

Virus-based vectors for gene expression in mammalian cells: Coronavirus

Luis Enjuanes, Fernando Almazán, and Javier Ortego

*Department of Molecular and Cell Biology, Centro Nacional de Biotecnología, CSIC, Campus Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain; Tel.: +34 (91) 585-4555; Fax: +34 (91) 585-4915.
E-mail: lenjuanes@cnb.uam.es*

1. Introduction

The coronavirus and the torovirus genera form the *Coronaviridae* family, which is closely related to the *Arteriviridae* family. Both families are included in the *Nidovirales* order [1,2]. Recently, a new group of invertebrate viruses, the *Roniviridae*, with a genetic structure and replication strategy similar to those of coronaviruses, has been described [3]. This new virus family has been included within the *Nidovirales* [4].

Coronaviruses have several advantages as vectors over other viral expression systems: (i) coronaviruses are single-stranded RNA viruses that replicate within the cytoplasm without a DNA intermediary, making integration of the virus genome into the host cell chromosome unlikely [5]; (ii) these viruses have the largest RNA virus genome and, in principle, have room for the insertion of large foreign genes [1,6]; (iii) a pleiotropic secretory immune response is best induced by the stimulation of gut associated lymphoid tissues. Since coronaviruses in general infect the mucosal surfaces, both respiratory and enteric, they may be used to target the antigen to the enteric and respiratory areas to induce a strong secretory immune response; (iv) the tropism of coronaviruses may be modified by manipulation of the spike (S) protein allowing engineering of the tropism of the vector [7,8]; (v) non-pathogenic coronavirus strains infecting most species of interest (human, porcine, bovine, canine, feline, and avian) are available to develop expression systems; and (vi) infectious coronavirus cDNA clones are available to design expression systems.

Within the coronavirus two types of expression vectors have been developed (Fig. 1), one requires two components (helper-dependent expression system) (Fig. 1A) and the other a single genome that is modified either by targeted recombination [6] (Fig. 1B.1) or by engineering a cDNA encoding an infectious RNA. Infectious cDNA clones are available for porcine [9,10] (Fig. 1B.2 and B.3), human (Fig. 1B.4) [11], murine [12] and avian (infectious bronchitis virus, IBV) coronavirus [13], and also for the arteriviruses equine infectious anemia virus (EAV) [14], porcine respiratory and reproductive syndrome virus (PRRSV) [15], and simian hemorrhagic fever virus (SHFV) [16]. The availability of these cDNAs and the application of target recombination to coronaviruses [6] have been essential for the development of vectors based on coronaviruses and arteriviruses.

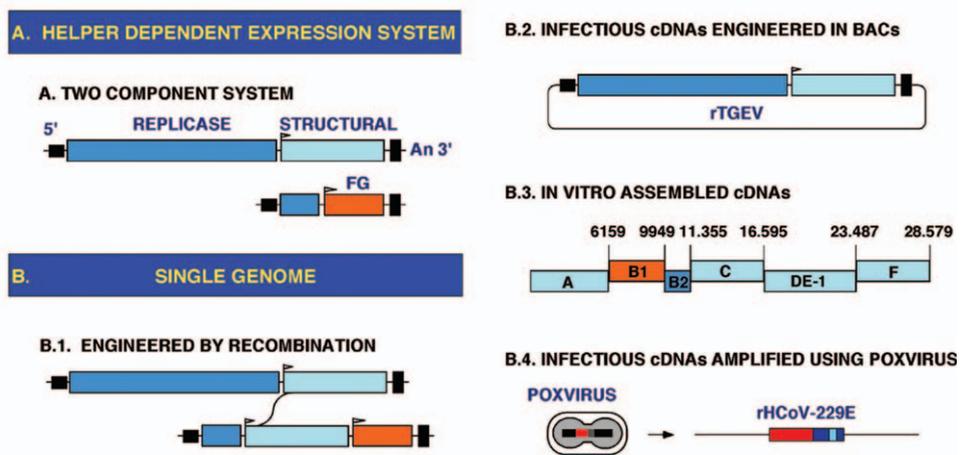


Fig. 1. Coronavirus-derived expression systems. (A) Helper dependent expression system based on two components, the helper virus and a minigenome carrying the foreign gene (FG). An, poly A. (B) Single genome engineered by targeted recombination (B.1), by assembling an infectious cDNA clone derived from TGEV genome in BACs (B.2), by the *in vitro* ligation of six cDNA fragments (B.3), or by using poxviruses as the cloning vehicle (B.4).

