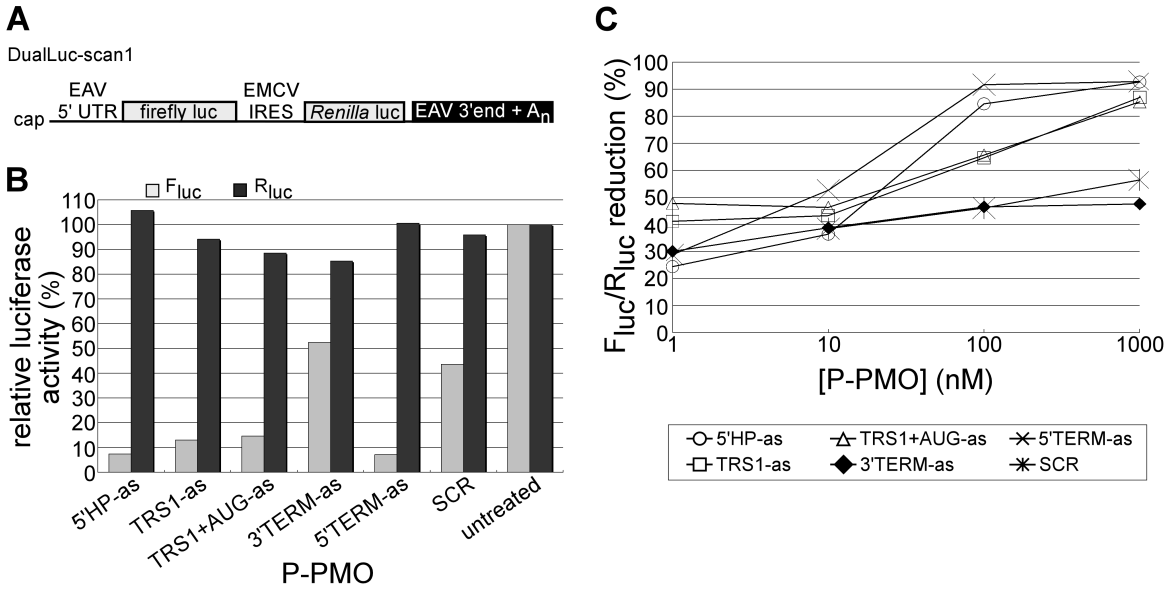


Figure 6-4. Analysis of translation inhibition by P-PMOs targeting the 5' UTR of the EAV genome. *(A) Schematic representation of the DualLuc-scan1 mRNA reporter (van den Born et al., 2005). Depicted are the EAV genomic 5' UTR, the F_{luc} reporter gene, the EMCV IRES element that drives R_{luc} expression, the R_{luc} gene (internal standard), and the 3'-terminal region of the EAV genome (starting from nt 11,737 and including a poly(A) tail). (B) Cell-free translation assay using DualLuc-scan1 mRNA in the presence of 1 µM P-PMOs. The F_{luc} and R_{luc} expression relative to that of the untreated control are presented. (C) Cell-free translation assay. The F_{luc} activity was normalized to the R_{luc} activity and the decrease in this ratio relative to that of the untreated control was plotted against the P-PMO concentration.*

four compounds are all complementary to sequences in the genomic 5' UTR, indicating that this region is the most effectual target area for this class of compounds, at least in EAV. It is noteworthy that the small amount of residual virus from the inoculum that is usually detectable in low dilutions of culture medium, was not recovered in plaque assays (Fig 6-3C; see also Fig. 6-2). The most likely explanation is that even residual amounts of highly efficacious P-PMOs prevented detectable plaque formation. Therefore, the P-PMO concentration at which virus particle production was completely prevented could not be exactly determined.

Some of the P-PMOs that do not have target sequences in the 5' UTR were able to inhibit virus amplification in a dose-dependent manner, but significant effects were observed only at relatively high concentrations (≥10 µM). TRS1-s, which targets the leader TRS complement, was completely ineffective (Fig. 6-3B & 6-3C). Moderate effects of up to one log-reduction were observed for 5'HP-s and TRS7-s (Fig. 6-3C), both of which target the minus strand. The result obtained with 5'HP-s was particularly interesting because it targeted the complement of the conserved hairpin B in the 5' UTR (Fig. 6-1A), suggesting that this sequence (and/or putative structure), located in the 3' end of the antigenome, plays a role in genome replication. Fairly strong inhibitory effects of

approximately two log-reduction were obtained with 3'TERM-as, 5'TERM-s, and TRS7-as. The former two target the 3'-terminus of the genome and antigenome, respectively, and these results suggested that they interfere with the interaction between EAV RNA and the replicase complex.

P-PMOs targeting the EAV 5' UTR can markedly interfere with translation

To investigate the effect of the above-identified highly active compounds on EAV genome translation, a cell-free translation assay was performed using the previously described pDualLuc-scan1 construct (van den Born *et al.*, 2005). Briefly, the T7 promoter of this vector directs the synthesis of a bicistronic mRNA harboring the EAV 5' UTR fused to the firefly luciferase (F_{luc}) reporter gene followed by an *Encephalomyocarditis virus* (EMCV) IRES-driven *Renilla* luciferase (R_{luc}) gene, which serves as an internal standard in the experiments. *In vitro* transcribed, capped pDualLuc-scan1 mRNA was added to rabbit reticulocyte lysate and the translation of the two reporter genes was measured. The expression of both F_{luc} and R_{luc} was analyzed in the presence of the P-PMOs targeting the 5' UTR (5'TERM-as, 5'HP-as, TRS1+AUG-as, and TRS1-as), SCR, and 3'TERM-as. The latter compound is complementary to the 3'-terminal region of the EAV genome that is also present in DualLuc-scan1 mRNA (Fig. 6-4A). Unless mRNA circularization is required for efficient translation *in vitro*, 3'TERM-as would not be expected to have activity in this system. At the highest P-PMO concentration used in this assay (1 μM), expression from the R_{luc} control cistron was not affected and could be used to normalize the F_{luc} activity (Fig. 6-4B). At nanomolar concentrations all P-PMOs targeting the 5' UTR generated dose-dependent reduction of F_{luc} expression, with a 50% inhibition concentration (IC_{50}) in the 10-100 nM range. Both controls, the SCR and 3'TERM-as P-PMOs, did not show nearly as high a reduction in translation, with IC_{50} values of approximately 1000 nM (Fig. 6-4C).

P-PMOs are unable to specifically affect sg mRNA synthesis

The discontinuous minus strand extension model for sg RNA synthesis (Sawicki & Sawicki, 1995) defined several interesting targets for antisense studies. P-PMOs TRS1-as, TRS1+AUG-as, TRS1-s, TRS7-s and TRS7-as were designed to affect sg RNA synthesis by interfering with leader TRS-to-body TRS base-pairing. Notably, the first two of these P-PMOs inhibited virus amplification likely by interfering with translation, whereas TRS7-as probably affected genome replication. To assess the effect of these P-PMOs on sg RNA production, EAV-GFP-infected cells were treated with P-PMO at the concentration that gives 50% reduction of GFP-positive cells ($IC_{50(GFP)}$; Table 6-1), which should reveal any major effects on viral sg mRNA synthesis. Intracellular RNA was

isolated at 36 h p.i. and subjected to hybridization analysis (Fig. 6-5A). The band intensities for genomic RNA and the three major sg mRNAs (2, 6 and 7) were measured by phosphorimager analysis. For each lane the ratio of sg RNA to genome synthesis was calculated and depicted in Fig. 6-5B. Although, due to common variation in experimental procedures, some fluctuation in this ratio was observed, none of the TRS-specific P-PMOs showed a specific reduction in the synthesis of all sg mRNAs (TRS1-as and TRS1&AUG-as) or sg mRNA7 (TRS7-as and TRS7-s).

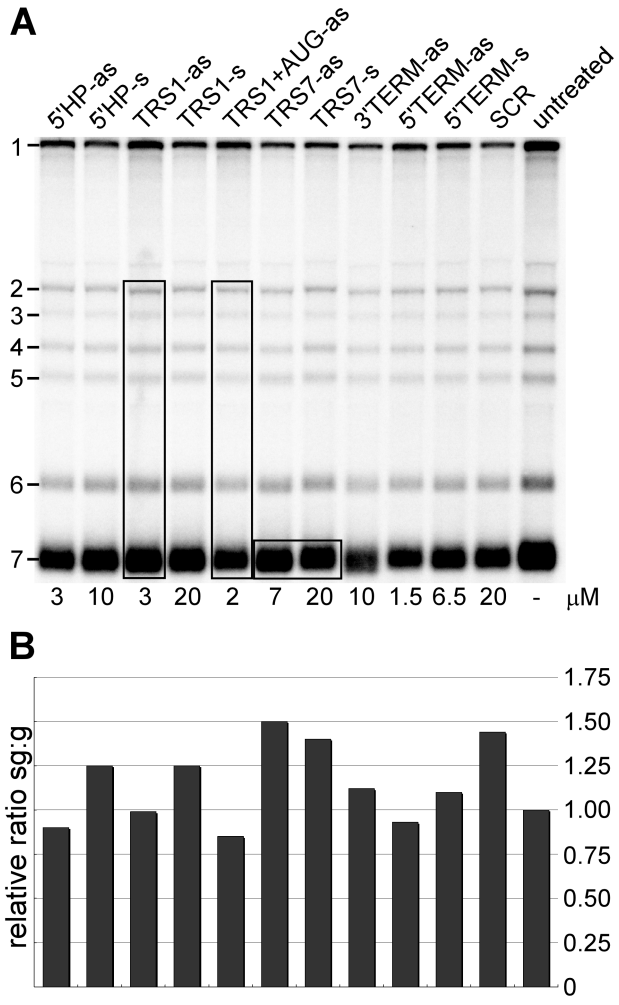


Figure 6-5. Effect of P-PMO treatment on EAV RNA synthesis. *(A) Gel hybridization analysis of EAV-specific mRNAs isolated at 36 h p.i. from EAV-GFP infected Vero-E6 cells. P-PMOs were added at the $IC_{50(GFP)}$ level determined for each of them (Table 6-1). The P-PMO concentration used is indicated below each lane. The position of the RNAs is indicated: 1, genome RNA; 2-7, sg mRNAs. Four P-PMOs were targeted to TRSs (leader TRS or RNA7 body TRS) in an attempt to specifically reduce synthesis of the sg mRNAs that are boxed. (B) Quantification of viral RNA synthesis upon P-PMO treatment, based on the hybridization experiment shown in panel A. The amount of sg mRNAs was normalized to that of the genome. This ratio (sg:g) was compared to that of the untreated control, which was set to 1.*

Discussion

Employing morpholino oligomers to inhibit RNA virus amplification has become an appealing prospect after many reports of the successful knockdown of cellular proteins via inhibition of pre-mRNA splicing and/or translation of mRNAs (Nasevicius & Ekker, 2000; Boguslavsky *et al.*, 2003; Liu *et al.*, 2003b; Muto *et al.*, 2004). RNA virus genomes offer a variety of interesting regions for PMO-targeting other than the 5' UTR, which is ordinarily selected for the purpose of inhibiting translation. Recently, two studies reported that P-PMOs targeted to RNA *cis*-acting elements involved in flavivirus genome replication successfully suppressed virus amplification (Kinney *et al.*, 2005; Deas *et al.*, 2005). In our study, we tested the ability of several P-PMOs to inhibit virus amplification by targeting RNA elements involved in three different aspects of the EAV life cycle: replication, sg mRNA synthesis, and translation. Importantly, the P-PMOs generated little cytopathic effects at concentrations lower than 40 μM, as assessed by cell proliferation assays. Moreover, SinV growth was not suppressed by any of the P-PMOs and the post-infection treatment protocol used in all of our experiments (Fig. 6-2).

The 5' UTR of the EAV genome was found to contain the most sensitive targets for inhibition with P-PMOs (Table 6-1 & Fig. 6-1). All four P-PMOs directed there were able to profoundly inhibit EAV amplification at low micromolar concentrations (Fig. 6-3B & 6-3C), with the P-PMO targeting the 5' genomic terminus (5'TERM-as) consistently generating the strongest effect. For arteriviruses and coronaviruses, the genomic 5' UTR provides an appealing target area, because all viral mRNA share a large part of this region, the common leader sequence, due to the unique mechanism of sg RNA synthesis. Consequently, P-PMOs 5'TERM-as and 5'HP-as may inhibit the translation of replicase-coding EAV genomic RNA as well as all six sg mRNAs. *In vitro* translation assays indicated that their antiviral effect was exerted, at least partly, by inhibition of translation (Fig. 6-4). A cell culture-based translation assay was also attempted, using transient expression of the luciferase reporters driven by the CMV promoter that is present, in addition to the T7 promoter, in pDualLuc-scan1. However, for unexplained reasons, P-PMO treatment equally affected the expression of both F_{luc} and the R_{luc} internal standard in a sequence-specific and dose-dependent manner (data not shown). This suggested a problem with plasmid DNA transfection or reporter mRNA synthesis or stability in the presence of P-PMO and, consequently, we had to refrain from using this assay.

Our results indicate that genome replication can also be reduced by targeting the genome termini (P-PMOs: 3'TERM-as and 5'TERM-s). Since translation assays indicated that the mode of action of 3'TERM-as was unlikely to be through interference with translation (Fig. 6-4), we assume that it exerted its inhibitory effect by acting on particular events during RNA synthesis. Rather unexpectedly, TRS7-as also significantly suppressed virus amplification (Fig. 6-3C), despite its target sequence being located approximately 450 nt from the genomic 3' end. Whether the P-PMO duplexing to this target affected the initiation of replication or perhaps blocked the processing RdRp complex is unclear.

No effect on sg mRNA synthesis was observed when leader or body TRS motifs were targeted. A likely explanation is that the target sequences for these P-PMOs were not accessible due to RNA structural or intracellular ultrastructural constraints. Genome replication and sg mRNA synthesis are thought to occur in association with dedicated double membrane vesicles (DMVs) that are believed to provide a suitable platform for viral RNA synthesis (Pedersen *et al.*, 1999). This protective microenvironment may physically prevent P-PMOs from entering replication complexes. Incoming viral genomes, which are translated in the cytoplasm in a similar manner to cellular mRNAs, are likely to be more accessible and thus susceptible to antisense P-PMO hybridization. Another possibility is that virus-encoded enzymes such as the nsp10 helicase can displace the P-PMO during sg RNA synthesis prior to the interaction between the leader TRS and a body TRS.

In all our experiments a post-treatment procedure was performed, although it has been found to be less effective compared to pre-treatment (Kinney *et al.*, 2005). However, we encountered several undesired side effects when using a P-PMO pre-treatment, probably due to the positively charged arginine-rich peptide conjugate. To a certain extent, pre-treatment with P-PMOs protected cells against virus infection, an observation that was made for both EAV and SinV. This nonspecific effect was also noted in a previous study (Kinney *et al.*, 2005). Secondly, in some cases high P-PMO concentrations actually increased virus titers as determined by plaque assays. For example, at both 4 and 10 μM of 5'TERM-as there were no eGFP-positive cells detected, but the amount of residual virus from the inoculum recovered from the 10 μM sample was approximately ten times higher than in the 4 μM sample (data not shown). This suggested that virus particles from the inoculum may adhere to the outside of cells pre-treated with certain P-PMO concentrations more than they do to non-treated cells. Finally, in cell culture-based translation studies, the efficiency of cationic-lipid mediated plasmid DNA transfection appeared to be influenced by P-PMO (data not shown). Peptide-conjugated PMOs have been shown to accumulate in cellular membranes (Moulton *et al.*, 2003; Moulton *et al.*, 2004). This accumulation of charged molecules may interfere with virus entry (and release) as well as cationic-lipid uptake. A post-infection P-PMO treatment strategy can be recommended to avoid these problems, and was done for all the experiments reported here.

