13

THE ARTERIVIRUS REPLICASE

The Road from RNA to Protein(s), and Back Again

Eric J. Snijder

Department of Virology
Leiden University Medical Center
AZL P4–26, Postbus 9600
2300 RC Leiden, The Netherlands
Telephone: ++ 31 715261657; fax: ++ 31 715266761; email: snijder@virology.azl.nl.

1. INTRODUCTION

About seven years ago, the identification of arteriviruses (den Boon *et al.*, 1991) and toroviruses (Snijder *et al.*, 1990) as distant relatives of "traditional" coronaviruses incited a discussion on the taxonomic position of these three virus groups (Cavanagh *et al.*, 1994). As a first result, the *genus torovirus* was included in the *Coronaviridae* family (Cavanagh *et al.*, 1993). The taxonomic debate ended at the 1996 International Congress of Virology in Jerusalem with the establishment of the *Arteriviridae* family and the order of the *Nidovirales*, containing the *Coronaviridae* and *Arteriviridae* families (Cavanagh, 1997). These re-classifications acknowledged both the many unique properties of arteriviruses and coronaviruses as well as their intriguing ancestral relationship at the level of replicase genes, genome organization, and replication strategy (Snijder and Horzinek, 1993; Snijder and Spaan, 1995; Cavanagh, 1997; de Vries *et al.*, 1997; Snijder and Meulenberg, 1998).

At the center of nidovirus molecular biology is a nonstructural gene which encodes a polyprotein of between 3175 (for the arterivirus equine arteritis virus (EAV); den Boon *et al.*, 1991) and approximately 7200 amino acids (for the coronavirus mouse hepatitis virus (MHV); Lee *et al.*, 1991; Bonilla *et al.*, 1994; Brown and Brierley, 1995). This size variation of the replicase again illustrates how both conservation and variation have played a part in nidovirus evolution. In addition to a number of highly conserved domains/functions which can be considered to form the "core" of the nidovirus replicase, each virus (or cluster of viruses) appears to have developed its own set of accessory nonstructural functions (see also below). However, the basic genome expression strategy is the same for all nidoviruses (Figure 1): from the incoming genomic RNA they generate two large replicase

Coronaviruses and Arteriviruses, edited by Enjuanes *et al.*
Plenum Press, New York, 1998

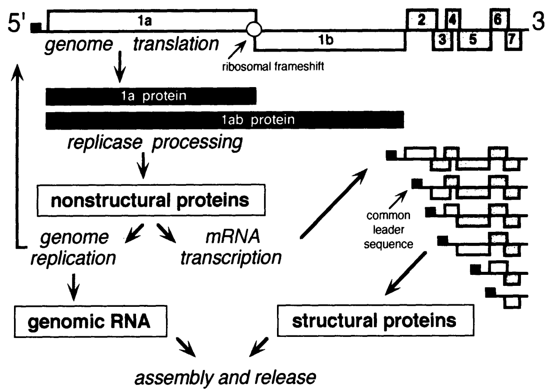

Figure 1. Overview of the nidovirus life cycle, illustrating the central role of the viral nonstructural proteins. EAV was used as an example in this figure.

polyproteins, which cleave themselves into smaller functional subunits. Subsequently, these nonstructural proteins can carry out the processes of genome replication and subgenomic mRNA transcription, the latter leading to the expression of a set of (mostly structural) genes from the 3' end of the genome. As expected from the common ancestry of the viral enzymes involved, the basic features of arterivirus mRNA transcription have recently been shown to be identical to those of coronaviruses (den Boon *et al.*, 1995b; den Boon *et al.*, 1996).

This chapter will focus on what is currently known about the structure and function of the arterivirus replicase polyprotein. Most of the data have been generated in Leiden using the arterivirus prototype EAV, but references to other arteriviruses and coronaviruses will be made when appropriate.

2. ORGANIZATION OF THE NIDOVIRUS REPLICASE

All nidovirus replicase genes consist of two open reading frames (ORFs), ORF1a and ORF1b. ORF1a is translated directly from the genomic RNA The expression of ORF1b requires a -1 ribosomal frameshift just upstream of the ORF1a termination codon, which leads to the synthesis of an ORF1ab polyprotein. In all nidovirus replicase genes two frameshift-promoting RNA signals have been identified (Jacks *et al.*, 1988; Brierley, 1995): a so-called "slippery" sequence, which is the actual frameshift site, and a downstream RNA pseudoknot structure. In EAV (den Boon *et al.*, 1991) the putative shift site 5' GUUAAAC 3' is followed immediately by the ORF1a termination codon and thus all 1727 residues of the ORF1a polyprotein are also present in the ORF1ab frameshift product. In many other nidoviruses, additional codons (up to 25) are present between frameshift site and ORF1a termination codon. This results in an ORF1a protein that contains unique C-terminal sequences which are lacking in the ORF1ab frameshift protein.