This review will focus on the advantages and limitations of these coronavirus expression systems, the attempts to increase their expression levels by studying the transcription-regulating sequences (TRSs), and the proven possibility of modifying their tissue and species-specificity.

2. Coronavirus pathogenicity

Coronaviruses comprise a large family of viruses infecting a broad range of vertebrates, from mammalian to avian species. Coronaviruses are associated mainly with respiratory, enteric, hepatic and central nervous system diseases. In humans and fowl, coronaviruses primarily cause upper respiratory tract infections, while porcine and bovine coronaviruses (BCoVs) establish enteric infections that result in severe economical loss. Human coronaviruses (HCoV) are responsible for 10–20% of all common colds, and have been implicated in gastroenteritis, high and low respiratory tract infections and rare cases of encephalitis. HCoV have also been associated with infant necrotizing enterocolitis and are tentative candidates for multiple sclerosis. In March 2003, a new group of HCoVs has emerged as the ethiological agent of the severe acute pneumonia syndrome (SARS) affecting thousands of people, mostly in China, Singapore, and Toronto. In addition, human infections by coronaviruses seem to be ubiquitous, as coronaviruses have been identified wherever they have been looked for, including North and South America, Europe, and Asia and no other human disease has been clearly associated with them with the exception of respiratory and enteric infections.

3. Molecular biology of coronavirus

3.1. Coronavirus genome

Virions contain a single molecule of linear, positive-sense, single-stranded RNA (Fig. 2). The coronavirus genome with a size ranging from 27.6 to 31.3 kb is the largest viral RNA known. Coronavirus RNA has a 5' terminal cap followed by a leader sequence of 65–98 nucleotides and an untranslated region of 200–400 nucleotides. At the 3' end of the genome there is an untranslated region of 200–500 nucleotides followed by a poly(A) tail. The virion RNA, which functions as an mRNA and is infectious, contains approximately 7–10 functional genes, four or five of which encode structural proteins. The genes are arranged in the order 5'-polymerase-(HE)-S-E-M-N-3', with a variable number of other genes that are believed to be non-structural and largely non-essential, at least in tissue culture [1].

About two-thirds of the entire RNA comprises the Rep1a and Rep1b genes. At the overlap between the Rep1a and 1b regions, there is a specific seven-nucleotide “slippery” sequence and a pseudoknot structure (ribosomal frameshifting signal), which are required for the translation of Rep1b as a single polypeptide (Rep1a/b). The 3' third of the genome comprises the genes encoding the structural proteins and the other non-structural ones.

Coronavirus transcription occurs via an RNA-dependent RNA synthesis process in which mRNAs are transcribed from negative-stranded templates.

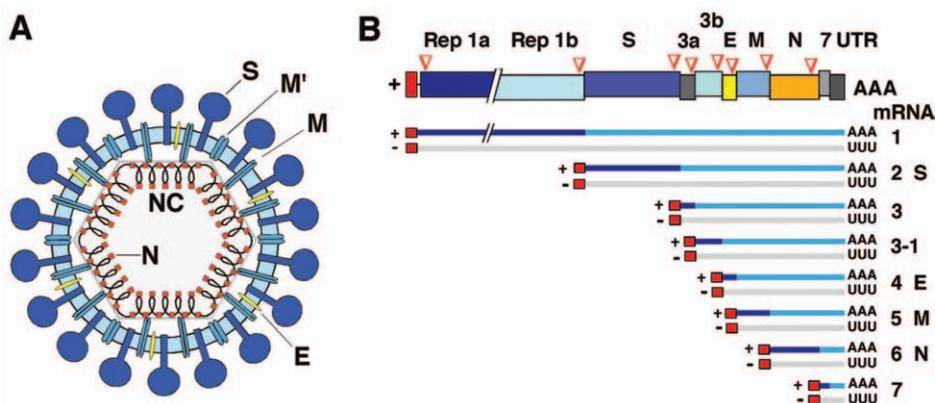


Fig. 2. Coronavirus structure and genome organization. (A) Schematic diagram of coronavirus structure using TGEV as a prototype. The diagram shows the envelope, the core and the nucleoprotein structure. S, spike protein; M and M', large membrane proteins with the amino-terminus facing the external surface of the virion and the carboxy-terminus towards the inside or the outside surface of the virion, respectively; E, small envelope protein; N, nucleocapsid protein; NC, nucleocapsid. (B) Representation of a prototype TGEV coronavirus genome and subgenomic RNAs. Beneath the top bar a set of positive- and negative-sense mRNA species synthesized in infected cells is shown. Poly(A) and Poly(U) tails are indicated by AAA or UUU. Rep1a and Rep1b, replicase genes; UTR, untranslated region; ∇, indicates presence of a TRS; other acronyms as in A.