We consistently observed that for any individual cell P-PMO treatment using our eGFP-expressing EAV-GFP recombinant generally resulted in an “all or nothing” outcome of infection. Either eGFP expression was completely absent, indicating that EAV-GFP was not able to replicate, or a bright fluorescence was observed, indicating that a vigorous virus infection was in progress. This phenomenon could be explained by low-level EAV-GFP replication in the presence of P-PMO, slowly generating more and more viral RNA. At some point, the number of viral targets could reach a threshold level over which the number of P-PMO molecules becomes inconsequential, allowing the virus to replicate robustly.

P-PMOs appear to have potential as antisense therapeutics *in vivo*. They behave with predictability and potency, and displayed little nonspecificity in most of our systems. The ability of a post-infection P-PMO treatment to significantly reduce the degree of viral amplification further supports the idea that this class of compounds may be suitable for therapeutic development. Delivery of PMOs into cells can clearly be enhanced by using peptide conjugation, which may prove to be an important factor in future drug design. One obstacle may be that serum reduces the cellular uptake of the type of arginine-rich peptide used in this study (Neuman *et al.*, 2004). However, new types of peptide conjugates may be able to overcome this problem (Deas *et al.*, 2005).

Arteri- and coronaviruses are a major economic problem in the swine and cattle breeding industry and coronaviruses also cause a considerable amount of human disease, such as SARS and a substantial percentage of cases of the common cold syndrome (Saif *et al.*, 1988; Albina, 1997; Makela *et al.*, 1998; Cook, 2002; Peiris *et al.*, 2003). The arterivirus EAV has proven to be a safe and convenient tool to study various aspects of nidovirus molecular biology, in particular viral RNA synthesis and replicase function. Continued *in vitro* studies and *in vivo* approaches with various types of P-PMOs will be required to assess the potential of these compounds to protect against and/or treat diseases caused by nidoviruses. In any case, it is clear from this study that these compounds represent useful reagents for exploring various aspects of molecular virology, especially translation mechanics.

Materials & Methods

Cells and viruses

Vero-E6 cells were cultured in Dulbecco's modified Eagle's medium DMEM containing 8% fetal calf serum (FCS) for maintenance or in DMEM without FCS or with 4% FCS, as stated, during P-PMO studies. EAV-GFP (see below for details) and *Sindbis virus* (SinV; HR strain) were both propagated in Vero-E6 cells at 37°C. Virus titration using plaque assays was described previously (Molenkamp *et al.*, 2000a).

RNA isolation and analysis

Intracellular RNA from Vero-E6 cells was isolated at 36 h post infection (p.i.), separated in denaturing agarose-formaldehyde gels and analyzed by hybridization to a ^{32}P-labeled probe, which is complementary to the 3' end of the EAV genome and all sg mRNAs (5'-TTGGTTCCTGGGTGGCTAATAACTACTT-3') (van Marle *et al.*, 1999a). Dried gels were exposed to phosphorimager screens, and scanned with a Personal Molecular Imager FX (Bio-Rad) after exposure. Band intensities were quantitated with Quantity One v4.2.2 (Bio-Rad).

P-PMO design and synthesis

Morpholino oligomers were designed to be complementary to target sequences in the EAV genome or antigenome. See Table 6-1, Fig. 6-1, and Results for P-PMO sequences and target locations. PMO is a single-stranded DNA analog that has a morpholine ring in place of each riboside moiety and phosphorodiamidate intersubunit linkages instead of phosphorodiester linkage (Summerton & Weller, 1997). The nucleoside bases attached to this backbone are the same as for DNA. In order to deliver the PMOs into cultured cells, an arginine-rich peptide NH_2-RRRRRRRRRFF-$CONH_2$, designated as R_9F_2C (Moulton *et al.*, 2004), was covalently linked to the 5' end of each PMO. The peptide-conjugated PMO is referred to herein as P-PMO. The performance of the R9F2C peptide varies with the amount of serum in the media (Neuman *et al.*, 2004; Moulton *et al.*, 2004). Hence, the 6-hour P-PMO treatments were carried out in the absence of serum. All PMOs were synthesized at AVI BioPharma Inc. by methods described previously (Summerton & Weller, 1997). The conjugation, purification and analysis of P-PMO compounds was carried out at AVI BioPharma according to methods described elsewhere (Moulton *et al.*, 2004).

P-PMO studies

Vero-E6 cells were grown to approximately 80% confluency. For pre-treatment experiments, Vero-E6 cells were incubated for 6 hours in serum-free DMEM containing different concentrations of P-PMOs. After removal of the medium and washing, the cells were infected with EAV-GFP or SinV at a multiplicity of infection (M.O.I.) of 0.5 PFU/cell for 1 hour. Subsequently, the inoculum was removed and fresh DMEM containing 4% FCS was added. For post-treatment experiments, Vero-E6 cells were infected as above, the inoculum was then removed, the cells rinsed, and P-PMO treatment applied for 6 hours, after which the compound was not removed, but 1 volume of fresh medium was added, resulting in a final concentration of FCS of 4%. As a consequence, the residual concentration of P-PMOs was halved.

Translation assays

To test the effect on translation by various P-PMOs, cell-free translation assays were carried out using the previously described pDualLuc-scan1 reporter gene construct for EAV genome translation (van den Born *et al.*, 2005). *In vitro* transcribed, capped bicistronic pDualLuc-scan1 transcripts were *in vitro* translated in the presence of various P-PMOs using the Rabbit Reticulocyte Lysate System according to the manufacture's protocol (Promega). Luciferase expression was measured using a luminometer (Turner Designs TD-20/20) and the Dual-luciferase Reporter Assay (Promega) on the basis of enzymatic activity.

Acknowledgements

We would like to thank the Chemistry Department at AVI BioPharma Inc. for expert synthesis, purification, and analysis of peptide-PMOs. We appreciate the technical assistance of Karin Wever and Erik van den Berg in the construction of the EAV-GFP recombinant. E. van den Born was supported by Council for Chemical Sciences grant CW 99-010 from the Netherlands Organisation for Scientific Research.

Chapter 7

Discussion

Development of reverse genetics systems to study arteri- and coronavirus RNA synthesis

Despite more than two decades of research into the molecular details of nidovirus subgenomic (sg) mRNA synthesis, and despite the recent contributions of *Severe acute respiratory syndrome-coronavirus* (SARS-CoV) inspired research, many mechanistic details of this transcription mechanism remain to be unraveled. Our current understanding of coronavirus sg mRNA synthesis is largely based on studies using defective interfering (DI) RNA replicons containing one or more (often artificial) transcription-regulating sequences (TRSs). Although these systems may represent sg mRNA synthesis as it occurs during the life cycle of the wild-type virus, conclusions have to be drawn with caution. High-frequency recombination between the DI RNA and the helper virus genome has been frequently observed. In addition to that complication, the amount of sg mRNAs produced by DI RNA replicons is generally much lower than that generated from the corresponding TRS in the genome of the parental virus. Consequently, it is not clear how directly the conclusions obtained with DI RNA systems can be extrapolated to sg RNA synthesis from the full-length genome.

The recent development of full-length infectious cDNA clones for several coronaviruses has created the possibility to study transcription in a more authentic setting, which is a significant step forward (Yount *et al.*, 2000; Almazan *et al.*, 2000; Casais *et al.*, 2001; Yount *et al.*, 2002; Yount *et al.*, 2003; Coley *et al.*, 2005). On the other hand, the reverse genetics system developed for coronaviruses thus far are quite laborious and the large size of the coronavirus genome often hampers straightforward genetic engineering. The development of a reverse genetic system for SARS-CoV was an important achievement (Yount *et al.*, 2003), although the pathogenicity of SARS-CoV for humans is a major disadvantage in experimental terms. Working with this virus, or its cloned full-length cDNA derivative, requires dedicated bio-safety facilities and well-trained staff.

Over the past decade, the arterivirus *Equine arteritis virus* (EAV) has proven to be a safe and convenient tool to study various aspects of nidovirus molecular biology, in particular viral RNA synthesis and replicase function. The reverse genetics system developed at LUMC (van Dinten *et al.*, 1997) is based on a convenient and stable plasmid (pUC18) that can be propagated in *E. coli*. The system allows the modification of essentially every nucleotide of the genome and rapid screening of the consequences of these changes in a first cycle analysis. Reverse genetics systems for EAV and other nidoviruses have significantly facilitated the functional dissection of RNA elements and proteins involved in nidovirus sg mRNA synthesis, and will certainly continue to do so in the near future. However, it should be noted that reverse genetics systems also have their limitations, because intermediate steps in sg mRNA synthesis are difficult to pinpoint and analyze. For the systematic dissection of the mechanistic details of nidovirus transcription *in vitro* systems for RNA synthesis are required.

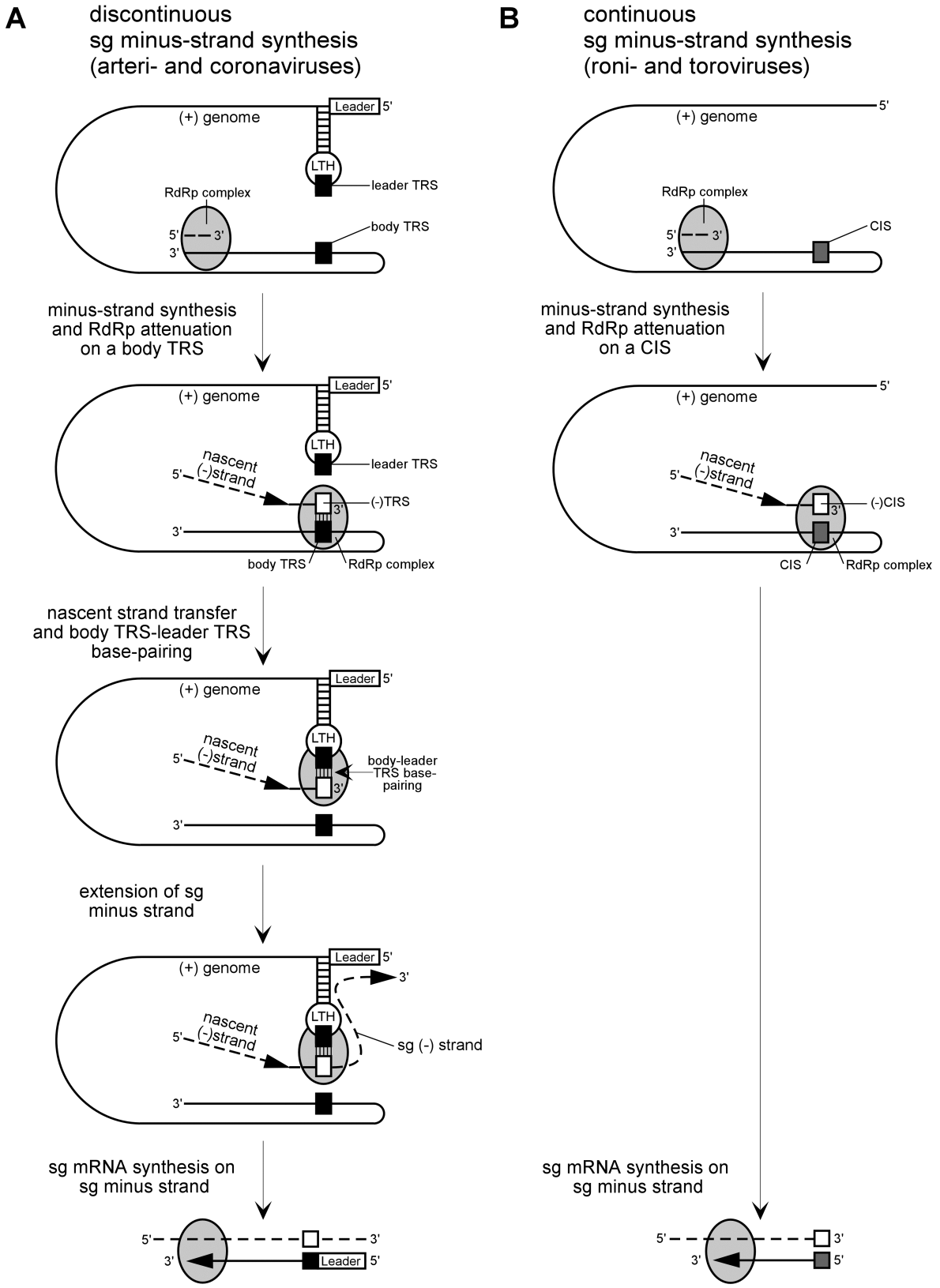


Figure 7-1. Model for nidovirus discontinuous (A) and continuous (B) sg minus strand synthesis. *See text for details.*

Discontinuous and continuous RNA synthesis in the *Nidovirales*

Nidoviruses share basic features of their genome organization and expression. One feature that unites nidoviruses, and that is reflected in the name of the taxon (*nidus* in Latin means nest), is the generation of a 3'-coterminal nested set of mRNAs during replication in infected cells. The sg mRNA produced by members of the genera *Arterivirus* and *Coronavirus* also possess a common 5' sequence, the so-called leader sequence. The leader and body segments of sg mRNAs, which are located at opposite ends of the genome, are fused following (+) leader TRS to (-) body TRS base-pairing, which most likely occurs during the discontinuous synthesis of sg minus strands (Fig. 7-1A) (Sawicki & Sawicki, 1995; van Marle *et al.*, 1999a; Pasternak *et al.*, 2001). During this process, the body TRS and its flanking sequences in the genome may act as a termination signal for the complex involved in minus-strand RNA synthesis. The attenuated nascent minus strand is extended with the leader complement following its transfer to the leader TRS, thus generating the chimeric template for sg mRNA synthesis. The fidelity of this translocation process is promoted by a base-pairing interaction between the body TRS complement in the nascent minus strand and the leader TRS in the genomic plus strand (Chapter 3). Thus, the body TRS may fulfil a dual role as attenuation signal and as director of the recombination-like strand transfer event. Since minus strand synthesis presumably terminates prior to nascent strand transfer, the arteri- and coronavirus discontinuous minus strand extension model shares characteristics with the premature termination model of sg RNA synthesis (Sawicki & Sawicki, 1995; Sit *et al.*, 1998; Miller & Koev, 2000; White, 2002). The presence of a 5’ common leader on arteri- and coronavirus sg mRNAs implies that at least a part of the mechanism underlying their synthesis is unique.