The arrangement of conserved domains (Figure 2) within the nidovirus replicase polyprotein is unique, and clearly separates the nidoviruses from other groups of plus-stranded RNA viruses. Sequence comparison between arterivirus and coronaviruses replicases revealed up to 30% amino acid sequence identity in the most conserved domains, a percentage that cannot be due to convergent evolution. The conservation of the putative RNA-dependent RNA polymerase and the putative RNA helicase, two domains which are

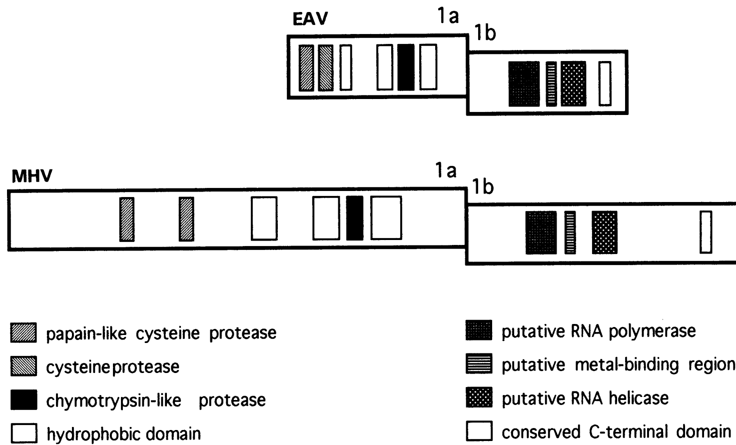

Figure 2. Comparison of the organization of the smallest (EAV) and largest (MHV) nidovirus replicase polyproteins.

common to positive-stranded RNA viruses, is not very surprising. It is remarkable, however, that only in nidovirus replicases the helicase domain is located downstream of the polymerase motif. The polymerase motif also carries another nidovirus trademark: the substitution of the canonical GDD in the core of the motif by SDD. Additional domains were identified that are conserved in nidovirus replicases (both in sequence and in position): e.g. a conserved domain in the C-terminal part of the ORF1b protein (den Boon *et al.*, 1991; Snijder *et al.*, 1990) and a Cys/His-rich domain upstream of the helicase motif. The latter was previously proposed to have metal-binding properties (Gorbalenya *et al.*, 1989b; Lee *et al.*, 1991) and was recently implicated in a remarkable mRNA transcription defect (van Dinten *et al.*, 1997).

The nidovirus replicase polyproteins are all processed by multiple ORF1a-encoded proteases. Despite the fact that our understanding of coronavirus replicase processing is far from complete, a common pattern for coronaviruses and arteriviruses is emerging (Snijder and Spaan, 1995; de Vries *et al.*, 1997; Snijder and Meulenberg, 1998). In comparable positions, the ORF1a proteins of both virus groups contain a "main protease", which is responsible for most replicase processing steps. These proteases belong to the superfamily that comprises the chymotrypsin-like and picornavirus 3C-like proteolytic enzymes (Fig. 2) (Bazan and Fletterick, 1988; Gorbalenya *et al.*, 1989a). Although the catalytic nucleophile of the coronavirus 3C-like protease (Cys; Ziebuhr *et al.*, 1995; Lu *et al.*, 1995; Liu and Brown, 1995) differs from that in the arterivirus nsp4 protease (Ser; Snijder *et al.*, 1996), these domains may still have a common ancestry (Snijder *et al.*, 1996). The exchange of Cys for Ser (or vice versa) at the active site of the enzyme is considered to be feasible (Bazan and Fletterick, 1988; Gorbalenya *et al.*, 1989a; Bazan and Fletterick, 1990; Dougherty and Semler, 1993).

The functions encoded from the central region of ORF1a to the 3'-end of ORF1b appear to form the "core" of the nidovirus replicase polyprotein: the well-conserved domains (main protease – polymerase – metal-binding domain – helicase – C-terminal ORF1b domain) are within this area, and only small insertions and deletions in this part of the replicase can be detected within the coronavirus or arterivirus groups. The presence of

hydrophobic domains on either side of the main protease domain is another property shared by arteriviruses and coronaviruses. The N-terminal half of the ORF1a protein is very variable, both in size and in sequence. A comparison between coronaviruses and arteriviruses in this region does not yield any significant similarities, and even within the coronavirus and arterivirus groups there is little conservation. However, one striking observation can be made. Both corona- and arteriviruses contain protease domains in the amino-terminal half of the ORF1a protein which have been shown, or are predicted, to belong to the papain-like cysteine protease superfamily (Fig. 2) (Gorbalenya *et al.*, 1991; Gorbalenya and Snijder, 1996). A number of these nidovirus papain-like proteases have recently been characterized experimentally (Gao *et al.*, 1996; Bonilla *et al.*, 1995; Baker *et al.*, 1993; Lee *et al.*, 1991; Snijder *et al.*, 1995; den Boon *et al.*, 1995a; Snijder *et al.*, 1992).

3. ARTERIVIRUS PROTEASES

Proteolytic processing of nonstructural proteins fulfils a key role in the life cycle of most viruses. In the course of virus evolution, highly specific virus-encoded proteases have evolved and their importance for the regulation of virus replication is becoming more and more evident (Dougherty and Semler, 1993; Gorbalenya and Snijder, 1996). For EAV, protease domains have been identified in three replicase cleavage products: nonstructural protein 1 (nsp1), nsp2, and nsp4 (Figure 3). These proteases and their corresponding cleavage sites are well-conserved in other arteriviruses (Meulenberg *et al.*, 1993; Godeny *et al.*, 1993; Snijder and Spaan, 1995; de Vries *et al.*, 1997; Snijder and Meulenberg, 1998).