Coronavirus mRNAs consist of six to eight types of varying sizes, depending on the coronavirus strain and the host species. The largest mRNA is the genomic RNA that also serves as the mRNA for Rep1a and 1b, the remainder are subgenomic mRNAs (sgmRNAs). The mRNAs have a nested-set structure in relation to the genome structure (Fig. 2B).

3.2. Coronavirus proteins

Coronaviruses are enveloped viruses containing a core that includes the ribonucleoprotein formed by the RNA and nucleoprotein N (Fig. 2A). The core is formed by the genomic RNA, the N protein and the membrane (M) protein carboxy-terminus. Most of the M protein is embedded within the membrane but its carboxy-terminus is integrated within the core and seems essential to maintain the core structure [17,18]. At least in the transmissible gastroenteritis virus (TGEV) the M protein presents two topologies. In one (M'), both the amino and the carboxyl termini face the outside of the virion, while in the other (M) the carboxy-terminus is inside [18]. In addition, the virus envelope contains two or three other proteins, the spike (S) protein that is responsible for cell attachment, the small membrane protein (E) and, in some strains, the hemagglutinin-esterase (HE) [1].

The replicase gene encodes a protein of approximately 740–800 kDa which is co-translationally processed. Several domains within the replicase have predicted functions based on regions of nucleotide homology [19].

4. Helper-dependent expression systems

The coronaviruses have been classified into three groups (1, 2 and 3) based on sequence analysis of a number of coronavirus genes [1]. Helper-dependent expression systems have been developed using members of the three groups of coronaviruses (Fig. 1).

4.1. Group 1 coronaviruses

Group 1 coronaviruses include porcine, canine, feline and HCoV. Expression systems have been developed for the porcine and HCoV since minigenomes are only available for these two coronaviruses. One expression system has been developed using TGEV-derived minigenomes [20]. The TGEV-derived RNA minigenomes were successfully expressed *in vitro* using T7 polymerase and amplified after *in vivo* transfection using a helper virus. TGEV-derived minigenomes of 3.3, 3.9 and 5.4 kb were efficiently used for the expression of heterologous genes [21,22]. The smallest minigenome replicated by the helper virus and efficiently packaged was 3.3 kb in length [20].

Using M39 minigenome, a two-step amplification system was developed based on the cloning of a cDNA copy of the minigenome after the immediate-early cytomegalovirus promoter (CMV) [20]. Minigenome RNA is first amplified in the

nucleus by the cellular RNA *pol II* and then, the RNAs are translocated into the cytoplasm where they are amplified by the viral replicase of the helper virus. The β -glucuronidase (GUS) and a surface glycoprotein (ORF5), that is the major protective antigen of the PRSSV, have been expressed using this vector [22]. TGEV-derived helper expression systems have a limited stability and minigenomes without the foreign gene replicate about 50-fold more efficiently than those with the heterologous gene [22]. Expression of GUS gene and PRSSV ORF5 with these minigenomes has been demonstrated in the epithelial cells of alveoli and in scattered pneumocytes of swine lungs, which led to the induction of a strong immune response to these antigens [22].

The HCoV-229E has also been used to express new sgmRNAs [23]. It was demonstrated that a synthetic RNA including 646 nt from the 5' end plus 1465 nt from the 3' end was amplified by the helper virus.

4.2. Group 2 coronaviruses

Most of the work has been done with mouse hepatitis virus (MHV) defective RNAs (minigenomes) [24,25]. Three heterologous genes have been expressed using the MHV system, chloramphenicol acetyltransferase (CAT), HE, and interferon γ (IFN- γ). Expression of CAT or HE was detected only in the first two passages because the minigenome used lacks the packaging signal [26]. When virus vectors expressing CAT and HE were inoculated intracerebrally into mice, HE- or CAT-specific sgmRNAs were only detected in the brains at days 1 and 2, indicating that the genes in the minigenome were expressed only in the early stage of viral infection [27].

A MHV minigenome RNA was also developed as a vector for expressing IFN- γ . The murine IFN- γ gene was secreted into culture medium as early as 6 h post-transfection and reached a peak level at 12 h post-transfection. No inhibition of virus replication was detected when the cells were treated with IFN- γ produced by the minigenome RNA, but infection of susceptible mice with a minigenome producing IFN- γ caused significantly milder disease, accompanied by less virus replication than that caused by virus containing a control vector [25,28].