In contrast to arteri- and coronaviruses, viruses belonging to the genera *Ronivirus* and *Torovirus* produce sg mRNAs lacking a common leader sequence, with the exception of the largest sg mRNA (encoding the spike protein) of toroviruses (Snijder *et al.*, 1990b; Snijder *et al.*, 1991; van Vliet *et al.*, 2002; Cowley *et al.*, 2002). Thus, the latter virus group has been proposed to combine two different transcription strategies, discontinuous and continuous minus-strand RNA synthesis, placing them in an intermediate and interesting position in the evolution of nidovirus sg mRNA synthesis (van Vliet *et al.*, 2002). Consequently, the production of the sg mRNAs of roniviruses- and the three smallest sg mRNAs of toroviruses does not involve a discontinuous step and may therefore be fully compatible with the premature termination model (Fig. 7-1B) (Miller & Koev, 2000; White, 2002). The mechanism for sg mRNA synthesis used by roni- and toroviruses appears to be similar to that of the closteroviruses, a family of plus-strand RNA viruses of plants that also has a polycistronic genome and produces an extensive set of 3'-coterminal sg mRNAs (Ayllon *et al.*, 2004).

The LTH and the leader switching phenomenon

During mixed infections with different MHV strains, genomic leader sequences can be exchanged at a high frequency (Makino *et al.*, 1986). Moreover, in a number of studies, it was shown that leader sequences could be swapped between the genomes of helper viruses and DI RNAs, a phenomenon termed "leader switching". The cross-over site of this high-frequency recombination event was mapped to a region directly downstream of the leader TRS (Makino & Lai, 1989; Chang *et al.*, 1994; Chang *et al.*, 1996; Stirrups *et al.*, 2000). In a MHV DI RNA system, a palindromic 9-nt sequence motif (5'-UUUAUAAAC-3') was found to be essential for leader switching (Makino & Lai, 1989). Furthermore, there are some indications that this motif in MHV is an enhancer element in sg mRNA synthesis (Zhang *et al.*, 1994; Wang & Zhang, 2000). The occurrence of leader switching underlines once again that data obtained with DI RNA replicon systems have to be interpreted with caution, because these systems are prone to recombination. Interestingly, the cross-over site, including the 9-nt sequence motif, maps to the 3' portion of the stem of the predicted coronavirus LTHs (Fig. 4-8). This suggests that the LTH region is a hot spot acceptor or donor site during recombination between co-replicating genomes or co-replicating genome and DI RNA. This notion is in agreement with the proposed function of the LTH as an acceptor site in a similarity-assisted recombination-like process during discontinuous minus strand synthesis (Sawicki & Sawicki, 1995; Brian & Spaan, 1997; Nagy & Simon, 1997; Pasternak *et al.*, 2001). Additional support comes from our finding that the arterivirus LTH could maintain its functionality after being moved to a position in the EAV genome approximately 300 nt more downstream of its original location (chapter 5).

Verification of the conformation of the coronavirus LTH structures

Chapter 4 and 5 provide genetic support for the idea that the LTH structure is a key regulator of EAV sg mRNA synthesis, and is the first identified RNA structure involved in nidovirus discontinuous sg RNA synthesis. Similar structures were predicted for the 5'-proximal region of other arterivirus genomes and the genomes of a number of coronaviruses, suggesting a common function in sg mRNA synthesis in both virus groups. It should be mentioned that the proposed LTH structure of several arteri- and coronaviruses (Chapter 4) conflicts with some of the RNA secondary structure models of the 5'-proximal region of the genome predicted by others (Stirrups *et al.*, 2000; Tan *et al.*, 2001; Raman *et al.*, 2003). If both predictions would represent biologically relevant RNA elements, this would suggest that the coronavirus LTH is involved in a conformational switch similar to the one that was proposed for the arterivirus LTH (Chapter 5). Alternatively, the observed inconsistency may result from inadequate structure predictions and lack of supporting biochemical data. The limited number of available viral sequences

prevented a detailed phylogenetic analysis, using nucleotide variations between virus isolates, to support the structures depicted in Chapter 4 . Recently, two almost identical and novel human coronaviruses, HCoV-NL and HCoV-NL63, were isolated from patient material in The Netherlands and were found to be related to HCoV-229E (Fouchier *et al.*, 2004; van der Hoek *et al.*, 2004). Their genome sequences were used to verify the predicted HCoV-229E LTH structure (Fig. 4-8). The analysis revealed that a modified structure model could accommodate the observed sequence variations, including two covariations, more adequately than the previous structure prediction (Fig. 7-1A). Using a similar approach, the *Transmissible gastroenteritis virus* (TGEV) LTH structure was adapted (Fig. 7-1B) to accommodate some sequence differences with the related *Feline coronavirus* (FCoV; database accession number DQ010921). After publication of Chapter 4, a few other new coronavirus sequences have become available. The sequence of a novel human coronavirus (HCoV-HKU1) was obtained from an elderly Chinese patient suffering from pneumonia, and was found to be most closely related to HCoV-OC43 (Woo *et al.*, 2005). Furthermore, sequences of *Equine coronavirus* (ECoV), *Porcine hemagglutinating encephalomyelitis virus* (HEV), and *Human enteric coronavirus* (HECoV) were deposited in the database (accession numbers AF523846, AF523845, and AF523844, respectively). RNA secondary structure modeling of the leader TRS region in those sequences, predicted similar LTH structures in each case (Fig. 7-1C). The fact that a LTH can be predicted in the 5'-proximal region of the genome of each member of the *Coronavirus* genus (Fig. 4-8 & 7-1) and the fact that the prediction of the *Bovine coronavirus* (BCoV) LTH was corroborated by biochemical assays (Chang *et al.*, 1996; Raman *et al.*, 2003), support the relevance of this structure in the coronavirus life cycle. Nevertheless, biochemical and genetic data will be required to establish reliable RNA secondary structure models of the 5'-proximal region of additional arteri- and coronavirus genomes and the function of these RNA elements will have to be confirmed in biological assays. The strategy (applied in Chapter 5) to functionally separate otherwise overlapping RNA signals may prove extremely useful during further studies of this region and the analysis of the existence and function of a coronavirus LTH-equivalent in particular.

Determinants of the attenuation step during discontinuous minus-strand RNA synthesis

Several studies have provided evidence that the base-pairing interaction between the leader TRS and a body TRS complement is an essential step during arteri- and coronavirus sg mRNA synthesis (Joo & Makino, 1992; van Marle *et al.*, 1999a; Pasternak *et al.*, 2001; Pasternak *et al.*, 2003; Zuniga *et al.*, 2004). In addition to this base-pairing interaction, Chapter 3 also provides genetic support that body TRSs have a sequence-specific function in sg mRNA synthesis that is distinct from the formation of a duplex with the leader TRS. However, it is still unclear whether the TRS nucleotides determine

the site where synthesis of the nascent minus strand is attenuated. Preliminary data demonstrated that, during *in vitro* reverse transcription from plus-stranded EAV genomic templates, the reverse transcriptase could be specifically stalled at body TRSs in the absence of additional proteins, suggesting that the body TRS region can attenuate the progressing RNA-dependent DNA polymerase (data not shown). RNA secondary structure can be a crucial element of sg RNA promoters and has been found to be involved in the termination of minus strand synthesis (premature termination model) and the

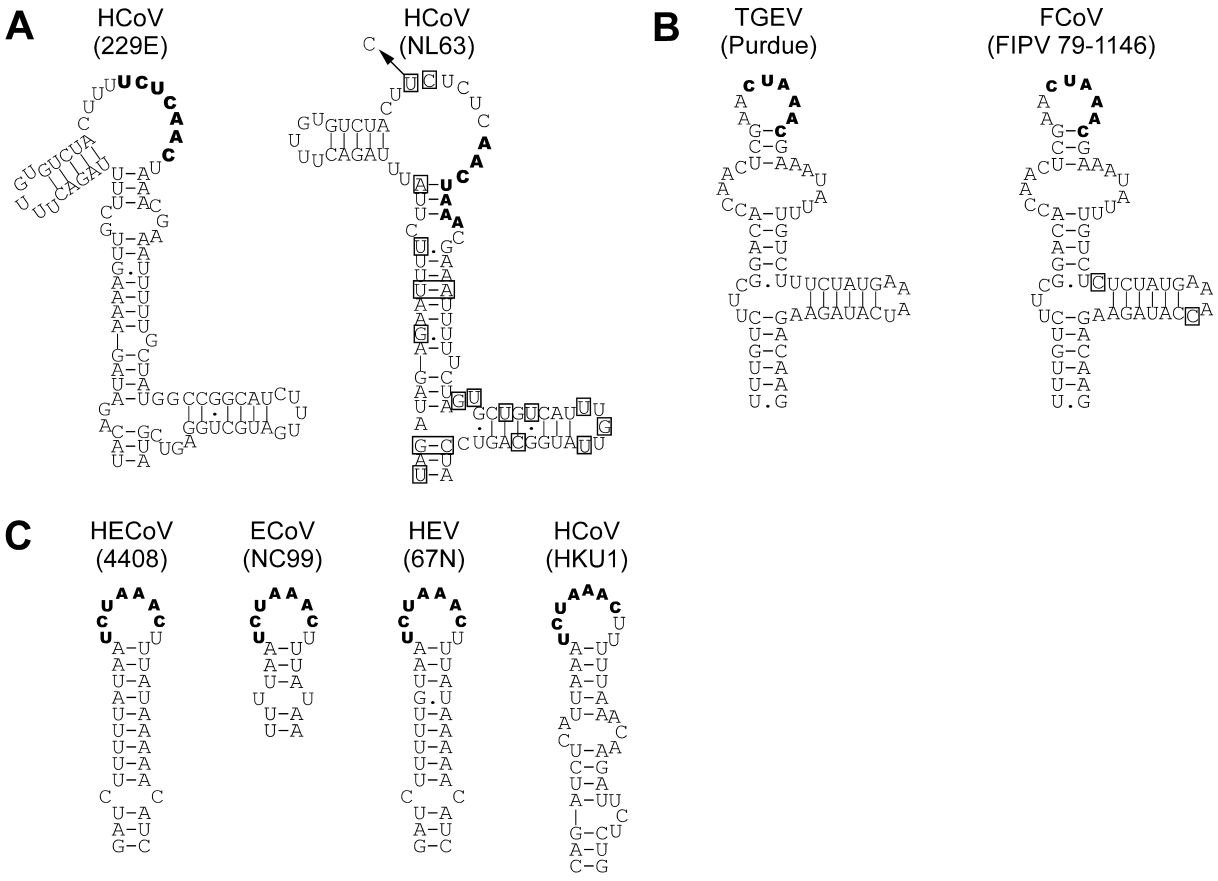


Figure 7-2. RNA secondary structure predictions of a number of coronavirus LTHs. *(A) Two similar LTH structures are presented for HCoV strains 229E and NL63. Note that the former is an alternative conformation of the structure depicted in Fig. 4-8. Conserved leader TRS nucleotides are shown in bold font (Herold et al., 1993; Pyrc et al., 2004). Boxed nucleotides indicated sequence variation between HCoV-229E and HCoV-NL63. The single nucleotide difference between HCoV strains NL63 and NL (Fouchier et al., 2004) is indicated with an arrow. (B) Two similar LTH structures are presented for TGEV strain Purdue and FCoV strain FIPV 79-1146 (database accession number DQ010921), the former is an alternative conformation of the LTH depicted in Fig. 4-8. Conserved leader TRS nucleotides are shown in bold font, and are for FCoV based on sequence comparison with TGEV (Alonso et al., 2002). Boxed nucleotides indicate sequence variation between TGEV and FCoV. (C) LTH predictions for HECoV strain 4408, ECoV strain NC99, HEV strain 67N and HCoV strain HKU1 (database accession numbers AF523844, AF523846, and AF523845, and Woo et al. (2005), respectively). Conserved leader TRS nucleotides are shown in bold font, and are based on sequence comparison with BCoV (Ozdarendeli et al., 2001).*

initiation of sg plus strand synthesis (internal initiation model) (Sit *et al.*, 1998; Lindenbach *et al.*, 2002; Ayllon *et al.*, 2004; Lin & White, 2004). Although proposed in several studies, RNA secondary or higher order structures have not been convincingly implicated in chain termination during nidovirus RNA synthesis. It was suggested that the body TRS nucleotides and their flanking sequences may contribute to stalling the RdRp (van Marle *et al.*, 1995; Alonso *et al.*, 2002; Pasternak *et al.*, 2004; Curtis *et al.*, 2004). The RNA sequence or structure may exert this function directly or indirectly, through an interaction with a regulatory protein factor. Obvious candidates for a role as a regulatory factor are components of the viral replicase. In the case of EAV, replicase subunits nsp10, in particular nsp1 have been implicated, to a variable degree, in sg mRNA synthesis (van Dinten *et al.*, 1997; van Marle *et al.*, 1999b; Tijms *et al.*, 2001).

After premature termination of minus-strand RNA synthesis and prior to leader-body TRS base-pairing, the nascent strand with the body TRS complement at its 3' end presumably is transferred to the leader TRS, presented by the LTH. The LTH structure may act as a platform to facilitate the leader TRS-body TRS interaction. The template for the synthesis of the sg mRNA is completed after the successful addition of the leader complement, using as template the genomic leader, including the part of the LTH located upstream of the leader TRS. The relatively low stability of the top of the LTH stem, which is characterized by a large number of A-U base-pairs, and/or the presence of a large bulge a few base-pairs below the LTH loop, may promote the unwinding of the LTH structure that is required for its use as template. In a variant of this model, recently proposed for coronaviruses, a 5'-3' interaction in the genome would promote the formation of a leader TRS-containing complex that would scan the nascent minus strand to identify the body TRS complement for base-pairing, whereafter the RdRp would switch to the leader template (Zuniga *et al.*, 2004; Sola *et al.*, 2005).

Previously, a model was proposed to explain the termination of sg minus strand synthesis of *Red clover necrotic mosaic virus* (RCNMV) (Sit *et al.*, 1998). It was demonstrated that a hairpin structure in the trans-activator element of RNA-1 base-pairs with a region within the sg promoter in RNA-2, thereby physically blocking the progressing RdRp involved in RNA2 minus strand synthesis. When in close proximity to a body TRS, the LTH may be involved in a similar interaction to terminate or attenuate sg minus strand synthesis. Such a long distance RNA-RNA interaction is reminiscent of the "RNA network" identified in tombusviruses, which is involved in premature termination of minus-strand RNA synthesis to produce the template for sg mRNA synthesis (Lin & White, 2004). However, in the case of arteriviruses and coronaviruses, a direct RNA-RNA interaction between the genomic leader TRS in the LTH loop and body TRSs in the 3'-proximal part of the genome is not possible since these sequences are identical and not complementary. Interacting proteins or additional regulatory RNA elements may therefore direct this process. A protein candidate is polypyrimidine-tract-binding protein (PTB), which was found to have affinity for the leader TRS (Furuya & Lai, 1993; Li *et al.*, 1999).

Furthermore, the body TRS-flanking sequences may determine the accessibility of the body TRS and may provide a structural platform for an interaction with the LTH.