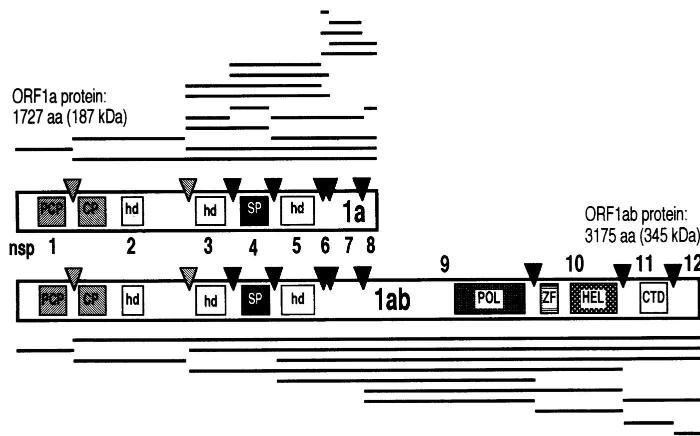

Figure 3. Proteolytic processing of the EAV replicase polyproteins. The processing schemes for the EAV ORF1a and ORF1ab proteins are depicted. The lines above and below the polyproteins represent the processing end products and intermediates which are currently known to be present in infected cells. Most of these protein species have been identified in immunoprecipitation analyses (Snijder *et al.*, 1994; van Dinten *et al.*, 1996; Wassenaar *et al.*, 1997). The three protease domains in the ORF1a polyprotein and their corresponding cleavage sites are indicated: PCP, nsp1 papainlike cysteine protease; CP, nsp2 cysteine protease; SP, nsp4 serine protease; hd, hydrophobic domain. In the ORF1b-encoded polypeptide the four major domains conserved in nidoviruses have been depicted: POL, putative RNA-dependent RNA polymerase; ZF, putative zinc finger domain; HEL, putative RNA helicase; CTD, conserved C-terminal domain specific for nidoviruses.

The internal papain-like cysteine protease (PCP) in EAV nsp1 is responsible for the production of this 29 kDa amino-terminal ORF1a cleavage product (Snijder *et al.*, 1992). Residues Cys-164 and His-230 have been proposed to form the catalytic dyad of the nsp1 PCP, which cleaves between Gly-260 and Gly-261 (Snijder *et al.*, 1992). Also the production of the next amino-terminal cleavage product, nsp2 (61 kDa), was shown to be an autoproteolytic event, which is mediated by a cysteine protease (CP) in the nsp2 N-terminal region (putative active site residues in EAV: Cys-270 and His-332; putative cleavage site: between Gly-831 and Gly-832). Although the nsp2 CP is most similar to the papain-like group of viral proteases, it possesses a number of unique properties and has been proposed to belong to a new subgroup of viral cysteine proteases (Snijder *et al.*, 1995). Two other arteriviruses, lactate dehydrogenase-elevating virus (LDV) and porcine reproductive and respiratory syndrome virus (PRRSV), have been found to use even a third N-terminal cysteine protease domain (Meulenberg *et al.*, 1993; Godeny *et al.*, 1993; den Boon *et al.*, 1995a). During arterivirus evolution, EAV appears to have lost this proteolytic activity, which is responsible for the production of an additional amino-terminal cleavage product from the nsp1 region of PRRSV and LDV (den Boon *et al.*, 1995a).

The arterivirus main protease, the nsp4 serine protease (SP), is a member of a relatively rare group of proteolytic enzymes, the 3C-like serine proteases (Snijder *et al.*, 1996). It combines the catalytic triad of classical chymotrypsinlike proteases (His-1103, Asp-1129, and Ser-1184 in EAV) with the substrate specificity of the 3C-like cysteine proteases, a subgroup of chymotrypsinlike enzymes named after the picornavirus 3C proteases. The latter property, which is assumed to be determined by a number of residues in the substrate-binding region of the nsp4 SP, explains the specificity for cleavage sites containing a Glu↓Gly/Ser dipeptide (Wassenaar *et al.*, 1997; Snijder *et al.*, 1996). Five of these cleavage sites have now been identified in the ORF1a protein: Glu-1064↓Gly-1065, Glu-1268↓Ser-1269, Glu-1430↓Gly-Gly1431, Glu-1452↓Ser-1453, and Glu-1677↓Gly-1678. The presence of three additional SP cleavage sites in the ORF1b-encoded polyprotein has been predicted (van Dinten *et al.*, 1996), but remains to be confirmed experimentally.

4. PROTEOLYTIC PROCESSING OF THE EAV REPLICASE

The post-translational fate of the EAV ORF1a (1727 aa) and ORF1ab (3175 aa) polyproteins has now been studied extensively and an apparently complete processing scheme has recently been obtained (Fig. 3) (Snijder *et al.*, 1994; van Dinten *et al.*, 1996; Wassenaar *et al.*, 1997). The ORF1ab polyprotein is cleaved ten times by the three ORF1a-encoded proteases described above. In combination with the ribosomal frameshift, this leads to the generation of 12 processing end products (named nsp 1 to 12). In the case of EAV, the C-terminal cleavage product of the ORF1a protein (nsp8) is identical to the N-terminal domain of the first ORF1b-encoded subunit (nsp9). In addition to the processing end products, a large number of intermediates has been detected, many of which have considerable half-lives and may therefore fulfil a specific role in the viral life cycle. The processing analysis of other viral replicases has revealed that such intermediates can be functional subunits themselves, e.g. in the replication of poliovirus (Ypma-Wong *et al.*, 1988; Jore *et al.*, 1988) and Sindbis virus (de Groot *et al.*, 1990; Lemm and Rice, 1993a; Lemm and Rice, 1993b; Shirako and Strauss, 1994; Lemm *et al.*, 1994).