4.3. Group 3 coronaviruses

IBV is an avian coronavirus with a single-stranded, positive-sense RNA genome of 27,608 nt. A defective RNA (CD-61) derived from the Beaudette strain of the IBV virus was used as an RNA vector for the expression of two reporter genes, luciferase and CAT [29].

4.4. Heterologous gene expression levels in helper-dependent expression systems

Helper-dependent expression systems have a limited stability probably due to the foreign gene since TGEV minigenomes of 9.7, 3.9 and 3.3 kb, in the absence of the heterologous gene, are amplified and efficiently packaged for at least 30 passages, without generating new dominant subgenomic RNAs [20]. The expression of GUS,

PRSSV ORF5, or CAT using TGEV- or IBV-derived minigenomes in general increases until passages three or four, expression levels are maintained for about four additional passages, and steadily decrease during successive passages [20–22,29]. Using IBV minigenomes CAT expression levels between 1 and 2 μg per 10^6 cells have been described. The highest expression levels (2–8 μg of GUS per 10^6 cells) have been obtained using a two-step amplification system based on TGEV derived minigenomes with optimized TRSs [20,21]. Using minigenomes derived from TGEV and IBV expression was highly dependent on the nature of the heterologous gene used. Luciferase expression with TGEV and IBV minigenomes was reduced to almost background levels, while expression of GUS or CAT was at least 100–1000-fold higher than background levels, respectively.

5. *Single genome coronavirus vectors*

5.1. *Group 1 coronaviruses*

The construction of cDNA clones encoding full-length coronavirus RNAs has considerably improved the genetic manipulation of coronaviruses. The enormous length of the coronavirus genome and the instability of plasmids carrying coronavirus replicase sequences have hampered, until recently, the construction of a full-length cDNA clone. Infectious coronavirus cDNA clones have been described for coronaviruses [9,10,13] and for arteriviruses [14,15].

The strategy used to clone TGEV infectious cDNA was based on three points [9]: (i) the construction was started from a minigenome that was stably and efficiently replicated by the helper virus [20]. During the filling in of minigenome deletions a cDNA fragment that was toxic to the bacterial host was identified. This fragment was reintroduced into the cDNA in the last cloning step; (ii) in order to express the long coronavirus genome and to add the 5' cap required for TGEV RNA infectivity, a two-step amplification system that couples transcription in the nucleus from the CMV promoter, with a second amplification in the cytoplasm driven by the viral polymerase, was used; and (iii) to increase viral cDNA stability within bacteria, the cDNA was cloned as a bacterial artificial chromosome (BAC), that produces a maximum of two plasmid copies per cell. A fully functional infectious TGEV cDNA clone, leading to a virulent virus infecting both the enteric and respiratory tract of swine was engineered. The stable propagation of a TGEV full-length cDNA in bacteria as a BAC has been considerably improved by the insertion of an intron to disrupt a toxic region identified in the viral genome (Fig. 3) [30]. The viral RNA was expressed in the cell nucleus under the control of the CMV promoter and the intron was efficiently removed during translocation of this RNA to the cytoplasm. Intron insertion in two different positions (nt 9466 and 9596) allowed stable plasmid amplification for at least 200 generations. Infectious TGEV was efficiently recovered from cells transfected with the modified cDNAs. The great advantage of this system is that the performance of coronavirus reverse genetics only involves recombinant DNA technologies carried out within the bacteria.