Replication and translation signals in the 5'-proximal region of the EAV genome

The minimal replication signals in the 5'-proximal region of the EAV genome presumably reside in the first 296 nucleotides (Tijms *et al.*, 2001). Chapter 4 provides a detailed RNA secondary structure model for this region, supported by biochemical and genetic data, that contains the LTH. Stem and loop mutations in the LTH structure were designed to alter its conformation and had a detrimental effect on genome replication, in addition to the postulated effect on sg mRNA synthesis (Fig. 4-5). In agreement with our findings, there are some indications that the BCoV LTH is also involved in DI RNA replication, since its deletion completely abrogated the amplification of a DI RNA replicon (Chang *et al.*, 1996). It should be noted that the mutations designed to alter the EAV LTH conformation may have had an indirect effect on replication due to conformational changes in the (overall) RNA secondary or tertiary structure of the 5'-proximal domain. Although the EAV LTH stem harbors the replicase translation initiation codon, a role of the LTH in translation is rather unlikely, because modification of the initiation codon context in the leader duplication constructs described in Chapter 5 did not affect the viability of the corresponding mutants. The effects observed at the level of replication (Chapter 4) and transcription (Chapter 4 and 5) emphasize once more that this region of the EAV genome is multifunctional. The RNA structure model and the leader duplication mutants now available will be important tools in a further assessment of the function of RNA elements located in the 5'-proximal region of the EAV genome.

From fundamental to applied science

The arterivirus prototype EAV is a safe, convenient, and powerful tool for fundamental research into nidovirus RNA synthesis, as illustrated in Chapters 3, 4 and 5. This model system has provided us with valuable insights in the molecular biology of nidoviruses in general and arteriviruses in particular. In this thesis, I have characterized the function of RNA elements involved in arterivirus sg mRNA synthesis, have studied the translation mechanism of the EAV genome, and have identified RNA elements putatively involved in replication. This knowledge was used for the rational selection of EAV RNA targets in an inhibition study using antisense morpholino oligomers (Chapter 6). Some of these compounds were able to block EAV amplification completely, and may have antiviral potential at relatively low concentrations. This finding is in agreement with

two recent studies in which antisense morpholino oligomers were applied to inhibit MHV and SARS-CoV amplification (Neuman *et al.*, 2004; Neuman *et al.*, 2005). SARS-CoV replication was also reduced by another type of antisense oligonucleotide (Shi *et al.*, 2004), and by small interfering RNA molecules (Wang *et al.*, 2004; Wu *et al.*, 2005), indicating that these compound may be developed into candidate drugs against this human pathogen. With one antisense drug (Vitravene) marketed in 2004 and approximately 20 others in clinical development, it appears that antisense drugs may prove of value in the treatment of a wide range of diseases (Holmlund, 2003; Crooke, 2004). Future studies are required to test the efficacy and toxicity of nidovirus-specific morpholino oligomers in animal model systems. In this manner, these studies may provide a basis for the development of antiviral strategies that may aid in combating diseases caused by human and animal nidoviruses.

References

Aaziz, R. & Tepfer, M. (1999). Recombination in RNA viruses and in virus-resistant transgenic plants. *J Gen Virol* **80 (Pt 6)**, 1339-1346.

Adam, M. A. & Miller, A. D. (1988). Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J Virol* **62**, 3802-3806.

Adkins, S., Siegel, R. W., Sun, J. H. & Kao, C. C. (1997). Minimal templates directing accurate initiation of subgenomic RNA synthesis in vitro by the brome mosaic virus RNA-dependent RNA polymerase. *RNA* **3**, 634-647.

Agrawal, S., Gupta, D. & Panda, S. K. (2001). The 3' end of hepatitis E virus (HEV) genome binds specifically to the viral RNA-dependent RNA polymerase (RdRp). *Virology* **282**, 87-101.

Albina, E. (1997). Epidemiology of porcine reproductive and respiratory syndrome (PRRS): an overview. *Vet Microbiol* **55**, 309-316.

Almazan, F., Galan, C. & Enjuanes, L. (2004). The nucleoprotein is required for efficient coronavirus genome replication. *J Virol* **78**, 12683-12688.

Almazan, F., Gonzalez, J. M., Penzes, Z., Izeta, A., Calvo, E., Plana-Duran, J. & Enjuanes, L. (2000). Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5516-5521.

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**, 1293-1308.

An, S. W. & Makino, S. (1998). Characterizations of coronavirus cis-acting RNA elements and the transcription step affecting its transcription efficiency. *Virology* **243**, 198-207.

Archambault, D., St Louis, M. C. & Martin, S. (2005). Binding of cellular proteins to the leader RNA of equine arteritis virus. *Virus Genes* **30**, 121-125.

Ayllon, M. A., Gowda, S., Satyanarayana, T. & Dawson, W. O. (2004). cis-acting elements at opposite ends of the Citrus tristeza virus genome differ in initiation and termination of subgenomic RNAs. *Virology* **322**, 41-50.

Baker, S. C. & Lai, M. M. C. (1990). An Invitro System for the Leader-Primed Transcription of Coronavirus Messenger-Rnas. *Embo Journal* **9**, 4173-4179.

Baranov, P. V., Henderson, C. M., Anderson, C. B., Gesteland, R. F., Atkins, J. F. & Howard, M. T. (2005). Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* **332**, 498-510.

Baric, R. S., Curtis, K. M. & Yount, B. (2001). MHV subgenomic negative strand function. *Adv Exp Med Biol* **494**, 459-465.

Baric, R. S., Shieh, C. K., Stohlman, S. A. & Lai, M. M. (1987). Analysis of intracellular small RNAs of mouse hepatitis virus: evidence for discontinuous transcription. *Virology* **156**, 342-354.

Baric, R. S., Stohlman, S. A. & Lai, M. M. (1983). Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. *J Virol* **48**, 633-640.

Baric, R. S., Stohlman, S. A., Razavi, M. K. & Lai, M. M. (1985). Characterization of leader-related small RNAs in coronavirus-infected cells: further evidence for leader-primed mechanism of transcription. *Virus Res* **3**, 19-33.

Baric, R. S. & Yount, B. (2000). Subgenomic negative-strand RNA function during mouse hepatitis virus infection. *J Virol* **74**, 4039-4046.

Berger, A., Drosten, C., Doerr, H. W., Sturmer, M. & Preiser, W. (2004). Severe acute respiratory syndrome (SARS)--paradigm of an emerging viral infection. *J Clin Virol* **29**, 13-22.

Blackwell, J. L. & Brinton, M. A. (1997). Translation elongation factor-1 alpha interacts with the 3' stem-loop region of West Nile virus genomic RNA. *J Virol* **71**, 6433-6444.

Boguslavsky, D., Ierusalimsky, V., Malyshev, A., Balaban, P. & Belyavsky, A. (2003). Selective blockade of gene expression in a single identified snail neuron. *Neuroscience* **119**, 15-18.

Bol, J. F. (2005). Replication of Alfamo- and Ilarviruses: Role of the Coat Protein. *Annu Rev Phytopathol* .

Bos, E. C., Dobbe, J. C., Luytjes, W. & Spaan, W. J. (1997). A subgenomic mRNA transcript of the coronavirus mouse hepatitis virus strain A59 defective interfering (DI) RNA is packaged when it contains the DI packaging signal. *J Virol* **71**, 5684-5687.

Brault, V. & Miller, W. A. (1992). Translational frameshifting mediated by a viral sequence in plant cells. *Proc Natl Acad Sci U S A* **89**, 2262-2266.

Bredenbeek, P. J., Pachuk, C. J., Noten, A. F., Charite, J., Luytjes, W., Weiss, S. R. & Spaan, W. J. (1990). The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res* **18**, 1825-1832.

Brian, D. A. & Spaan, W. J. M. (1997). Recombination and coronavirus defective interfering RNAs. *Semin Virol* **8**, 101-111.

Brierley, I., Boursnell, M. E., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. & Inglis, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J* **6**, 3779-3785.

Brierley, I., Digard, P. & Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537-547.

Brierley, I., Jenner, A. J. & Inglis, S. C. (1992). Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* **227**, 463-479.

Brierley, I., Rolley, N. J., Jenner, A. J. & Inglis, S. C. (1991). Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* **220**, 889-902.

Casais, R., Thiel, V., Siddell, S. G., Cavanagh, D. & Britton, P. (2001). Reverse genetics system for the avian coronavirus infectious bronchitis virus. *Journal of Virology* **75**, 12359-12369.

Cavanagh, D. (1997). Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Arch Virol* **142**, 629-633.

Cech, T. R. (2002). Ribozymes, the first 20 years. *Biochem Soc Trans* **30**, 1162-1166.

Chamorro, M., Parkin, N. & Varmus, H. E. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc Natl Acad Sci U S A* **89**, 713-717.

Chang, R. Y. & Brian, D. A. (1996). cis Requirement for N-specific protein sequence in bovine coronavirus defective interfering RNA replication. *J Virol* **70**, 2201-2207.

Chang, R. Y., Hofmann, M. A., Sethna, P. B. & Brian, D. A. (1994). A cis-acting function for the coronavirus leader in defective interfering RNA replication. *J Virol* **68**, 8223-8231.

Chang, R. Y., Krishnan, R. & Brian, D. A. (1996). The UCUAAAC promoter motif is not required for high-frequency leader recombination in bovine coronavirus defective interfering RNA. *J Virol* **70**, 2720-2729.

Choi, K. S., Huang, P. & Lai, M. M. (2002a). Polypyrimidine-tract-binding protein affects transcription but not translation of mouse hepatitis virus RNA. *Virology* **303**, 58-68.

Choi, Y. G., Dreher, T. W. & Rao, A. L. (2002b). tRNA elements mediate the assembly of an icosahedral RNA virus. *Proc Natl Acad Sci U S A* **99**, 655-660.

Choi, Y. G. & Rao, A. L. (2003). Packaging of brome mosaic virus RNA3 is mediated through a bipartite signal. *J Virol* **77**, 9750-9757.

Christiansen, J., J. Egebjerg, N. Larsen & R. A. Garrett (2003). In *Ribosomes and Protein Synthesis: A Practical Approach*, Edited by G. Spedding . Oxford: IRL Press.

Coley, S. E., Lavi, E., Sawicki, S. G., Fu, L., Schelle, B., Karl, N., Siddell, S. G. & Thiel, V. (2005). Recombinant mouse hepatitis virus strain A59 from cloned, full-length cDNA replicates to high titers in vitro and is fully pathogenic in vivo. *J Virol* **79**, 3097-3106.

Cologna, R. & Hogue, B. G. (2000). Identification of a bovine coronavirus packaging signal. *J Virol* **74**, 580-583.

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

Compton, S. R., Rogers, D. B., Holmes, K. V., Fertsch, D., Remenick, J. & McGowan, J. J. (1987). In vitro replication of mouse hepatitis virus strain A59. *J Virol* **61**, 1814-1820.

Cook, J. (2002). In *Poultry diseases*, 5th edn, pp. 298-306. Edited by F. Jordan, M. Pattison, D. Alezander & T. Faragher. London: WB Saunders.

Cowley, J. A., Dimmock, C. M., Spann, K. M. & Walker, P. J. (2000). Gill-associated virus of Penaeus monodon prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. *J Gen Virol* **81**, 1473-1484.

Cowley, J. A., Dimmock, C. M. & Walker, P. J. (2002). Gill-associated nidovirus of Penaeus monodon prawns transcribes 3'-coterminal subgenomic mRNAs that do not possess 5'-leader sequences. *J Gen Virol* **83**, 927-935.

Crooke, S. T. (2004). Antisense strategies. *Curr Mol Med* **4**, 465-487.

Curtis, K. M., Yount, B., Sims, A. C. & Baric, R. S. (2004). Reverse genetic analysis of the transcription regulatory sequence of the coronavirus transmissible gastroenteritis virus. *J Virol* **78**, 6061-6066.

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

Damgaard, C. K., Tange, T. O. & Kjems, J. (2002). hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA* **8**, 1401-1415.

Darnell, J. E., Jr. (1985). RNA. *Sci Am* **253**, 68-78.

de Groot, R. J., van der Most, R. G. & Spaan, W. J. M. (1992). The fitness of defective interfering murine coronavirus DI-a and its derivatives is decreased by nonsense and frameshift mutations. *J Virol* **66**, 5898-5905.

de Vries, A. A., Chirnside, E. D., Bredenbeek, P. J., Gravestein, L. A., Horzinek, M. C. & Spaan, W. J. M. (1990). All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. *Nucleic Acids Res* **18**, 3241-3247.

de Vries, A. A., Chirnside, E. D., Horzinek, M. C. & Rottier, P. J. (1992). Structural proteins of equine arteritis virus. *J Virol* **66**, 6294-6303.

de Vries, A. A., Horzinek, M. C., Rottier, P. J. & de Groot, R. J. (1997). The genome organization of the Nidovirales: Similarities and differences between arteri-, toro-, and coronaviruses. *Semin Virol* **8**, 33-47.

de Vries, A. A., Glaser, A. L., Raamsman, M. J. & Rottier, P. J. (2001). Recombinant equine arteritis virus as an expression vector. *Virology* **284**, 259-276.

Deas, T. S., Binduga-Gajewska, I., Tilgner, M., Ren, P., Stein, D. A., Moulton, H. M., Iversen, P. L., Kauffman, E. B., Kramer, L. D. & Shi, P. Y. (2005). Inhibition of flavivirus infections by antisense oligomers specifically suppressing viral translation and RNA replication. *J Virol* **79**, 4599-4609.

den Boon, J. A. 1996. Ph.D. thesis. Leiden University. Equine arteritis virus. Replication and transcription of a coronavirus-like virus.

den Boon, J. A., Kleijnen, M. F., Spaan, W. J. M. & Snijder, E. J. (1996). Equine arteritis virus subgenomic mRNA synthesis: analysis of leader- body junctions and replicative-form RNAs. *J Virol* **70**, 4291-4298.

den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A., Horzinek, M. C. & Spaan, W. J. M. (1991). Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J Virol* **65**, 2910-2920.

DeRose, V. J. (2002). Two decades of RNA catalysis. *Chem Biol* **9**, 961-969.

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 U S A* **97**, 3913-3918.

Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding--a perspective from simple exact models. *Protein Sci* **4**, 561-602.

Dinman, J. D., Icho, T. & Wickner, R. B. (1991). A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc Natl Acad Sci U S A* **88**, 174-178.

Dirac, A. M., Huthoff, H., Kjems, J. & Berkhout, B. (2002). Regulated HIV-2 RNA dimerization by means of alternative RNA conformations. *Nucleic Acids Res* **30**, 2647-2655.

Drosten, C., Gunther, S., Preiser, W., van der, W. S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz, H. & Doerr, H. W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **348**, 1967-1976.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res* **15**, 9109-9128.

Eleouet, J. F., Rasschaert, D., Lambert, P., Levy, L., Vende, P. & Laude, H. (1995). Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* **206**, 817-822.

Escors, D., Izeta, A., Capiscol, C. & Enjuanes, L. (2003). Transmissible gastroenteritis coronavirus packaging signal is located at the 5 ' end of the virus genome. *Journal of Virology* **77**, 7890-7902.

Felsenfeld, G. (1985). DNA. *Sci Am* **253**, 58-67.

Fischer, F., Stegen, C. F., Koetzner, C. A. & Masters, P. S. (1997). Analysis of a recombinant mouse hepatitis virus expressing a foreign gene reveals a novel aspect of coronavirus transcription. *J Virol* **71**, 5148-5160.

Fosmire, J. A., Hwang, K. & Makino, S. (1992). Identification and characterization of a coronavirus packaging signal. *J Virol* **66**, 3522-3530.

Fouchier, R. A., Hartwig, N. G., Bestebroer, T. M., Niemeyer, B., De Jong, J. C., Simon, J. H. & Osterhaus, A. D. (2004). A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci U S A* .