An interesting recent observation was the fact that two alternative pathways can be followed for the processing of the C-terminal part of the ORF1a protein (Wassenaar *et al.*,

1997). After the rapid autoproteolytic release of nsp1 and nsp2 from the ORF1a polyprotein, a 96kDa nsp3–8 processing intermediate remains which contains the viral main protease, the nsp4 SP. It has now been found that nsp3–8 can cleave itself following either a "major pathway" (i.e. the pathway used most abundantly in infected cells) or following a "minor pathway". Co-expression experiments in the vaccinia virus/T7 system strongly suggested that the association with cleaved nsp2 is required for nsp3–8 to enter the major pathway. This leads to cleavage of the nsp4/5 and nsp7/8 junctions, but the nsp5/6 and nsp6/7 sites are not processed. When nsp2 is lacking the nsp4/5 site is no longer cleaved and instead the nsp5/6 and nsp6/7 sites are processed. This yields a set of "alternative" processing products which are not produced from the major pathway. An interesting consequence of the use of the two pathways is the generation of multiple SP-containing proteins. In a number of intermediates (e.g. nsp3–4 and nsp4–5) the SP is linked to a hydrophobic domain and may therefore be membrane-associated. The fully cleaved nsp4 is probably cytoplasmic and was shown to be efficient in processing other parts of the replicase polyprotein *in trans* (van Dinten *et al.*, unpublished observations).

Cofactors that strongly influence polyprotein processing have previously been identified in several other animal positive-stranded RNA virus systems. In most of these cases, the cofactor probably interacts (more or less) directly with the protease domain. In the case of EAV, however, the nsp2 cofactor role appears to be limited to just one of the eight cleavages carried out by the nsp4 SP. It may rather affect the conformation of the region containing the nsp4/5 junction than the activity of the nsp4 protease domain itself, since the SP was perfectly capable of processing other sites (nsp3/4, nsp5/6, nsp6/7, and nsp7/8) in the absence of nsp2. Nsp2 was previously reported to interact very strongly with nsp3 and nsp3-containing intermediates (Snijder *et al.*, 1994). Possibly, this interaction is the basis for the cofactor role of nsp2 in the "major" pathway for SP-directed nsp3–8 processing. After processing of the nsp4/5 junction in nsp3–8 (yielding nsp3–4 and nsp5–8), the conformation of nsp5–8 (51 kDa) appears to be such that cleavage of its internal nsp5/6 and nsp6/7 sites is prevented. However, the Glu-1677↓Gly-1678 site in the nsp5–8 C-terminal region is accessible to the SP, which results in the production of nsp5–7 and nsp8. Both nsp2 and nsp3–8 (and its derivatives) are complex molecules containing several clusters of conserved Cys residues (Snijder *et al.*, 1994) and multiple hydrophobic domains. The latter have recently been implicated in the membrane-association of the EAV replication complex (van Dinten *et al.*, 1996; van der Meer *et al.*, unpublished observations). Together, these results suggest that translation, processing, and membrane-association of the nsp2 to nsp8 region are interconnected processes.

5. MEMBRANE-ASSOCIATION OF THE EAV REPLICATION COMPLEX

The first indications for the membrane-association of the EAV replication complex were obtained when the rabbit antisera directed against various replicase regions (Snijder *et al.*, 1994; van Dinten *et al.*, 1996) were used for immunofluorescence assays on EAV-infected cells. With the exception of the anti-nsp1 serum, all antisera which could be used in this assay gave the same perinuclear staining (Figure 4). The co-localization of most of the replicase processing products, including those that contain the putative RNA polymerase (nsp9) and helicase (nsp10) functions (van Dinten *et al.*, 1996), indicated that these proteins assemble into a large replication/transcription complex. Their concentration in the

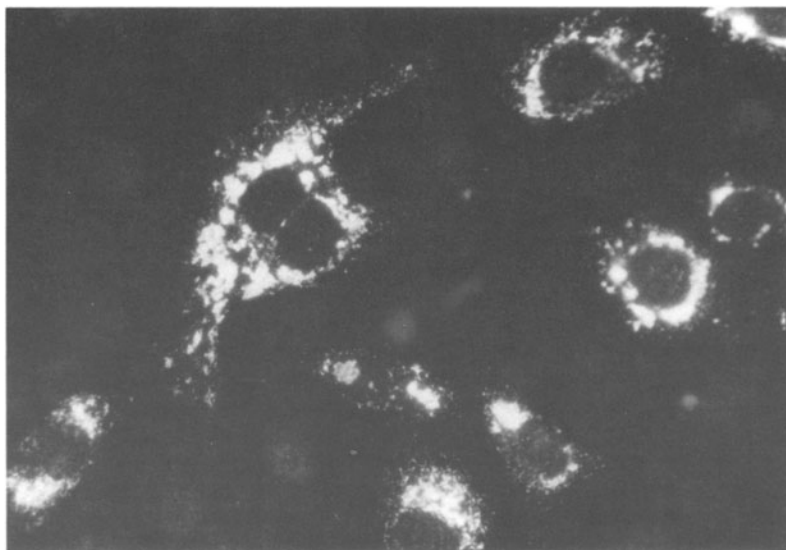

Figure 4. Immunofluorescence staining of the EAV replication complex in infected Vero cells using an antiserum directed against the C-terminal 220 residues of the ORF1a protein (Snijder *et al.*, 1994).

perinuclear region suggested association with intracellular membranes, which was subsequently confirmed by double-labeling experiments with marker antibodies directed against cellular membrane proteins (van der Meer *et al.*, unpublished observations). Using a metabolic RNA labeling, it was recently confirmed that the localization of the replicase proteins indeed coincides with the site of viral RNA transcription.