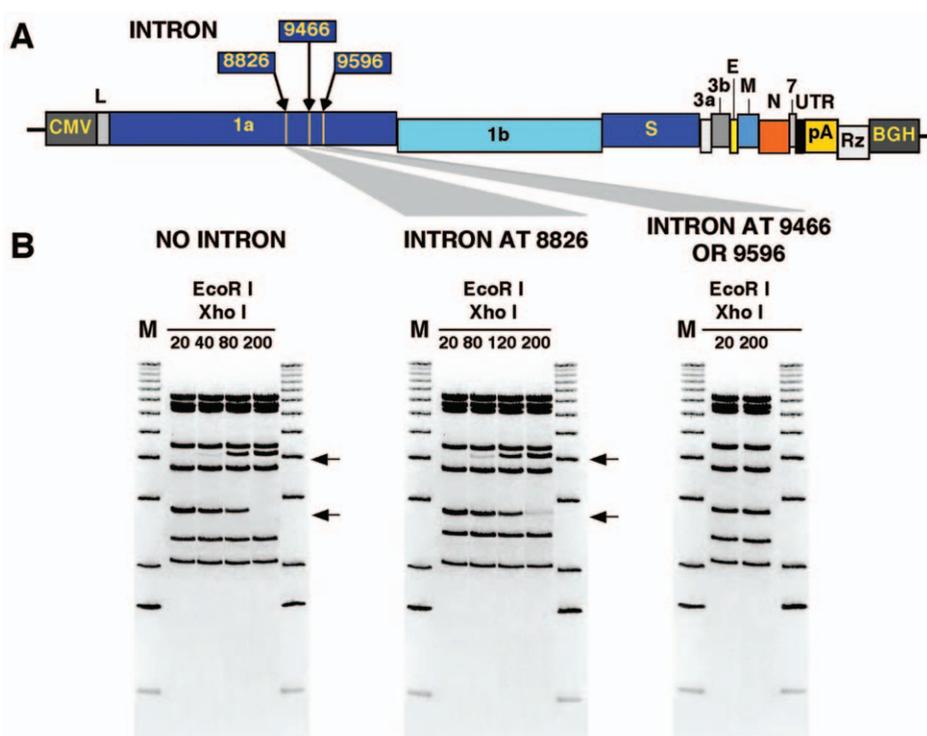


Fig. 3. Intron insertion to stabilize TGEV full-length cDNA. (A) Strategy for the insertion of the 133-nt intron in the indicated positions of the TGEV sequence. (B) Analysis of the three intron-containing TGEV full-length cDNAs in *Escherichia coli* cells. The *EcoRI*-*XhoI* restriction patterns of the three plasmids extracted from *E. coli* cells grown for the indicated number of generations are shown. Arrows indicate disappearance or appearance of a band. M, molecular mass markers. UTR, untranslated region. Rz, hepatitis delta virus ribozyme; BGH, bovine growth hormone termination and polyadenylation sequences. Other acronyms as in Fig. 2.

Using TGEV cDNA the green fluorescent protein (GFP) gene of 0.72 kb was cloned in two positions of the RNA genome: either by replacing the non-essential 3a and 3b genes or between genes N and 7. The engineered genome was very stable (> 30 passages in cultured cells) and led to the production of high expression levels ($50 \mu\text{g}/10^6$ cells) when the GFP replaced genes 3a/b but was unstable when cloned between genes N and 7 [31]. In this case, the GFP gene was eliminated by homologous recombination between preexisting TRS sequences and those introduced to express GFP. Using the most stable vector, the acquisition of immunity by newborn piglets breast fed by immunized sows (lactogenic immunity) was demonstrated [31]. GUS expression levels using coronavirus based vectors are similar (Fig. 4) to those described for vectors derived from other positive strand RNA viruses such as Sindbis virus ($50 \mu\text{g}$ per 10^6 cells) [32].

A second procedure to assemble a full-length infectious construct of TGEV was based on the *in vitro* ligation of six adjoining cDNA subclones that span the entire

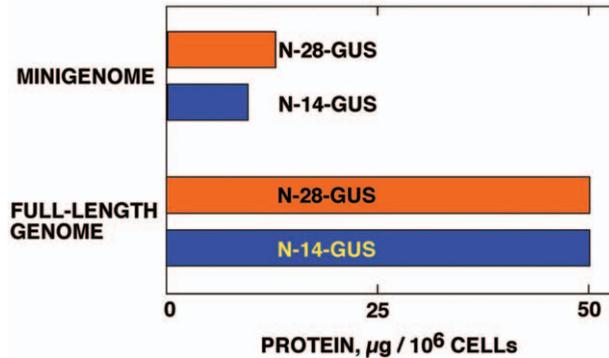


Fig. 4. GUS expression levels using TGEV-based vectors. The amount of protein expressed with the helper dependent expression system (minigenome) is compared with the expression levels from full-length genomes using the same TRSs (N-14 or N-28) from N gene.

TGEV genome [10]. Each clone was engineered with unique flanking interconnecting junctions that determine a precise assembly with only the adjacent cDNA subclones, resulting in a TGEV cDNA. *In vitro* transcripts derived from the full-length TGEV construct were infectious. Using this construct, a recombinant TGEV was assembled that replaced ORF 3a with the GFP gene, leading to the production of a recombinant TGEV that grew with titers of 10^8 pfu/ml and expressed GFP in a high proportion of cells [33].