Frolov, I., Hardy, R. & Rice, C. M. (2001). Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA* **7**, 1638-1651.

Fuerst, T. R. & Moss, B. (1989). Structure and stability of mRNA synthesized by vaccinia virus-encoded bacteriophage T7 RNA polymerase in mammalian cells. Importance of the 5' untranslated leader. *J Mol Biol* **206**, 333-348.

Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. (1986). Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci U S A* **83**, 8122-8126.

Furuya, T. & Lai, M. M. (1993). Three different cellular proteins bind to complementary sites on the 5'-end-positive and 3'-end-negative strands of mouse hepatitis virus RNA. *J Virol* **67**, 7215-7222.

Gallie, D. R. (1998). A tale of two termini: a functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation. *Gene* **216**, 1-11.

Giedroc, D. P., Theimer, C. A. & Nixon, P. L. (2000). Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* **298**, 167-185.

Godeny, E. K., Chen, L., Kumar, S. N., Methven, S. L., Koonin, E. V. & Brinton, M. A. (1993). Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV). *Virology* **194**, 585-596.

Godeny, E. K., de Vries, A. A., Wang, X. C., Smith, S. L. & de Groot, R. J. (1998). Identification of the leader-body junctions for the viral subgenomic mRNAs and organization of the simian hemorrhagic fever virus genome: evidence for gene duplication during arterivirus evolution. *J Virol* **72**, 862-867.

Goebel, S. J., Hsue, B., Dombrowski, T. F. & Masters, P. S. (2004a). Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. *J Virol* **78**, 669-682.

Goebel, S. J., Taylor, J. & Masters, P. S. (2004b). The 3' cis-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. *J Virol* **78**, 7846-7851.

González, J. M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A. E. & Enjuanes, L. (2003). A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Archives of Virology* **148**, 2207-2235.

Gora-Sochacka, A. (2004). Viroids: unusual small pathogenic RNAs. *Acta Biochim Pol* **51**, 587-607.

Gorbalenya, A. E., Snijder, E. J. & Spaan, W. J. (2004). Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J Virol* **78**, 7863-7866.

Gosline, J., Lillie, M., Carrington, E., Guerette, P., Ortlepp, C. & Savage, K. (2002). Elastic proteins: biological roles and mechanical properties. *Philos Trans R Soc Lond B Biol Sci* **357**, 121-132.

Gowda, S., Satyanarayana, T., Ayllon, M. A., Moreno, P., Flores, R. & Dawson, W. O. (2003). The conserved structures of the 5' nontranslated region of Citrus tristeza virus are involved in replication and virion assembly. *Virology* **317**, 50-64.

Gultyaev, A. P., van Batenburg, F. H. & Pleij, C. W. (1995). The computer simulation of RNA folding pathways using a genetic algorithm. *J Mol Biol* **250**, 37-51.

Guo, J., Henderson, L. E., Bess, J., Kane, B. & Levin, J. G. (1997). Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J Virol* **71**, 5178-5188.

Guo, J., Wu, T., Anderson, J., Kane, B. F., Johnson, D. G., Gorelick, R. J., Henderson, L. E. & Levin, J. G. (2000). Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J Virol* **74**, 8980-8988.

Haasnoot, P. C., Brederode, F. T., Olsthoorn, R. C. & Bol, J. F. (2000). A conserved hairpin structure in Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. *RNA* **6**, 708-716.

Hayashi, T., Shioda, T., Iwakura, Y. & Shibuta, H. (1992). RNA packaging signal of human immunodeficiency virus type 1. *Virology* **188**, 590-599.

Heasman, J. (2002). Morpholino oligos: making sense of antisense? *Dev Biol* **243**, 209-214.

Hellen, C. U. T. & Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev* **15**, 1593-1612.

Herold, J. & Andino, R. (2001). Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol Cell* **7**, 581-591.

Herold, J., Raabe, T. & Siddell, S. (1993). Molecular analysis of the human coronavirus (strain 229E) genome. *Arch Virol Suppl* **7**, 63-74.

Herold, J. & Siddell, S. G. (1993). An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res* **21**, 5838-5842.

Hertzig, T., Scandella, E., Schelle, B., Ziebuhr, J., Siddell, S. G., Ludewig, B. & Thiel, V. (2004). Rapid identification of coronavirus replicase inhibitors using a selectable replicon RNA. *J Gen Virol* **85**, 1717-1725.

Hofmann, M. A., Chang, R. Y., Ku, S. & Brian, D. A. (1993). Leader-mRNA junction sequences are unique for each subgenomic mRNA species in the bovine coronavirus and remain so throughout persistent infection. *Virology* **196**, 163-171.

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

Holmlund, J. T. (2003). Applying antisense technology: Affinitak and other antisense oligonucleotides in clinical development. *Ann N Y Acad Sci* **1002**, 244-251.

Hsue, B., Hartshorne, T. & Masters, P. S. (2000). Characterization of an essential RNA secondary structure in the 3' untranslated region of the murine coronavirus genome. *J Virol* **74**, 6911-6921.

Hsue, B. & Masters, P. S. (1997). A bulged stem-loop structure in the 3' untranslated region of the genome of the coronavirus mouse hepatitis virus is essential for replication. *J Virol* **71**, 7567-7578.

Hsue, B. & Masters, P. S. (1999). Insertion of a new transcriptional unit into the genome of mouse hepatitis virus. *J Virol* **73**, 6128-6135.

Huang, P. & Lai, M. M. (1999). Polypyrimidine tract-binding protein binds to the complementary strand of the mouse hepatitis virus 3' untranslated region, thereby altering RNA conformation. *J Virol* **73**, 9110-9116.

Huang, P. & Lai, M. M. (2001). Heterogeneous nuclear ribonucleoprotein a1 binds to the 3'-untranslated region and mediates potential 5'-3'-end cross talks of mouse hepatitis virus RNA. *J Virol* **75**, 5009-5017.

Hwang, Y. K. & Brinton, M. A. (1998). A 68-nucleotide sequence within the 3' noncoding region of simian hemorrhagic fever virus negative-strand RNA binds to four MA104 cell proteins. *J Virol* **72**, 4341-4351.

Izeta, A., Smerdou, C., Alonso, S., Penzes, Z., Mendez, A., Plana-Duran, J. & Enjuanes, L. (1999). Replication and packaging of transmissible gastroenteritis coronavirus-derived synthetic minigenomes. *J Virol* **73**, 1535-1545.

Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55**, 447-458.

Jacks, T. & Varmus, H. E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* **230**, 1237-1242.

Jackson, R. J. & Kaminski, A. (1995). Internal initiation of translation in eukaryotes: The picornavirus paradigm and beyond. *RNA* **1**, 985-1000.

Jang, S. K., Davies, M. V., Kaufman, R. J. & Wimmer, E. (1989). Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA in vivo. *J Virol* **63**, 1651-1660.

Jendrach, M., Thiel, V. & Siddell, S. (1999). Characterization of an internal ribosome entry site within mRNA 5 of murine hepatitis virus. *Arch Virol* **144**, 921-933.

Jeong, Y. S. & Makino, S. (1992). Mechanism of Coronavirus Transcription - Duration of Primary Transcription Initiation Activity and Effects of Subgenomic Rna-Transcription on Rna Replication. *Journal of Virology* **66**, 3339-3346.

Jeong, Y. S., Repass, J. F., Kim, Y. N., Hwang, S. M. & Makino, S. (1996). Coronavirus transcription mediated by sequences flanking the transcription consensus sequence. *Virology* **217**, 311-322.

Jonassen, C. M., Jonassen, T. O. & Grinde, B. (1998). A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol* **79 (Pt 4)**, 715-718.

Jonassen, C. M., Kofstad, T., Larsen, I. L., Lovland, A., Handeland, K., Follestad, A. & Lillehaug, A. (2005). Molecular identification and characterization of novel coronaviruses infecting graylag geese (Anser anser), feral pigeons (Columbia livia) and mallards (Anas platyrhynchos). *J Gen Virol* **86**, 1597-1607.

Joo, M. & Makino, S. (1992). Mutagenic analysis of the coronavirus intergenic consensus sequence. *J Virol* **66**, 6330-6337.

Joo, M. & Makino, S. (1995). The effect of two closely inserted transcription consensus sequences on coronavirus transcription. *J Virol* **69**, 272-280.

Jubin, R., Vantuno, N. E., Kieft, J. S., Murray, M. G., Doudna, J. A., Lau, J. Y. & Baroudy, B. M. (2000). Hepatitis C virus internal ribosome entry site (IRES) stem loop IIId contains a phylogenetically conserved GGG triplet essential for translation and IRES folding. *J Virol* **74**, 10430-10437.

Ke, N., Gao, X., Keeney, J. B., Boeke, J. D. & Voytas, D. F. (1999). The yeast retrotransposon Ty5 uses the anticodon stem-loop of the initiator methionine tRNA as a primer for reverse transcription. *RNA* **5**, 929-938.

Kheyar, A., St Laurent, G., Diouri, M., Dufresne, J. & Archambault, D. (1998). Sequence determination and genetic analysis of the leader region of various equine arteritis virus isolates. *Adv Exp Med Biol* **440**, 805-812.

Kim, K. H. & Hemenway, C. L. (1999). Long-distance RNA-RNA interactions and conserved sequence elements affect potato virus X plus-strand RNA accumulation. *RNA* **5**, 636-645.

Kim, Y. N., Jeong, Y. S. & Makino, S. (1993). Analysis of cis-acting sequences essential for coronavirus defective interfering RNA replication. *Virology* **197**, 53-63.

Kim, Y. N. & Makino, S. (1995). Characterization of a murine coronavirus defective interfering RNA internal cis-acting replication signal. *J Virol* **69**, 4963-4971.

Kinney, R. M., Huang, C. Y., Rose, B. C., Kroeker, A. D., Dreher, T. W., Iversen, P. L. & Stein, D. A. (2005). Inhibition of dengue virus serotypes 1 to 4 in vero cell cultures with morpholino oligomers. *J Virol* **79**, 5116-5128.

Kirkegaard, K. & Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. *Cell* **47**, 433-443.

Kocherhans, R., Bridgen, A., Ackermann, M. & Tobler, K. (2001). Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* **23**, 137-144.

Koev, G., Mohan, B. R. & Miller, W. A. (1999). Primary and secondary structural elements required for synthesis of barley yellow dwarf virus subgenomic RNA1. *J Virol* **73**, 2876-2885.

Koning, R., van den, W. S., Plaisier, J. R., van Duin, J., Pieter, A. J. & Koerten, H. (2003). Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *J Mol Biol* **332**, 415-422.

Kontos, H., Napthine, S. & Brierley, I. (2001). Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol* **21**, 8657-8670.

Kozak, M. (1989). The scanning model for translation - an update. *J Cell Biol* **108**, 229-241.

Kozak, M. (1991). Structural features in eukaryotic messenger-RNAs that modulate the initiation of translation. *J Biol Chem* **266**, 19867-19870.

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**, 1-34.
Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-157.
Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J. & Anderson, L. J. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**, 1953-1966.
Lai, M. M. (1998). Cellular factors in the transcription and replication of viral RNA genomes: a parallel to DNA-dependent RNA transcription. *Virology* **244**, 1-12.
Lai, M. M., Baric, R. S., Brayton, P. R. & Stohlman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc Natl Acad Sci U S A* **81**, 3626-3630.
Lai, M. M. & Cavanagh, D. (1997). The molecular biology of coronaviruses. *Adv Virus Res* **48**, 1-100.
Lai, M. M., Patton, C. D. & Stohlman, S. A. (1982a). Further characterization of mRNA's of mouse hepatitis virus: presence of common 5'-end nucleotides. *J Virol* **41**, 557-565.
Lai, M. M., Patton, C. D. & Stohlman, S. A. (1982b). Replication of mouse hepatitis virus: negative-stranded RNA and replicative form RNA are of genome length. *J Virol* **44**, 487-492.
Lai, M. M. & Stohlman, S. A. (1981). Comparative analysis of RNA genomes of mouse hepatitis viruses. *J Virol* **38**, 661-670.
Lai, M. M. C. & K. V. Holmes (2001). In *Fields virology*, 4th edn, pp. 1163-1185. Edited by D. M. Knipe & P. M. Howlay. Philidelphia, PA: Lippincott, Williams & Wilkins.
Latchman, D. S. (1999). Regulation of DNA virus transcription by cellular POU family transcription factors. *Rev Med Virol* **9**, 31-38.
Le, S. Y., Sonenberg, N. & Maizel, J. V., Jr. (1994). Distinct structural elements and internal entry of ribosomes in mRNA3 encoded by infectious bronchitis virus. *Virology* **198**, 405-411.
Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A. & Lai, M. M. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180**, 567-582.
Lewis, T. L. & Matsui, S. M. (1996). Astrovirus ribosomal frameshifting in an infection-transfection transient expression system. *J Virol* **70**, 2869-2875.
Li, H. P., Huang, P., Park, S. & Lai, M. M. (1999). Polypyrimidine tract-binding protein binds to the leader RNA of mouse hepatitis virus and serves as a regulator of viral transcription. *J Virol* **73**, 772-777.
Li, H. P., Zhang, X., Duncan, R., Comai, L. & Lai, M. M. (1997). Heterogeneous nuclear ribonucleoprotein A1 binds to the transcription- regulatory region of mouse hepatitis virus RNA. *Proc Natl Acad Sci U S A* **94**, 9544-9549.
Liao, C. L. & Lai, M. M. (1994). Requirement of the 5'-end genomic sequence as an upstream cis-acting element for coronavirus subgenomic mRNA transcription. *J Virol* **68**, 4727-4737.
Lin, H. X. & White, K. A. (2004). A complex network of RNA-RNA interactions controls subgenomic mRNA transcription in a tombusvirus. *EMBO J* **23**, 3365-3374.
Lin, Y. J. & Lai, M. M. (1993). Deletion mapping of a mouse hepatitis virus defective interfering RNA reveals the requirement of an internal and discontiguous sequence for replication. *J Virol* **67**, 6110-6118.
Lin, Y. J., Liao, C. L. & Lai, M. M. (1994). Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus: implications for the role of minus-strand RNA in RNA replication and transcription. *J Virol* **68**, 8131-8140.
Lin, Y. J., Zhang, X., Wu, R. C. & Lai, M. M. (1996). The 3' untranslated region of coronavirus RNA is required for subgenomic mRNA transcription from a defective interfering RNA. *J Virol* **70**, 7236-7240.
Lindenbach, B. D., Sgro, J. Y. & Ahlquist, P. (2002). Long-distance base-pairing in flock house virus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. *J Virol* **76**, 3905-3919.