The ORF1a cleavage products nsp2, nsp3, and nsp5 contain hydrophobic domains which are likely to play a role in anchoring the EAV replication complex to intracellular membranes (Snijder *et al.*, 1994; van Dinten *et al.*, 1996). To confirm this biochemically, we recently isolated the membrane fraction from EAV-infected cells and compared it with the cytoplasmic fraction for the presence of replicase cleavage products (van der Meer *et al.*, unpublished observations). A number of major processing products (nsp2, nsp3–8, nsp3–4, nsp5–8, nsp5–7) was indeed predominantly recovered from the membrane fraction. This was also the case at pH 11, which suggests that these proteins probably contain transmembrane domains (Fujiki *et al.*, 1982). Upon Triton X-114 extraction (Bordier, 1981), the major part of the same proteins was retrieved from the detergent phase, which again confirmed their affinity for membranes. As expected, a number of cleavage products that lack hydrophobic domains displayed no affinity for the Triton X-114 detergent phase.

A typical feature of arterivirus replication is the formation of paired membranes and double membrane vesicles (DMV) at 3 to 6 hours post infection (Wood *et al.*, 1970; Breese and McCollum, 1970; Stueckemann *et al.*, 1982; Pol *et al.*, 1997). Although the origin and function of these membrane structures are unclear, they do not appear to be involved in virus assembly. Their possible relationship to the viral replication complex is currently under investigation using immuno electron microscopy and EAV replicase-specific antisera.

6. DEVELOPMENT OF AN INFECTIOUS cDNA CLONE FOR EAV

The molecular analysis and experimental manipulation of RNA virus genomes can be greatly facilitated by the availability of full-length cDNA clones from which functional RNA transcripts can be generated *in vitro*. This is especially true for positive-stranded RNA viruses since their genomes function as the mRNA for viral replicase translation. Thus, a virus infection can be initiated by the "simple" introduction of an infectious recombinant RNA molecule into a susceptible host cell. The availability of an infectious cDNA clone and the subsequent application of recombinant DNA technology ("reverse genetics") can enormously increase our understanding of many (molecular) aspects of the viral life cycle. This is e.g. illustrated by the advances which were made in the research on several animal RNA virus groups after infectious cDNA clones became available (5 to 15 years ago). Furthermore, full-length cDNA clones have proven to be valuable tools for pathogenesis studies and the development of vaccines and RNA virus expression vectors (Bredenbeek and Rice, 1992; Boyer and Haenni, 1994).

Since 1992, the development of a nidovirus infectious cDNA clone has been a major goal in our Department, since it was expected to be an essential tool to study (among other things) the common features of the replication and transcription of coronaviruses and arteriviruses. Also the development of nidovirus-based expression vectors would be very interesting because of the polycistronic genome organization and the "natural" generation of a set of subgenomic mRNAs. Compared to coronaviruses, the arteriviruses offered the obvious advantage of their considerably smaller genome size. In 1996, researchers in three Institutes in the Netherlands independently assembled arterivirus infectious cDNA clones. Full-length EAV cDNA clones were generated at Leiden (van Dinten *et al.*, 1997) and Utrecht (de Vries *et al.*, 1997), whereas an infectious PRRSV cDNA clone was developed at Lelystad (Meulenberg *et al.*, 1997).

The EAV infectious clone from Leiden (pEAV030; EMBL database accession number Y07862) was assembled from fully sequenced cDNA clones which had previously been used to determine the EAV genome sequence (den Boon *et al.*, 1991). The 17 most 5' nucleotides, which were lacking in the cDNA library, were determined (van Dinten *et al.*, 1997) and attached to the 5' end using a PCR. In the same PCR, a T7 RNA polymerase promoter was placed upstream of nucleotide G-1 of the EAV genome.

Unfortunately, RNA transcribed from the initial reconstruction (pEAV030F) was not infectious. When differences between two cDNA clones had been detected during the sequence analysis of the EAV genomic library, a third cDNA clone had always been analyzed to determine the consensus sequence. However, for two of these positions, nucleotide 5,899 and nucleotide 7,508, a third clone had not been available. One of these point mutations (C instead of T at position 7,508), which induces a Ser-2429 to Pro substitution in the ORF1b-encoded replicase part, was eventually identified as the reason why pEAV030F-transcripts are not infectious (van Dinten *et al.*, 1997). Upon correction of this error, BHK-21 cells could be infected by transfection (electroporation) of pEAV030 RNA transcripts and the first recombinant arterivirus, carrying a translationally silent mutation to discriminate it from the wild-type virus, was generated (van Dinten *et al.*, 1997). Subsequent experiments confirmed the potential for heterologous gene expression from arterivirus RNA vectors. It was shown to be possible to express a foreign gene (chloramphenicol acetyl transferase; CAT) from two different EAV subgenomic mRNAs. These CAT-expressing vectors were not infectious, due to the fact that (in both cases) the insertion prevented the expression of one of the viral structural proteins. Modified expression vectors which should retain their infectivity are currently under construction.