An infectious cDNA clone has also been constructed for HCoV-229E, another member of group 1 coronaviruses [11]. In this case, the system is based on the *in vitro* transcription of infectious RNA from a cDNA copy of the HCoV-229E genome that has been cloned and propagated in vaccinia virus (Fig. 5). Briefly, the full-length genomic cDNA clone of HCoV-229E was assembled by *in vitro* ligation, and then cloned into the vaccinia virus DNA under control of the T7 promoter. Recombinant vaccinia viruses containing the HCoV-229E genome were recovered after transfection of the recombinant vaccinia virus DNA into cells infected with fowlpox virus. In a second phase, the recombinant vaccinia virus DNA was purified and used as a template for *in vitro* transcription of HCoV-229E genomic RNA that was transfected into susceptible cells for the recovery of infectious recombinant coronavirus (Fig. 5).

A coronavirus replicon has been derived from the HCoV genome using the same procedure described for the full-length genome construction [34]. This replicon included the 5' and 3' ends of the HCoV-229E genome, the replicase gene of this virus and a single reporter gene coding for GFP located downstream of a TRS element for coronavirus mRNA transcription. When RNA transcribed from this cDNA was transfected into BHK-21 cells, only 0.1% of the cells showed strong fluorescence. This data shows that the coronavirus replicase gene products suffice for discontinuous sgRNA transcription, in agreement with the requirements for the arterivirus replicase [35].

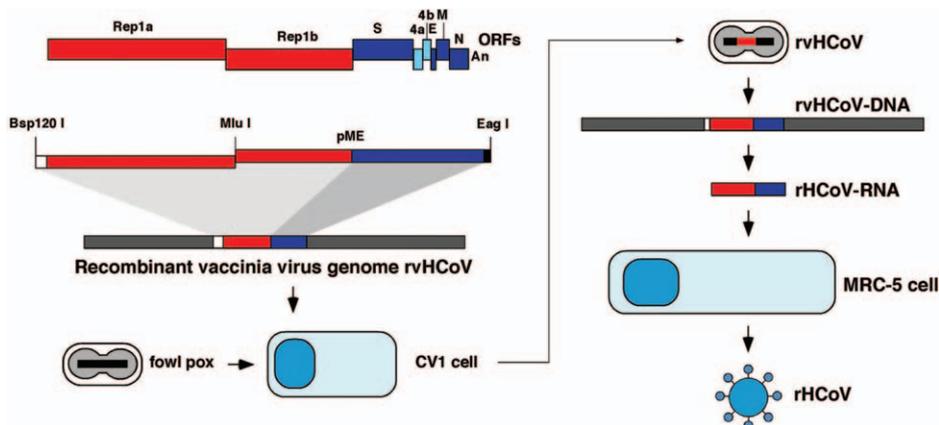


Fig. 5. Infectious HCoV-229E recovery from cDNA. The genetic map of HCoV-229E (top) is shown. ORFs encoding virus replicase proteins and structural or non-structural proteins are colored in light and dark gray, respectively. Vaccinia virus DNA (black box) and the T7 promoter (white box) are also shown [11].

The expression of a heterologous gene (GFP) by a TGEV replicon was increased between 300- and 400-fold when TGEV N protein was *in cis* co-expressed [36]. In addition, expression from a TGEV replicon was also observed when N protein was *in trans* co-expressed using the Venezuelan equine encephalitis virus vector [36]. Furthermore, expression from HCoV based vectors also was significantly increased by co-expression of N gene. Therefore, it seems that N protein either stabilizes coronavirus replicons or increases their replication, transcription or translation.

5.2. Group 2 coronaviruses

Reverse genetics in this coronavirus group has been efficiently performed by targeted recombination between a helper virus and either non-replicative or replicative coronavirus-derived RNAs (Fig. 2B.1). This approach, developed by Masters' group [6,37], was first applied to the engineering of a five-nucleotide insertion into the 3' untranslated region (3' UTR) of MHV [37]. This approach was facilitated by the availability of an N gene mutant, designated Alb4, that was both temperature-sensitive and thermolabile. Alb4 forms tiny plaques at restrictive temperature that are easily distinguishable from wild-type plaques. In addition, incubation of Alb4 virions at non-permissive temperature results in a 100-fold greater loss of titer than for wild-type virions. These phenotypic traits allowed the selection of recombinant viruses generated by a single crossover event following cotransfection into mouse cells of Alb4 genomic RNA together with a synthetic copy of the smallest subgenomic RNA (RNA7) tagged with a marker in the 3' UTR.

An improvement of the recombination frequency was obtained between the helper virus and replicative defective RNAs as the donor species. Whereas between replication competent MHV and non-replicative RNAs a recombination frequency

of the order of 10^{-5} was estimated, the use of replicative donor RNA yielded recombinants at a rate of some three orders of magnitude higher [38]. This higher efficiency made it possible to screen for recombinants even in the absence of selection. In this manner, the transfer of silent mutation in Rep1a gene of a minigenome to wild-type MHV at a frequency of about 1% was demonstrated.