Liphardt, J., Napthine, S., Kontos, H. & Brierley, I. (1999). Evidence for an RNA pseudoknot loop-helix interaction essential for efficient -1 ribosomal frameshifting. *J Mol Biol* **288**, 321-335.
Liu, D. X., Cavanagh, D., Green, P. & Inglis, S. C. (1991). A polycistronic mRNA specified by the coronavirus infectious bronchitis virus. *Virology* **184**, 531-544.
Liu, D. X. & Inglis, S. C. (1992). Internal entry of ribosomes on a tricistronic mRNA encoded by infectious bronchitis virus. *J Virol* **66**, 6143-6154.
Liu, H., Naismith, J. H. & Hay, R. T. (2003a). Adenovirus DNA replication. *Curr Top Microbiol Immunol* **272**, 131-164.
Liu, Q., Johnson, R. F. & Leibowitz, J. L. (2001). Secondary structural elements within the 3' untranslated region of mouse hepatitis virus strain JHM genomic RNA. *J Virol* **75**, 12105-12113.
Liu, Q., Yu, W. & Leibowitz, J. L. (1997). A specific host cellular protein binding element near the 3' end of mouse hepatitis virus genomic RNA. *Virology* **232**, 74-85.
Liu, Y., Sinha, S. & Owens, G. (2003b). A transforming growth factor-beta control element required for SM alpha-actin expression in vivo also partially mediates GKLF-dependent transcriptional repression. *J Biol Chem* **278**, 48004-48011.
Lopinski, J. D., Dinman, J. D. & Bruenn, J. A. (2000). Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol Cell Biol* **20**, 1095-1103.
Luytjes, W., Gerritsma, H. & Spaan, W. J. M. (1996). Replication of synthetic defective interfering RNAs derived from coronavirus mouse hepatitis virus-A59. *Virology* **216**, 174-183.
Maines, T. R., Young, M., Dinh, N. N. & Brinton, M. A. (2005). Two cellular proteins that interact with a stem loop in the simian hemorrhagic fever virus 3'(+)NCR RNA. *Virus Res* **109**, 109-124.
Makela, M. J., Puhakka, T., Ruuskanen, O., Leinonen, M., Saikku, P., Kimpimaki, M., Blomqvist, S., Hyypia, T. & Arstila, P. (1998). Viruses and bacteria in the etiology of the common cold. *J Clin Microbiol* **36**, 539-542.
Makino, S. & Joo, M. (1993). Effect of intergenic consensus sequence flanking sequences on coronavirus transcription. *J Virol* **67**, 3304-3311.
Makino, S., Joo, M. & Makino, J. K. (1991). A system for study of coronavirus mRNA synthesis: a regulated, expressed subgenomic defective interfering RNA results from intergenic site insertion. *J Virol* **65**, 6031-6041.
Makino, S. & Lai, M. M. (1989). High-frequency leader sequence switching during coronavirus defective interfering RNA replication. *J Virol* **63**, 5285-5292.
Makino, S., Shieh, C. K., Keck, J. G. & Lai, M. M. (1988a). Defective-interfering particles of murine coronavirus: mechanism of synthesis of defective viral RNAs. *Virology* **163**, 104-111.
Makino, S., Shieh, C. K., Soe, L. H., Baker, S. C. & Lai, M. M. (1988b). Primary structure and translation of a defective interfering RNA of murine coronavirus. *Virology* **166**, 550-560.
Makino, S., Stohlman, S. A. & Lai, M. M. (1986). Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. *Proc Natl Acad Sci U S A* **83**, 4204-4208.
Makino, S., Yokomori, K. & Lai, M. M. (1990). Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. *J Virol* **64**, 6045-6053.
Mandal, M. & Breaker, R. R. (2004). Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463.
Marczinke, B., Bloys, A. J., Brown, T. D., Willcocks, M. M., Carter, M. J. & Brierley, I. (1994). The human astrovirus RNA-dependent RNA polymerase coding region is expressed by ribosomal frameshifting. *J Virol* **68**, 5588-5595.
Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattra, J., Asano, J. K., Barber, S. A., Chan, S. Y., Cloutier, A., Coughlin, S. M., Freeman, D., Girn, N., Griffith, O. L., Leach, S. R., Mayo, M., McDonald, H., Montgomery, S. B., Pandoh, P. K., Petrescu, A. S., Robertson, A. G., Schein, J. E., Siddiqui, A. & Smailus, D. E. (2003). The Genome Sequence of the SARS-Associated Coronavirus. *Science* **may 2003**, 1-5.
Masters, P. S., Koetzner, C. A., Kerr, C. A. & Heo, Y. (1994). Optimization of targeted RNA recombination and mapping of a novel nucleocapsid gene mutation in the coronavirus mouse hepatitis virus. *J Virol* **68**, 328-337.

Mathews, M. B. (2002). Lost in translation. *Trends Biochem Sci* **27**, 267-269.
McCaffrey, A. P., Meuse, L., Karimi, M., Contag, C. H. & Kay, M. A. (2003). A potent and specific morpholino antisense inhibitor of hepatitis C translation in mice. *Hepatology* **38**, 503-508.
Melchers, W. J., Hoenderop, J. G., Bruins Slot, H. J., Pleij, C. W., Pilipenko, E. V., Agol, V. I. & Galama, J. M. (1997). Kissing of the two predominant hairpin loops in the coxsackie B virus 3' untranslated region is the essential structural feature of the origin of replication required for negative-strand RNA synthesis. *J Virol* **71**, 686-696.
Meselson, M. & Stahl, F. W. (1958). The replication of DNA. *Cold Spring Harb Symp Quant Biol* **23**, 9-12.
Meulenberg, J. J., Hulst, M. M., de Meijer, E. J., Moonen, P. L., den Besten, A., de Kluyver, E. P., Wensvoort, G. & Moormann, R. J. (1993a). Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. *Virology* **192**, 62-72.
Meulenberg, J. J. M., Bos-De Ruijter, J. N. A., van de Graaf, R., Wensvoort, G. & Moormann, R. J. M. (1998). Infectious transcripts from cloned genome-length cDNA of porcine reproductive and respiratory syndrome virus. *J Virol* **72**, 380-387.
Meulenberg, J. J. M., de Meijer, E. J. & Moormann, R. J. (1993b). Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence. *J Gen Virol* **74 (Pt 8)**, 1697-1701.
Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. & Ruger, W. (2003). Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table.
Miller, W. A., Dreher, T. W. & Hall, T. C. (1985). Synthesis of brome mosaic virus subgenomic RNA in vitro by internal initiation on (-)-sense genomic RNA. *Nature* **313**, 68-70.
Miller, W. A. & Koev, G. (2000). Synthesis of subgenomic RNAs by positive-strand RNA viruses. *Virology* **273**, 1-8.
Molenkamp, R., Greve, S., Spaan, W. J. M. & Snijder, E. J. (2000a). Efficient homologous RNA recombination and requirement for an open reading frame during replication of equine arteritis virus defective interfering RNAs. *J Virol* **74**, 9062-9070.
Molenkamp, R., Rozier, B. C., Greve, S., Spaan, W. J. M. & Snijder, E. J. (2000b). Isolation and characterization of an arterivirus defective interfering RNA genome. *J Virol* **74**, 3156-3165.
Molenkamp, R. & Spaan, W. J. M. (1997). Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. *Virology* **239**, 78-86.
Molenkamp, R., van Tol, H., Rozier, B. C., van der Meer, Y., Spaan, W. J. M. & Snijder, E. J. (2000c). The arterivirus replicase is the only viral protein required for genome replication and subgenomic mRNA transcription. *J Gen Virol* **81**, 2491-2496.
Moulton, H. M., Hase, M. C., Smith, K. M. & Iversen, P. L. (2003). HIV Tat peptide enhances cellular delivery of antisense morpholino oligomers. *Antisense Nucleic Acid Drug Dev* **13**, 31-43.
Moulton, H. M., Nelson, M. H., Hatlevig, S. A., Reddy, M. T. & Iversen, P. L. (2004). Cellular uptake of antisense morpholino oligomers conjugated to arginine-rich peptides. *Bioconjug Chem* **15**, 290-299.
Muto, E., Tabata, Y., Taneda, T., Aoki, Y., Muto, A., Arai, K. & Watanabe, S. (2004). Identification and characterization of Veph, a novel gene encoding a PH domain-containing protein expressed in the developing central nervous system of vertebrates. *Biochimie* **86**, 523-531.
Nagel, J. H. A. & Pleij, C. W. (2002). Self-induced structural switches in RNA. *Biochimie* **84**, 913-923.
Nagy, P. D. & Simon, A. E. (1997). New insights into the mechanisms of RNA recombination. *Virology* **235**, 1-9.
Nagy, P. D. & Simon, A. E. (1998). In vitro characterization of late steps of RNA recombination in turnip crinkle virus. I. Role of motif1-hairpin structure. *Virology* **249**, 379-392.
Nagy, P. D., Zhang, C. & Simon, A. E. (1998). Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase. *EMBO J* **17**, 2392-2403.
Nanda, S. K., Johnson, R. F., Liu, Q. & Leibowitz, J. L. (2004). Mitochondrial HSP70, HSP40, and HSP60 bind to the 3' untranslated region of the Murine hepatitis virus genome. *Arch Virol* **149**, 93-111.
Nanda, S. K. & Leibowitz, J. L. (2001). Mitochondrial aconitase binds to the 3' untranslated region of the mouse hepatitis virus genome. *J Virol* **75**, 3352-3362.

Napthine, S., Liphardt, J., Bloys, A., Routledge, S. & Brierley, I. (1999). The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *J Mol Biol* **288**, 305-320.
Narayanan, K., Chen, C. J., Maeda, J. & Makino, S. (2003). Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. *J Virol* **77**, 2922-2927.
Narayanan, K. & Makino, S. (2001). Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J Virol* **75**, 9059-9067.
Nasevicius, A. & Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat Genet* **26**, 216-220.
Neuman, B. W., Stein, D. A., Kroeker, A. D., Paulino, A. D., Moulton, H. M., Iversen, P. L. & Buchmeier, M. J. (2004). Antisense morpholino-oligomers directed against the 5' end of the genome inhibit coronavirus proliferation and growth. *J Virol* **78**, 5891-5899.
Nixon, K. S., Raman, S. & Brian, D. A. (2004). A predicted SARS-CoV 5' UTR-located stem-loop IV homolog of BCoV stem-loop IV appears more group I-like than group 2-like. *23rd Annual Meeting of the American Society for Virology* , 108-109.
O'Connor, J. B. & Brian, D. A. (2000). Downstream ribosomal entry for translation of coronavirus TGEV gene 3b. *Virology* **269**, 172-182.
Olsthoorn, R. C., Mertens, S., Brederode, F. T. & Bol, J. F. (1999). A conformational switch at the 3' end of a plant virus RNA regulates viral replication. *EMBO J* **18**, 4856-4864.
Ozdarendeli, A., Ku, S., Rochat, S., Williams, G. D., Senanayake, S. D. & Brian, D. A. (2001). Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J Virol* **75**, 7362-7374.
Panavas, T., Pogany, J. & Nagy, P. D. (2002). Analysis of minimal promoter sequences for plus-strand synthesis by the Cucumber necrosis virus RNA-dependent RNA polymerase. *Virology* **296**, 263-274.
Pasternak, A. O. 2003. Ph.D. thesis. Leiden University. Nidovirus transcription-regulating sequences.
Pasternak, A. O., Gultyaev, A. P., Spaan, W. J. M. & Snijder, E. J. (2000). Genetic manipulation of arterivirus alternative mRNA leader-body junction sites reveals tight regulation of structural protein expression. *J Virol* **74**, 11642-11653.
Pasternak, A. O., Spaan, W. J. & Snijder, E. J. (2004). Regulation of relative abundance of arterivirus subgenomic mRNAs. *J Virol* **78**, 8102-8113.
Pasternak, A. O., van den Born, E., Spaan, W. J. M. & Snijder, E. J. (2001). Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J* **20**, 7220-7228.
Pasternak, A. O., van den Born, E., Spaan, W. J. M. & Snijder, E. J. (2003). The stability of the duplex between sense and antisense transcription- regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis. *J Virol* **77**, 1175-1183.
Paul, N. & Joyce, G. F. (2002). Inaugural Article: a self-replicating ligase ribozyme. *Proc Natl Acad Sci U S A* **99**, 12733-12740.
Pedersen, K. W., van der, M. Y., Roos, N. & Snijder, E. J. (1999). Open reading frame 1a-encoded subunits of the arterivirus replicase induce endoplasmic reticulum-derived double-membrane vesicles which carry the viral replication complex. *J Virol* **73**, 2016-2026.
Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K. & Yuen, K. Y. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319-1325.
Pelletier, J. & Sonenberg, N. (1985). Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic messenger-RNA reduces translational efficiency. *Cell* **40**, 515-526.
Plant, E. P., Perez-Alvarado, G. C., Jacobs, J. L., Mukhopadhyay, B., Hennig, M. & Dinman, J. D. (2005). A Three-Stemmed mRNA Pseudoknot in the SARS Coronavirus Frameshift Signal. *PLoS Biol* **3**, e172.
Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**, 847-858.

Prufer, D., Tacke, E., Schmitz, J., Kull, B., Kaufmann, A. & Rohde, W. (1992). Ribosomal frameshifting in plants: a novel signal directs the -1 frameshift in the synthesis of the putative viral replicase of potato leafroll luteovirus. *EMBO J* **11**, 1111-1117.

Pyrc, K., Jebbink, M. F., Berkhout, B. & van der, H. L. (2004). Genome structure and transcriptional regulation of human coronavirus NL63. *Virol J* **1**, 7.

Raman, S., Bouma, P., Williams, G. D. & Brian, D. A. (2003). Stem-loop III in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *J Virol* **77**, 6720-6730.

Ramos, F. D., Carrasco, M., Doyle, T. & Brierley, I. (2004). Programmed -1 ribosomal frameshifting in the SARS coronavirus. *Biochem Soc Trans* **32**, 1081-1083.

Reeder, R. H. & Lang, W. H. (1997). Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. *Trends Biochem Sci* **22**, 473-477.

Repass, J. F. & Makino, S. (1998a). Importance of the positive-strand RNA secondary structure of a murine coronavirus defective interfering RNA internal replication signal in positive-strand RNA synthesis. *J Virol* **72**, 7926-7933.

Repass, J. F. & Makino, S. (1998b). The effect of deletion of a conserved 11 nucleotide sequence on mouse hepatitis virus RNA replication. *Adv Exp Med Biol* **440**, 247-252.

Reusken, C. B. E. M., Dalebout, T. J., Eerligh, P., Bredenbeek, P. J. & Spaan, W. J. M. (2003). Analysis of hepatitis C virus/classical swine fever virus chimeric 5 ' NTRs: sequences within the hepatitis C virus IRES are required for viral RNA replication. *J Gen Virol* **84**, 1761-1769.

Rich, A. & Kim, S. H. (1978). The three-dimensional structure of transfer RNA. *Sci Am* **238**, 52-62.

Robertson, M. P., Igel, H., Baertsch, R., Haussler, D., Ares, M., Jr. & Scott, W. G. (2005). The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* **3**, e5.

Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J. & McCaustland, K. (2003). Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* , 1-6.

Sagripanti, J. L., Zandomeni, R. O. & Weinmann, R. (1986). The cap structure of simian hemorrhagic fever virion RNA. *Virology* **151**, 146-150.

Saif, L. J., Redman, D. R., Brock, K. V., Kohler, E. M. & Heckert, R. A. (1988). Winter dysentery in adult dairy cattle: detection of coronavirus in the faeces. *Vet Rec* **123**, 300-301.

Sawicki, S. G. & Sawicki, D. L. (2005). Coronavirus transcription: a perspective. *Curr Top Microbiol Immunol* **287***,* 31-55.

Sawicki, D., Wang, T. & Sawicki, S. (2001). The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. *J Gen Virol* **82**, 385-396.

Sawicki, S. G. & Sawicki, D. L. (1995). Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. *Adv Exp Med Biol* **380**, 499-506.