Surprisingly, the non-infectious RNA from the pEAV030F clone (which carried the T-7,508 to C mutation) turned out to be capable of efficient self-replication. In contrast, subgenomic mRNA synthesis (and therefore also structural protein expression) did not occur at a detectable level, explaining why virus particles were not produced (van Dinten *et al.*, 1997). This intriguing phenotype is most likely due to the Ser-2429 to Pro mutation in the EAV replicase. Residue Ser-2,429 is conserved in the three available arterivirus replicase sequences. It is located between two well-conserved domains in the nidovirus replicase: the putative zinc finger region (residues 2,374–2,426 in EAV) and the putative RNA helicase domain (residues 2,500–2,800 in EAV). These two domains are both part of the 50 kDa nsp10 subunit of the EAV replicase (van Dinten *et al.*, 1996). This mutant shows that the requirements for nidovirus genome replication and discontinuous mRNA synthesis are (at least in part) different. Since the mutation is located only 3 residues downstream of the most C-terminal conserved Cys residue of the putative zinc finger domain, it is currently being tested whether replacement of conserved Cys and His residues in this region yields the same phenotype (genome replication without subgenomic mRNA transcription).

7. REVERSE GENETICS USING THE EAV INFECTIOUS cDNA CLONE

In addition to the development of arterivirus-based expression vectors, the EAV full-length cDNA clone is currently being used for a number of fundamental studies into the arterivirus life cycle. This work includes an analysis of the functions of various replicase subunits, the role of replicase processing in the regulation of arterivirus replication, and the details of the discontinuous transcription mechanism used to generate the subgenomic mRNAs.

Obviously, the fortuitously detected phenotype of the Ser-2429 to Pro mutant is receiving considerable attention, both at the RNA (van Marle *et al.*, unpublished observations) and at the protein level (van Dinten *et al.*, unpublished observations). The effects of a number of internal deletions in the full-length cDNA clone have been examined and these studies have yielded preliminary information about the RNA sequences required for genome replication. At the 3' end not more than 200 nt upstream of the poly(A) tail are required for RNA replication. However, transcripts containing longer 3'-terminal sequences seem to replicate considerably more efficient. Deletion of ORFs 2 to 7 yielded an RNA that could replicate and produce a subgenomic mRNA from the single remaining mRNA promoter (the RNA2 promoter in the 3' part of ORF1b). This showed that the products of EAV ORFs 2 to 7 (three known viral envelope proteins, the nucleocapsid protein, and two as yet unidentified proteins) are not required for genome replication or mRNA transcription. At the 5' end, deletion of the nsp1-encoding region yielded an RNA that could replicate but did not produce subgenomic mRNAs. Whether the deletion of certain RNA sequences from the 5'-terminal region of the genome or the absence of nsp1 forms the basis for this phenotype is currently being investigated.

8. CONCLUDING REMARKS

With the development of infectious cDNA clones, arterivirus (and nidovirus) research has entered a new era. The possibility to genetically modify two economically relevant arteriviruses could lead to the rapid development of (marker) vaccines and

arterivirus-based expression vectors. At the fundamental level, the functional dissection of the EAV replicase polyprotein and the RNA transcription processes are important challenges, especially in view of the similarities and differences between arteriviruses and coronaviruses. At present, the set of structural proteins employed by arteriviruses and the structure of the virion into which they assemble, appear to be unique. Still, common features of the assembly of coronaviruses and arteriviruses may emerge, once the structural proteins of the latter group are characterized in more detail. It is striking that both groups of viruses assemble at intracellular membranes and contain a triple-spanning membrane (M) protein, which is assumed to play a crucial role in this process. Since arteriviruses are now known to both replicate and assemble at intracellular membranes, one could speculate on the possibility of an integrated process of genome replication, genome encapsidation, and virus budding. In any case, the analysis of the interactions between the viral replication machinery and organelles and molecules of the host cell will be an important novel aspect of future nidovirus research.

ACKNOWLEDGMENTS

This chapter would not have been written without the continuous experimental, theoretical, and/or moral support of: Johan den Boon, Leonie van Dinten, Sasha Gorbalenya, Yvonne van der Meer, Sietske Rensen, Willy Spaan, Hans van Tol, and (most of all) Fred Wassenaar.

REFERENCES

Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A. E., Koonin, E. V., andLai, M. M., 1993, Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus, *J. Virol.* **67**:6056–6063.

Bazan, J. F., and Fletterick, R. J., 1988, Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications, *Proc. Natl. Acad. Sci. USA* **85**:7872–7876.

Bazan, J. F., and Fletterick, R. J., 1990, Structural and catalytic models of trypsin-like viral proteases, *Semin. Virol.* **1**:311–322.

Bonilla, P. J., Gorbalenya, A. E., and Weiss, S. R., 1994, Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: heterogeneity among MHV strains, *Virology* **198**:736–740.

Bonilla, P. J., Hughes, S. A., Pinon, J. D., and Weiss, S. R., 1995, Characterization of the leader papain-like proteinase of MHV A-59: identification of a new in vitro cleavage site, *Virology* **209**:489–497.