Targeted recombination has been applied to the generation of mutants in most of the coronavirus genes. Thus, two silent mutations have been created so far in gene 1a [38]. The S protein has also been modified by targeted recombination. Changes were introduced by one crossover event at the 5' end of the S gene that modified MHV pathogenicity [39]. Targeted recombination mediated by two cross-overs allowed the replacement of the S gene of a respiratory strain of TGEV by the S gene of enteric TGEV strain PUR-C11 leading to the isolation of viruses with a modified tropism and virulence [7]. In this case the recombinants were selected *in vivo* using their new tropism in piglets. A new strategy for the selection of recombinants within the S gene, after promoting targeting recombination, was based on elimination of the parental replicative TGEV by the simultaneous neutralization with two mAbs (I. Sola and L. Enjuanes, unpublished results). Mutations have also been introduced by targeted mutagenesis within the E and M genes. These mutants provided corroboration for the pivotal role of E protein in coronavirus assembly and identified the carboxyl terminus of the M molecule as crucial to assembly [40].

Targeted recombination was also used to express heterologous genes. For instance, the gene encoding GFP was inserted into MHV between genes S and E, resulting in the creation of the largest known RNA viral genome [41].

An infectious MHV cDNA clone has recently been assembled *in vitro*. A method similar to the one developed to assemble an infectious TGEV cDNA clone based on the *in vitro* ligation of seven contiguous cDNA subclones has been applied to the construction of a cDNA that spanned the 31.5 kb of the MHV A59 strain [12]. The ends of the cDNAs were engineered with unique junctions, which were directed to assembly with only the adjacent cDNAs subclone, resulting in an intact MHV-A59 cDNA construct. The interconnecting restriction site junctions that are located at the ends of each cDNA are systematically removed during the assembly of the complete full-length cDNA product, allowing reassembly without the introduction of nucleotide changes. RNA transcripts derived from the full-length MHV-A59 construct was infectious, although virus recovery was enhanced 10–15-fold in the presence of RNA transcripts encoding the nucleocapsid protein, N.

5.3. Group 3 coronaviruses

The infectious IBV cDNA clone was assembled using the same strategy reported for HCoV-229E with some modifications [13]. Similarly to HCoV-229E, the IBV genomic cDNA was assembled downstream of the T7 promoter by *in vitro* ligation and cloned into the vaccinia virus DNA. However, recovery of recombinant IBV was done after the *in situ* synthesis of infectious IBV RNA by transfection of restricted recombinant vaccinia virus DNA (containing the IBV genome) into primary chick

kidney cells previously infected with a recombinant fowlpox expressing T7 RNA polymerase.

Engineered cDNAs are having an important impact on the study of mechanisms of coronavirus replication and transcription and provide an invaluable tool for the experimental investigation of virus–host interactions.

5.4. Replication-competent propagation-deficient coronavirus-derived expression systems

Replication-competent propagation-deficient virus vectors based on TGEV genomes deficient in the essential gene E that are complemented in packaging cell lines have been developed [33,42]. Two types of cell lines expressing TGEV E protein have been established, one with transient expression using the non-cytopathic Sindbis virus replicon pSINrep21 (Fig. 6) and another stably expressing E gene under the CMV promoter. The rescue of recombinant TGEV deficient in the non-essential 3a and 3b genes, and the essential E gene reached high titers ($>6 \times 10^6$ pfu/ml) in cells transiently expressing the TGEV E protein, while this titer was up to 5×10^5 pfu/ml in packaging cell lines stably expressing protein E. Interestingly, virus titers were related to protein E expression levels [42]. Recovered virions showed the same morphology and stability at different pH and temperatures than the wild type virus.