Schaad, M. C., Stohlman, S. A., Egbert, J., Lum, K., Fu, K., Wei, T., Jr. & Baric, R. S. (1990). Genetics of mouse hepatitis virus transcription: identification of cistrons which may function in positive and negative strand RNA synthesis. *Virology* **177**, 634-645.

Schelle, B., Karl, N., Ludewig, B., Siddell, S. G. & Thiel, V. (2005). Selective replication of coronavirus genomes that express nucleocapsid protein. *J Virol* **79**, 6620-6630.

Scott, W. G. & Klug, A. (1996). Ribozymes: structure and mechanism in RNA catalysis. *Trends Biochem Sci* **21**, 220-224.

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

Shen, X. & Masters, P. S. (2001). Evaluation of the role of heterogeneous nuclear ribonucleoprotein A1 as a host factor in murine coronavirus discontinuous transcription and genome replication. *Proc Natl Acad Sci U S A* **98**, 2717-2722.

Shi, S. T., Huang, P., Li, H. P. & Lai, M. M. (2000). Heterogeneous nuclear ribonucleoprotein A1 regulates RNA synthesis of a cytoplasmic virus. *EMBO J* **19**, 4701-4711.

Shi, Y., Luo, H., Jia, J., Xiong, J., Yang, D., Huang, B. & Jin, Y. (2004). Antisense downregulation of SARS-CoV gene expression in Vero E6 cells. *J Gene Med* **7**, 97-107.

Shieh, C. K., Soe, L. H., Makino, S., Chang, M. F., Stohlman, S. A. & Lai, M. M. (1987). The 5'-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. *Virology* **156**, 321-330.

Siddell, S. G., J. Ziebuhr & E. J. Snijder (2005). In *Topley and Wilson's microbiology and microbial infections*, pp. 823-856. Edited by B. W. Mahy & V. ter Meulen. London: Hodder Arnold.

Siegel, R. W., Adkins, S. & Kao, C. C. (1997). Sequence-specific recognition of a subgenomic RNA promoter by a viral RNA polymerase. *Proc Natl Acad Sci U S A* **94**, 11238-11243.

Sit, T. L., Vaewhongs, A. A. & Lommel, S. A. (1998). RNA-mediated trans-activation of transcription from a viral RNA. *Science* **281**, 829-832.

Sivakumaran, K., Kim, C. H., Tayon, R., Jr. & Kao, C. C. (1999). RNA sequence and secondary structural determinants in a minimal viral promoter that directs replicase recognition and initiation of genomic plus-strand RNA synthesis. *J Mol Biol* **294**, 667-682.

Skinner, M. A., Ebner, D. & Siddell, S. G. (1985). Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *J Gen Virol* **66 (Pt 3)**, 581-592.

Smits, S. L., van Vliet, A. L., Segeren, K., el Azzouzi, H., van Essen, M. & de Groot, R. J. (2005). Torovirus non-discontinuous transcription: mutational analysis of a subgenomic mRNA promoter. *J Virol* **79**, 8275-8281.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., Guan, Y., Rozanov, M., Spaan, W. J. M. & Gorbalenya, A. E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of Molecular Biology* **331**, 991-1004.

Snijder, E. J., den Boon, J. A., Bredenbeek, P. J., Horzinek, M. C., Rijnbrand, R. & Spaan, W. J. (1990a). The carboxyl-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionarily related. *Nucleic Acids Res* **18**, 4535-4542.

Snijder, E. J., den Boon, J. A., Horzinek, M. C. & Spaan, W. J. M. (1991). Characterization of defective interfering RNAs of Berne virus. *J Gen Virol* **72 (Pt 7)**, 1635-1643.

Snijder, E. J., Horzinek, M. C. & Spaan, W. J. M. (1990b). A 3'-coterminal nested set of independently transcribed messenger-RNAs is generated during Berne Virus-replication. *J Virol* **64**, 331-338.

Snijder, E. J. & Meulenberg, J. J. M. (1998). The molecular biology of arteriviruses. *J Gen Virol* **79 (Pt 5)**, 961-979.

Snijder, E. J. & J. J. M. Meulenberg (2001). In *Fields virology*, 4th edn, pp. 1205-1220. Edited by D. M. Knipe & P. M. Howlay. Philadelphia, PA: Lippincott Williams & Wilkins.

Snijder, E. J., S. G. Siddell & A. E. Gorbalenya (2005). In *Topley and Wilson's microbiology and microbial infections*, pp. 390-404. Edited by B. W. Mahy & V. ter Meulen. London: Hodder Arnold.

Snijder, E. J., van Tol, H., Pedersen, K. W., Raamsman, M. J. & de Vries, A. A. (1999). Identification of a novel structural protein of arteriviruses. *J Virol* **73**, 6335-6345.

Snijder, E. J., Wassenaar, A. L. M. & Spaan, W. J. M. (1994). Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J Virol* **68**, 5755-5764.

Sola, I., Moreno, J. L., Zuniga, S., Alonso, S. & Enjuanes, L. (2005). Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J Virol* **79**, 2506-2516.

Somogyi, P., Jenner, A. J., Brierley, I. & Inglis, S. C. (1993). Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol* **13**, 6931-6940.

Spaan, W. J. M., D. Cavanagh, R. J. de Groot, L. Enjuanes, A. E. Gorbalenya, E. J. Snijder & P. J. Walker (2004). In *Virus Taxonomy, VIIIth Report of the ICTV*, pp. 935-943. Edited by C. M. Fauqet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier/Academic Press.

Spaan, W. J. M., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. & Siddell, S. G. (1983a). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J* **2**, 1839-1844.

Spaan, W. J. M., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. & Siddell, S. G. (1983b). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J* **2**, 1839-1844.

Spaan, W. J. M., Rottier, P. J., Horzinek, M. C. & van der Zeijst, B. A. (1981). Isolation and identification of virus-specific mRNAs in cells infected with mouse hepatitis virus (MHV-A59). *Virology* **108**, 424-434.

Spagnolo, J. F. & Hogue, B. G. (2000). Host protein interactions with the 3 ' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. *Journal of Virology* **74**, 5053-5065.

Spann, K. M., Vickers, J. E. & Lester, R. J. G. (1995). Lymphoid organ virus of *Penaeus Monodon* from Australia. *Dis Aqua Organ* **23**, 127-134.

Stadejek, T., Bjorklund, H., Bascunana, C. R., Ciabatti, I. M., Scicluna, M. T., Amaddeo, D., McCollum, W. H., Autorino, G. L., Timoney, P. J., Paton, D. J., Klingeborn, B. & Belak, S. (1999). Genetic diversity of equine arteritis virus. *J Gen Virol* **80 (Pt 3)**, 691-699.

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H. D. & Rappuoli, R. (2003). SARS--beginning to understand a new virus. *Nat Rev Microbiol* **1**, 209-218.

Steitz, T. A. & Moore, P. B. (2003). RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* **28**, 411-418.

Stirrups, K., Shaw, K., Evans, S., Dalton, K., Cavanagh, D. & Britton, P. (2000). Leader switching occurs during the rescue of defective RNAs by heterologous strains of the coronavirus infectious bronchitis virus. *J Gen Virol* **81**, 791-801.

Summerton, J. (1999). Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim Biophys Acta* **1489**, 141-158.

Summerton, J. & Weller, D. (1997). Morpholino antisense oligomers: design, preparation, and properties. *Antisense Nucleic Acid Drug Dev* **7**, 187-195.

Tan, C., Chang, L., Shen, S., Liu, D. X. & Kwang, J. (2001). Comparison of the 5' leader sequences of North American isolates of reference and field strains of porcine reproductive and respiratory syndrome virus (PRRSV). *Virus Genes* **22**, 209-217.

ten Dam, E. B., Pleij, C. W. & Bosch, L. (1990). RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes* **4**, 121-136.

Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001a). Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J Gen Virol* **82**, 1273-1281.

Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001b). Viral replicase gene products suffice for coronavirus discontinuous transcription. *Journal of Virology* **75**, 6676-6681.

Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E. & Ziebuhr, J. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol* **84**, 2305-2315.

Thiel, V. & Siddell, S. G. (1994). Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *J Gen Virol* **75 (Pt 11)**, 3041-3046.

Tijms, M. A., van Dinten, L. C., Gorbalenya, A. E. & Snijder, E. J. (2001). A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus. *Proc Natl Acad Sci U S A* **98**, 1889-1894.

Tortorici, M. A., Broering, T. J., Nibert, M. L. & Patton, J. T. (2003). Template recognition and formation of initiation complexes by the replicase of a segmented double-stranded RNA virus. *J Biol Chem* **278**, 32673-32682.

Tzeng, T. H., Tu, C. L. & Bruenn, J. A. (1992). Ribosomal frameshifting requires a pseudoknot in the Saccharomyces cerevisiae double-stranded RNA virus. *J Virol* **66**, 999-1006.

Vagner, S., Galy, B. & Pyronnet, S. (2001). Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. *EMBO Rep* **2**, 893-898.

van Berlo, M. F., Horzinek, M. C. & van der Zeijst, B. A. (1982). Equine arteritis virus-infected cells contain six polyadenylated virus-specific RNAs. *Virology* **118**, 345-352.
van den Born, E., Gultyaev, A. P. & Snijder, E. J. (2004). Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome. *RNA* **10**, 424-437.
van den Born, E., Posthuma, C. C., Gultyaev, A. P. & Snijder, E. J. (2005). Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. *J Virol* **79**, 6312-6324.
van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J., Wolthers, K. C., Wertheim-van Dillen, P. M., Kaandorp, J., Spaargaren, J. & Berkhout, B. (2004). Identification of a new human coronavirus. *Nat Med* **10**, 368-373.
van der Meer, Y., van Tol, H., Locker, J. K. & Snijder, E. J. (1998). ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. *J Virol* **72**, 6689-6698.
van der Most, R. G., Bredenbeek, P. J. & Spaan, W. J. (1991). A domain at the 3' end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. *J Virol* **65**, 3219-3226.
van der Most, R. G., de Groot, R. J. & Spaan, W. J. M. (1994). Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus: a study of coronavirus transcription initiation. *J Virol* **68**, 3656-3666.
van der Most, R. G., Heijnen, L., Spaan, W. J. & de Groot, R. J. (1992). Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. *Nucleic Acids Res* **20**, 3375-3381.
van der Most, R. G., Luytjes, W., Rutjes, S. & Spaan, W. J. (1995). Translation but not the encoded sequence is essential for the efficient propagation of the defective interfering RNAs of the coronavirus mouse hepatitis virus. *J Virol* **69**, 3744-3751.
van der Most, R. G. & W. J. M. Spaan (1995). In *The Coronaviridae*, pp. 11-31. Edited by S. G. Siddell . New York: Plenum Press.
van der Velden, A. W., van Nierop, K., Voorma, H. O. & Thomas, A. A. M. (2002). Ribosomal scanning on the highly structured insulin-like growth factor II-leader 1. *Int J Biochem Cell Biol* **34**, 286-297.
van Dinten, L. C., den Boon, J. A., Wassenaar, A. L. M., Spaan, W. J. M. & Snijder, E. J. (1997). An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc Natl Acad Sci U S A* **94**, 991-996.
van Dinten, L. C., Rensen, S., Gorbalenya, A. E. & Snijder, E. J. (1999). Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and is essential for virus replication. *J Virol* **73**, 2027-2037.
van Marle, G., Dobbe, J. C., Gultyaev, A. P., Luytjes, W., Spaan, W. J. M. & Snijder, E. J. (1999a). Arterivirus discontinuous mRNA transcription is guided by base-pairing between sense and antisense transcription-regulating sequences. *Proc Natl Acad Sci U S A* **96**, 12056-12061.
van Marle, G., Luytjes, W., van der Most, R. G., Van der straaten, T. & Spaan, W. J. M. (1995). Regulation of Coronavirus Messenger-Rna Transcription. *Journal of Virology* **69**, 7851-7856.
van Marle, G., van Dinten, L. C., Spaan, W. J. M., Luytjes, W. & Snijder, E. J. (1999b). Characterization of an equine arteritis virus replicase mutant defective in subgenomic mRNA synthesis. *J Virol* **73**, 5274-5281.
van Meerten, D., Girard, G. & van Duin, J. (2001). Translational control by delayed RNA folding: identification of the kinetic trap. *RNA* **7**, 483-494.
van Vliet, A. L., Smits, S. L., Rottier, P. J. & de Groot, R. J. (2002). Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J* **21**, 6571-6580.
Varani, G. & A. Pardi (1994). In *RNA-protein interactions*, pp. 1-24. Edited by K. Nagai & I. W. Mattaj. Oxford: Oxford University Press.
Verheije, M. H., Olsthoorn, R. C. L., Kroese, M. V., Rottier, P. J. M. & Meulenberg, J. J. M. (2002). Kissing interaction between 3 ' noncoding and coding sequences is essential for porcine arterivirus RNA replication. *Journal of Virology* **76**, 1521-1526.

Wagner, H. M., Balasuriya, U. B. R. & James, M. N. (2003). The serologic response of horses to equine arteritis virus as determined by competitive enzyme-linked immunosorbent assays (c-ELISAs) to structural and non-structural viral proteins. *Comp Immunol Microbiol Infect Dis* **26**, 251-260.

Walker, P. J., J. R. Bonami, V. Boonsaeng, P. S. Chang, J. A. Cowley, L. Enjuanes, T. W. Flegel, D. V. Lightner, P. C. Loh, E. J. Snijder & K. Tang (2004). In *Virus Taxonomy, VIIIth Report of the ICTV*, pp. 973-977. Edited by C. M. Fauqet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier/Academic Press.

Wang, J., Carpenter, C. D. & Simon, A. E. (1999). Minimal sequence and structural requirements of a subgenomic RNA promoter for turnip crinkle virus. *Virology* **253**, 327-336.

Wang, Y. & Zhang, X. (2000). The leader RNA of coronavirus mouse hepatitis virus contains an enhancer-like element for subgenomic mRNA transcription. *J Virol* **74**, 10571-10580.

Wang, Z., Ren, L., Zhao, X., Hung, T., Meng, A., Wang, J. & Chen, Y. G. (2004). Inhibition of severe acute respiratory syndrome virus replication by small interfering RNAs in mammalian cells. *J Virol* **78**, 7523-7527.

Watson, J. D. & Crick, F. H. (1953a). Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964-967.

Watson, J. D. & Crick, F. H. (1953b). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.

Weiss, M., Steck, F. & Horzinek, M. C. (1983). Purification and partial characterization of a new enveloped RNA virus (Berne virus). *J Gen Virol* **64 (Pt 9)**, 1849-1858.

White, K. A. (2002). The premature termination model: A possible third mechanism for subgenomic mRNA transcription in (+)-strand RNA viruses. *Virology* **304**, 147-154.

White, K. A. & Morris, T. J. (1995). RNA determinants of junction site selection in RNA virus recombinants and defective interfering RNAs. *RNA* **1**, 1029-1040.

Williams, G. D., Chang, R. Y. & Brian, D. A. (1995). Evidence for a pseudoknot in the 3' untranslated region of the bovine coronavirus genome. *Adv Exp Med Biol* **380**, 511-514.

Williams, G. D., Chang, R. Y. & Brian, D. A. (1999). A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. *J Virol* **73**, 8349-8355.

Wilson, K. S. & von Hippel, P. H. (1995). Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci U S A* **92**, 8793-8797.