Bordier, C., 1981, Phase separation of integral membrane proteins in Triton X-114 solution, *J. Biol. Chem.* **256**:1604–1607.

Boyer, J. C. and Haenni, A. L., 1994, Infectious transcripts and cDNA clones of RNA viruses, *Virology* **198**:415–426.

Bredenbeek, P. J., and Rice, C. M., 1992, Animal RNA virus expression systems, *Semin. Virol.* **3**:297–310.

Breese, S. S.,Jr. and McCollum, W. H., 1970, *Proceedings of the 2nd International Conference on Equine Infectious Diseases* (Bryans, J. T. and Gerber, H. Eds.) S. Karger, Basel. pp.133–139.

Brierley, I., 1995, Ribosomal frameshifting on viral RNAs, *J. Gen. Virol.* 76:1885–1892. Brown, T. D. K., and Brierley, I., 1995, *The Coronaviridae* (Siddell, S. G. Ed.) Plenum Press, New York. pp. 191–217.

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

Cavanagh, D., Brian, D. A., Brinton, M. A., Enjuanes, L., Holmes, K. V., Horzinek, M. C., Lai, M. M. C., Laude, H., Plagemann, P. G. W., Siddell, S. G., Spaan, W. J. M., Taguchi, F., and Talbot, P. J., 1993, The coronaviridae now comprises two genera, coronavirus and torovirus: report of the coronaviridae study group, *Adv. Exp. Med. Biol.* **342**:255–257.

Cavanagh, D., Brian, D. A., Brinton, M. A., Enjuanes, L., Holmes, K. V., Horzinek, M. C., Lai, M. M. C., Laude, H., Plagemann, P. G. W., Siddell, S. G., Spaan, W. J. M., Taguchi, F., and Talbot, P. J., 1994, Revision of the taxonomy of the Coronavirus, Torovirus, and Arterivirus genera, *Arch. Virol.* **135**:227–237.

de Groot, R. J., Hardy, W. R., Shirako, Y., and Strauss, J. H., 1990, Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo, *EMBO J.* **9**:2631–2638.

de Vries, A. A. F., Horzinek, M. C., Rottier, P. J. M., and de Groot, R. J., 1997, The genome organization of the Nidovirales: similarities and differences between arteri-, toro-, and coronaviruses, *Semin. Virol.*, in press.

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

den Boon, J. A., Faaberg, K. S., Meulenberg, J. J. M., Wassenaar, A. L. M., Plagemann, P. G. W., Gorbalenya, A. E., and Snijder, E. J., 1995a, Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases, *J. Virol.* **69**:4500–4505.

den Boon, J. A., Spaan, W. J. M., and Snijder, E. J., 1995b, Equine arteritis virus subgenomic RNA transcription: UV inactivation and translation inhibition studies, *Virology* **213**:364–372.

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

Dougherty, W. G., and Semler, B. L., 1993, Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes, *Microbiol. Rev.* **57**:781–822.

Fujiki, Y., Hubbard, A. L., Fowler, S., and Lazarow, P. B., 1982, Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum, *J. Cell Biol.* **93**:97–102.

Gao, H. Q., Schiller, J. J., and Baker, S. C., 1996, Identification of the polymerase polyprotein products p72 and p65 of the murine coronavirus MHV JHM, *Virus Res.* **45**:101–109.

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

Gorbalenya, A. E., and Snijder, E. J., 1996, Viral cysteine proteases, *Perspect. Drug Discov. Design* **6**:64–86.

Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M., and Koonin, E. V., 1989a, Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold, *FEBS Lett.* **243**:103–114.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M., 1989b, Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis, *Nucleic Acids Res.* **17**:4847–4861.

Gorbalenya, A. E., Koonin, E. V., and Lai, M. M., 1991, Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses, *FEBS Lett.* **288**:201–205.

Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E., 1988, Signals for ribosomal frameshifting in the Rous sarcoma virus gagpol region, *Cell* **55**:447–458.

Jore, J., De Geus, B., Jackson, R. J., Pouwels, P. H., and Enger Valk, B. E., 1988, Poliovirus protein 3CD is the active protease for processing of the precursor protein P1 in vitro, *J. Gen. Virol.* **69**:1627–1636.

Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., and Lai, M. M. C., 1991, The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase, *Virology* **180**: 567–582.

Lemm, J. A., and Rice, C. M., 1993a, Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription, *J. Virol.* **67**: 1916- 1926.

Lemm, J. A., and Rice, C. M., 1993b, Assembly of functional Sindbis virus RNA replication complexes: requirement for coexpression of p123 and p34, *J. Virol.* **67**:1905–1915.

Lemm, J. A., Rumenapf, T., Strauss, E. G., Strauss, J. H., and Rice, C. M., 1994, Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis, *EMBO J.* **13**:2925–2934.

Liu, D. X. and Brown, T. D. K., 1995, Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein, *Virology* **209**:420–427.

Lu, Y., Lu, X., and Denison, M. R., 1995, Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59, *J. Virol.* **69**:3554–3559.

Meulenberg, J. J. M., Hulst, M. M., de Meijer, E. J., Moonen, P. L., den Besten, A., de Kluyver, E. P., Wensvoort, G., and Moormann, R. J. M., 1993, 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., Wensvoort, G., and Moormann, R. J. M., 1997, Infectious transcripts from cloned genome-length cDNA of porcine reproductive respiratory syndrome virus, *J. Virol.*, submitted.

Pol, J. M., Wagenaar, F., and Reus, J. E. G., 1997, Comparative morphogenesis of three PRRS virus strains, *Vet. Microbiol.* **55**:203–208.

Shirako, Y., and Strauss, J. H., 1994, Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis, *J. Virol.* **68**:1874–1885.

Snijder, E. J., and Meulenberg, J. J. M., 1998, The molecular biology of arteriviruses, *J. Gen. Virol.*, submitted.

Snijder, E. J., and Spaan, W. J. M., 1995, *The Coronaviridae* (Siddell, S. G. Ed.) Plenum Press, New York. pp. 239–255.

Snijder, E. J., den Boon, J. A., Bredenbeek, P. J., Horzinek, M. C., Rijnbrand, R., and Spaan, W. J., 1990, 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., Wassenaar, A. L. M., and Spaan, W. J. M., 1992, The 5' end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease, *J. Virol.* **66**:7040–7048.

Snijder, E. J., Wassenaar, A. L. M., and Spaan, W. J. M., 1994, Proteolytic processing of the replicase ORF1a protein of equine arteritis virus, *J. Virol.* **68**:5755–5764.

Snijder, E. J., Wassenaar, A. L. M., Spaan, W. J. M., and Gorbalenya, A. E., 1995, The arterivirus nsp2 protease. an unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases, *J. Biol. Chem.* **270**:16671–16676.

Snijder, E. J., Wassenaar, A. L. M., van Dinten, L. C., Spaan, W. J. M., and Gorbalenya, A. E. 1996. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases, *J. Biol. Chem.* **271**:4864–4871.

Snijder, E. J. and Horzinek, M. C, 1993, Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily, *J. Gen. Virol.* **74**:2305–2316.

Stueckemann, J. A., Holth, M., Swart, W. J., Kowalchyk, K., Smith, M. S., Wolstenholme, A. J., Cafruny, W. A., and Plagemann, P. G. W., 1982, Replication of lactate dehydrogenase-elevating virus in macrophages. 2. mechanism of persistent infection in mice and cell culture, *J. Gen. Virol.* **59**:263–272.

van Dinten, L. C., Wassenaar, A. L. M., Gorbalenya, A. E., Spaan, W. J. M., and Snijder, E. J., 1996, Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains, *J. Virol.* **70**:6625–6633.

van Dinten, L. C., den Boon, J. A., Wassenaar, A. L. M., Spaan, W. J. M., and Snijder, E. J., 1997, An infectious arterivirus cDNA clone: identification of a replicase point mutation which abolishes discontinuous mRNA transcription, *Proc. Natl. Acad. Sci. USA* **94**:991–996.

Wassenaar, A. L. M., Spaan, W. J. M., Gorbalenya, A. E., and Snijder, E. J., 1997, Alternative proteolytic processing of the arterivirus ORF1a polyprotein: evidence that nsp2 acts as a cofactor for the nsp4 serine protease, *J. Virol.*, submitted.

Wood, O., Tauraso, N. M., and Liebhaber, H. 1970. Electron microscopic study of tissue cultures infected with simian haemorrhagic fever virus, *J. Gen. Virol.* **7**:129–136.

Ypma-Wong, M. F., Dewalt, P. G., Johnson, V. H., Lamb, J. G., and Semler, B. L,. 1988, Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor, *Virology* **166**:265–270.

Ziebuhr, J., Herold, J., and Siddell, S. G., 1995, Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity, *J. Virol.* **69**:4331–4338.

14

REPLICATION AND TRANSCRIPTION OF HCV 229E REPLICONS

Volker Thiel, Stuart G. Siddell, and Jens Herold

Institute of Virology and Immunology
University of Wuerzburg
Versbacherstr. 7
97078 Wuerzburg, Germany

1. ABSTRACT

Replicons based upon the human coronavirus 229E (HCV 229E) genome were transfected into HCV 229E infected cells. We demonstrate that a synthetic RNA comprised of 646 nucleotides from the 5' end and 1465 nucleotides from the 3' end of the HCV 229E genome is replication competent. We conclude that the cis-acting elements necessary for replication are located in these 5' and 3' genomic regions. Furthermore, we inserted the intergenic region of the HCV 229E nucleocapsid protein gene into this basic construct and were able to demonstrate the transcription of "subgenomic" RNAs

2. INTRODUCTION

The study of coronavirus replication has been greatly facilitated by the use of defective interfering (DI) RNAs that can replicate in coronavirus infected cells (Makino and Lai, 1989; Van der Most et al., 1991). The analysis of several different DI RNAs has revealed that cis-acting elements required for replication are located at the 5' and 3' end of the coronavirus genome. Furthermore, the insertion of an intergenic sequence into these DI RNAs leads to the synthesis of "subgenomic" RNA transcripts (Makino et al., 1991). In order to establish a similar system to study HCV 229E replication, we have constructed an HCV 229E replicon containing 5' and 3' sequences of the HCV 229E genome behind a T7 RNA polymerase promoter. Furthermore, we have inserted the intergenic sequence of the HCV 229E nucleocapsid protein gene at different positions within this construct and tested for replication and transcription. The potential use of this repli-

Coronaviruses and Arteriviruses, edited by Enjuanes *et al.*
Plenum Press, New York, 1998