Woo, K., Joo, M., Narayanan, K., Kim, K. H. & Makino, S. (1997). Murine coronavirus packaging signal confers packaging to nonviral RNA. *J Virol* **71**, 824-827.

Woo, P. C., Lau, S. K., Chu, C. M., Chan, K. H., Tsoi, H. W., Huang, Y., Wong, B. H., Poon, R. W., Cai, J. J., Luk, W. K., Poon, L. L., Wong, S. S., Guan, Y., Peiris, J. S. & Yuen, K. Y. (2005). Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J Virol* **79**, 884-895.

Wootton, S., Yoo, D. & Rogan, D. (2000). Full-length sequence of a Canadian porcine reproductive and respiratory syndrome virus (PRRSV) isolate. *Arch Virol* **145**, 2297-2323.

Worobey, M. & Holmes, E. C. (1999). Evolutionary aspects of recombination in RNA viruses. *J Gen Virol* **80 (Pt 10)**, 2535-2543.

Wu, C. J., Huang, H. W., Liu, C. Y., Hong, C. F. & Chan, Y. L. (2005). Inhibition of SARS-CoV replication by siRNA. *Antiviral Res* **65**, 45-48.

Wu, H. Y., Guy, J. S., Yoo, D., Vlasak, R., Urbach, E. & Brian, D. A. (2003). Common RNA replication signals exist among group 2 coronaviruses: evidence for in vivo recombination between animal and human coronavius molecules. *Virology* **315**, 174-183.

Yount, B., Curtis, K. M. & Baric, R. S. (2000). Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. *J Virol* **74**, 10600-10611.

Yount, B., Curtis, K. M., Fritz, E. A., Hensley, L. E., Jahrling, P. B., Prentice, E., Denison, M. R., Geisbert, T. W. & Baric, R. S. (2003). Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12995-13000.

Yount, B., Denison, M. R., Weiss, S. R. & Baric, R. S. (2002). Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J Virol* **76**, 11065-11078.
Yu, W. & Leibowitz, J. L. (1995). A conserved motif at the 3' end of mouse hepatitis virus genomic RNA required for host protein binding and viral RNA replication. *Virology* **214**, 128-138.
Zhang, G., Slowinski, V. & White, K. A. (1999). Subgenomic mRNA regulation by a distal RNA element in a (+)-strand RNA virus. *RNA* **5**, 550-561.
Zhang, X. & Lai, M. M. (1994). Unusual heterogeneity of leader-mRNA fusion in a murine coronavirus: implications for the mechanism of RNA transcription and recombination. *J Virol* **68**, 6626-6633.
Zhang, X. & Lai, M. M. (1995). Interactions between the cytoplasmic proteins and the intergenic (promoter) sequence of mouse hepatitis virus RNA: correlation with the amounts of subgenomic mRNA transcribed. *J Virol* **69**, 1637-1644.
Zhang, X. & Lai, M. M. (1996). A 5'-proximal RNA sequence of murine coronavirus as a potential initiation site for genomic-length mRNA transcription. *J Virol* **70**, 705-711.
Zhang, X., Liao, C. L. & Lai, M. M. (1994). Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription both in trans and in cis. *J Virol* **68**, 4738-4746.
Zhang, X. & Liu, R. (2000). Identification of a noncanonical signal for transcription of a novel subgenomic mRNA of mouse hepatitis virus: implication for the mechanism of coronavirus RNA transcription. *Virology* **278**, 75-85.
Zhao, X., Shaw, K. & Cavanagh, D. (1993). Presence of subgenomic mRNAs in virions of coronavirus IBV. *Virology* **196**, 172-178.
Ziebuhr, J. (2004). Molecular biology of severe acute respiratory syndrome coronavirus. *Curr Opin Microbiol* **7**, 412-419.
Ziebuhr, J. (2005). The coronavirus replicase. *Curr Top Microbiol Immunol* **287**, 57-94.
Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol* **81**, 853-879.
Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415.
Zuniga, S., Sola, I., Alonso, S. & Enjuanes, L. (2004). Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J Virol* **78**, 980-994.

Summary

Viruses depend on their host cell for the production of their progeny. The genetic information that is required to regulate this process is contained in the viral genome, which is a RNA molecule in a large group of viruses. In the case of plus-stranded RNA viruses, this RNA genome is also directly involved in translation (resulting in the synthesis of viral enzymes), replication, transcription and encapsidation into progeny virions. The multifunctional nature of these viral RNA genomes requires the tight control of all these processes. For this regulation, the viral RNA genomes are equipped with RNA sequence motifs and higher order RNA structures.

Nidoviruses are enveloped, positive-strand RNA viruses that share a similar polycistronic genome organization and replication strategy. The order *Nidovirales* includes three families, *Coronaviridae* (consisting of the genera *Coronavirus* and *Torovirus*), *Arteriviridae* and *Roniviridae*. At 25-32 kilobases, corona- and roniviruses possess the largest known RNA genomes. Arterivirus genomes (12-16 kilobases) are substantially smaller. One characteristic of nidoviruses is that in infected cells they produce a nested set of at least two (but usually five to nine) 3´ co-terminal subgenomic (sg) mRNAs to express the viral structural proteins (and sometimes also accessory proteins) from genes located in the 3'-proximal third of the genome. In addition, sg mRNAs of arteri- and coronaviruses are also 5' co-terminal, because they contain a common 5' leader sequence that is identical to that of the genome. Also for nidoviruses, RNA secondary structures have been recognized as regulatory signals in the viral life cycle. The aim of the work presented here was to identify RNA elements that are involved in the regulation of arteri- and coronavirus sg mRNA synthesis.

Transcription-regulating sequences (TRSs) play a crucial role during the synthesis of arteri- and coronavirus sg mRNAs. They are located both in the 5' end of the genome (the 'leader TRS') and precede each ORF in the 3'-terminal third of the genome ('body TRSs'). To address the role of the TRS more specifically, a site-directed mutagenesis approach was applied using the full-length infectious cDNA clone of *Equine arteritis virus* (EAV) to make all possible single nucleotide (nt) mutations in its hexanucleotide leader TRS and RNA7 body TRS (Chapter 3). Data from these experiments provided additional support for the importance of the base-pairing interaction between the leader TRS and the anti-body TRS, a step that is essential for sg RNA synthesis. In addition, evidence for body TRS-specific functions was obtained.

It was hypothesized that the leader TRS is located in the loop of a long RNA hairpin structure. To confirm this and obtain information on the role of the leader, a structural and functional characterization of the 5'-proximal region of the EAV genome was performed (Chapter 4). The result of this study was a detailed RNA secondary structure model of the 5'-terminal 313 nt of the EAV genome. The model supported the presence of a 'leader TRS hairpin' (LTH), which carries the leader TRS in its loop. In the same study, similar LTH structures were predicted in the 5'-proximal region of all arterivirus and most coronavirus genomes, including that of *Severe acute respiratory syndrome coronavirus* (SARS-CoV). In a systematic mutagenesis study (Chapter 5), the

LTH and its immediate flanking sequences were found to be essential for efficient sg mRNA synthesis. Since, the sequences that flank the LTH on either side are able to base-pair with each other and thus elongate the LTH stem, the extended LTH conformation was postulated to represent the structure that guides discontinuous sg RNA synthesis.

The RNA secondary structure model of the 5'-terminal 313 nt of the EAV genome and our current knowledge of EAV sg mRNA synthesis, in part based on the data presented in Chapters 3, 4, and 5, formed the basis for an EAV inhibitor study (Chapter 6) in which regulatory RNA sequences were targeted in an attempt to block virus replication in cell culture. The antiviral potential of antisense peptide-conjugated phosphorodiamidate morpholino oligomers (P-PMOs), targeting specific EAV RNA sequences, was evaluated. These uncharged single stranded DNA mimics were designed to interfere with several aspects of virus amplification. The 5’ untranslated region (5’ UTR) of the EAV genome represented the most effective P-PMO target, showing complete inhibition of virus amplification at P-PMO concentrations in the low micromolar range.

The theoretical chapters in this thesis introduce the reader to the role of higher order RNA structures in biological processes in general and in the life cycle of RNA viruses in particular (Chapter 1). An overview of the current knowledge of the regulatory role of RNA elements in the life cycle of nidoviruses is presented in Chapter 2. A general discussion of the findings described in the experimental chapters is presented in Chapter 7.

Samenvatting

Voor hun vermenigvuldiging zijn virussen afhankelijk van hun gastheercel. De genetische informatie die nodig is om dit proces te reguleren bevindt zich in het virale genoom, dat bij een grote groep virussen een RNA-molecuul is. In het geval van positiefstrengige RNA-virussen is dit RNA-genoom direct betrokken bij translatie (resulterend in de synthese van virale enzymen), replicatie, transcriptie en wordt het uiteraard ingepakt in nieuwe virusdeeltjes. Het multifunctionele karakter van deze virale RNA-genomen vereist een strikte regulatie van deze processen. Hiervoor zijn de virale RNA-genomen uitgerust met RNA-sequenties en structuren.

Nidovirussen zijn positiefstrengige RNA-virussen die eenzelfde polycistronische genoomorganisatie en replicatiestrategie hebben. Het virusdeeltje is omgeven door een envelop. De orde Nidovirales verenigt drie families, Coronaviridae (samengesteld uit de genera Coronavirus en Torovirus), Arteriviridae en Roniviridae. Met een lengte van 25 tot 32 kilobasen hebben corona- en ronivirussen, van de nu bekende virussen, de grootste RNA-genomen. Het genoom van arterivirussen (12 tot 16 kilobasen) is beduidend kleiner. Een specifieke eigenschap van nidovirussen is dat ze in geïnfecteerde cellen een overlappende set van tenminste twee (maar meestal vijf tot negen) subgenome (sg) mRNA's produceren die gebruikt worden voor de expressie van de virale structurele eiwitten (en soms ook additionele eiwitten), gecodeerd door genen die liggen in het 3'-gelegen deel van het genoom. Daarbij hebben de sg mRNA's van arteri- and coronavirussen ook hetzelfde 5' uiteinde, de zogenaamde 'leadersequentie', die identiek is aan het 5' uiteinde van het genoom. Ook bij nidovirussen zijn primaire en secundaire RNA-structuren betrokken bij de regulatie van de virale levenscyclus. Het doel van het in dit proefschrift beschreven onderzoek was het identificeren van dit soort RNA-elementen die betrokken zijn bij de regulatie van de sg mRNA synthese van arteri- and coronavirussen.

Transcriptie-regulerende sequenties (TRS's) spelen een cruciale rol in de synthese van arteri- and coronavirus sg mRNA's. TRS's zijn zowel gelokaliseerd in het 5' uiteinde van het genoom (de 'leader TRS') als ook voorafgaand aan elk gen in het 3'-gelegen derde deel van het genoom ('body TRS's'). Om de rol van de TRS in meer detail te bestuderen is een uitgebreide mutagenesestudie verricht waarbij gebruik is gemaakt van een zogenaamde infectieuze cDNA-kloon van het Equine arteritis virus (EAV), een plasmide dat de complete cDNA-sequentie van het virus bevat en dat kan worden gebruikt om infectieus RNA te produceren. In deze cDNA-sequentie werd elke nucleotide (nt) van de hexanucleotide leader TRS en RNA7 body TRS systematisch veranderd in de drie andere mogelijke nucleotiden (Hoofdstuk 3). De resultaten van deze experimenten ondersteunden het belang van de basenparing interactie tussen de leader TRS en de anti-body TRS, een stap die essentieel is voor sg RNA synthese. Ook werd bewijs verkregen voor body TRS-specifieke functies.

De hypothese is dat de leader TRS zich bevindt in de lus van een lange RNA haarspeldstructuur. Om dit te bevestigen, en om meer informatie te verkrijgen over de rol van de leader sequentie, werd een structurele en functionele karakterisering van het 5'-

gelegen deel van het EAV-genoom uitgevoerd (Hoofdstuk 4). Deze studie resulteerde in een gedetailleerd model van de secundaire RNA structuur van de 313- nt die het 5' uiteinde van het EAV-genoom vormen. Dit model ondersteunde de aanwezigheid en het belang van een 'leader TRS haarspeld' (LTH), die de leader TRS in haar lus heeft. In dezelfde studie, werden vergelijkbare LTH-structuren voorspeld in de 5'-gelegen delen van de genomen van alle arterivirussen en de meeste coronavirussen, inclusief het SARS-coronavirus. In een systematische mutagenese studie (Hoofdstuk 5), werden de LTH en de direct flankerende sequenties essentieel bevonden voor de efficiënte synthese van sg mRNA's. Omdat de sequenties die de LTH aan weerskanten flankeren kunnen basenparen en zo de LTH-stam kunnen verlengen, werd gepostuleerd dat deze verlengde LTH-conformatie de structuur is die betrokken is bij discontinue sg RNA synthese.

Het RNA-structuurmodel van het 5' uiteinde van het EAV-genoom en onze kennis van de EAV sg mRNA synthese, deels gebaseerd op de data gepresenteerd in Hoofdstukken 3, 4, en 5, vormden de basis voor een strategie om EAV-vermenigvuldiging (in celkweek) te remmen (Hoofdstuk 6). Hierbij werden regulerende RNA-sequenties gebruikt als doelwit waartegen complementaire peptide-geconjugeerde 'phosphorodiamidate morpholino oligomers' (P-PMO's) werden ontworpen. Wij evalueerden de antivirale potentie van deze ongeladen enkelstrengige DNA-analogen, gericht tegen specifieke EAV RNA-sequenties om te interfereren met bepaalde aspecten van virusreplicatie. P-PMO's bleken het meest effectief wanneer zij gericht waren tegen het niet-getransleerde deel van het 5'-uiteinde van het EAV-genoom, wat zich uitte in een volledige remming van virusreplicatie bij P-PMO concentraties in het laag micromolaire gebied.

De theoretische hoofdstukken in dit proefschrift zijn bedoeld om de lezer te introduceren in de rol van hogere orde RNA-structuren in biologische processen in het algemeen en in de levenscyclus van RNA-virussen in het bijzonder (Hoofdstuk 1). Hoofdstuk 2 presenteert een overzicht van de huidige kennis van regulerende RNA elementen in de levenscyclus van nidovirussen. De bevindingen beschreven in de experimentele hoofdstukken worden bediscussieerd in hoofdstuk 7.

Curriculum Vitae

Erwin van den Born werd geboren op 2 april 1977 te Wageningen. In 1995 behaalde hij het VWO diploma aan het Christelijk Streeklyceum te Ede. Vervolgens studeerde hij Bioprocestechnologie (specialisatie cellulair-moleculair) aan de Wageningen Universiteit, waar in januari 2001 het doctoraalexamen werd afgelegd met afstudeervakken in de richtingen moleculaire virologie en microbiologie. Hij was als Assistent in Opleiding in dienst van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek van februari 2001 tot januari 2005 en vervolgens in dienst van de Rijksuniversiteit Leiden tot oktober 2005. Onder begeleiding van dr. E.J. Snijder werd het in dit proefschrift beschreven onderzoek verricht bij de sectie Virologie van de afdeling Medische Microbiologie van het Leids Universitair Medisch Centrum. Vanaf oktober 2005 is hij werkzaam bij dezelfde afdeling waar hij als onderzoeker het onderzoek aan nidovirussen voortzet.