A second strategy for the construction of replication-competent propagation-defective TGEV genomes expressing heterologous genes, involves the assembly of an infectious cDNA from six cDNA fragments that are ligated *in vitro* [33]. The defective virus with the essential E gene deleted was complemented by the expression of E gene using the Venezuelan equine encephalitis replicon expression vector. However, titers of recombinant TGEV- Δ E expressing the GFP were at least

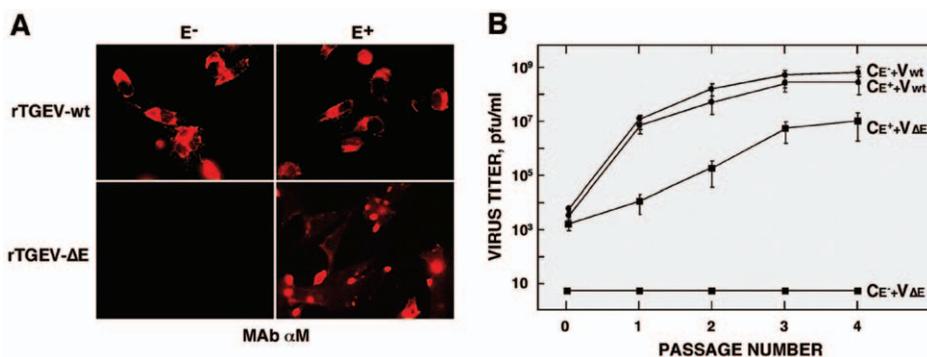


Fig. 6. Rescue of recombinant TGEV- Δ E from cDNA in cells transiently expressing E protein. (A) Analysis by immunofluorescence with M protein specific monoclonal antibodies of the rescue of full-length recombinant TGEV (rTGEV-wt) or TGEV cDNA with the E gene deleted (rTGEV- Δ E) in normal BHK-APN cells (E⁻) or cells expressing TGEV E protein (E⁺). (B) Titers through passage of recombinant TGEV rescued from cDNA in BHK-APN cells (CE⁻) or BHK-APN cells expressing TGEV E protein (CE⁺) transfected either with rTGEV-wt (V_{wt}) or with rTGEV- Δ E virus (V _{Δ E}). Error bars represent standard deviations of the mean from four experiments.

10- or 100-fold lower (around 10^4 pfu/ml) than with the system using stably transformed cells or the SIN vector to complement deletion of the E gene [42].

5.5. Cloning capacity of coronavirus expression vectors

Coronavirus minigenomes have a theoretical cloning capacity close to 27 kb, since their RNA with a size of about 3 kb is efficiently amplified and packaged by the helper virus and the virus genome has about 30 kb. In contrast, the theoretical cloning capacity for an expression system based on a single coronavirus genome like TGEV according to current available knowledge is between 3 and 3.5 kb taking into account that: (i) the non-essential genes 3a (0.2 kb), 3b (0.73 kb), and most of gene 7 [43] have been deleted leading to a viable virus; (ii) the standard S gene can be replaced by the S gene of a porcine respiratory coronavirus (PRCV) mutant with a deletion of 0.67 kb; and (iii) both DNA and RNA viruses may accept genomes with sizes up to 105% of the wild type genome. This cloning capacity will probably be enlarged by deleting non-essential domains of the replicase gene. These domains are being identified by comparing the arterivirus replicase gene (i.e., for EAV 9.7×10^3 nt) and that of coronavirus (i.e., for TGEV 20.3×10^3 nt) [19]. Differences in size between these two replicase genes could correspond to non-essential domains in the coronavirus replicase that may be dispensable.

6. Optimization of transcription levels

To optimize expression levels it is essential to improve virus vector replication levels without increasing virulence, to optimize the accumulation of total mRNA levels, and to improve mRNA translation. These results can only be achieved by determining the mechanism involved in these processes. A brief review of the mechanism of mRNA transcription in coronavirus and arterivirus is described to help achieve this goal.

Coronavirus RNA synthesis occurs in the cytoplasm via a negative-strand RNA intermediate that contains short stretches of oligo(U) at the 5' end. Both genome-size and subgenomic negative-strand RNAs, which correspond in number of species and size to those of the virus-specific mRNAs have been detected [44,45]. Coronavirus mRNAs have a leader sequence at their 5' ends. At the start site of every transcription unit on the viral genomic RNA, there is a TRS that includes a highly conserved core sequence (CS) that is nearly homologous to the 3' end of the leader RNA. This sequence constitutes part of the signal for sgRNA transcription. The common 5' leader sequence is only found at the very 5' terminus of the genome, which implies that the synthesis of sgRNAs involves fusion of non-contiguous sequences. The mechanism involved in this process is under debate. Nevertheless, the *discontinuous transcription during negative-strand RNA synthesis* model is compatible with most of the experimental evidence [45–47]. Because the leader–mRNA junction occurs during the synthesis of the negative strand within the sequence complementary to the CS (cCS) the nature of the CS is considered crucial for

mRNA synthesis. Transcription levels may be influenced by many factors. The three that we [21] consider most relevant are: (i) potential base pairing between the leader 3' end and sequences complementary to the TRS located at the 5' end of each nidovirus gene (cTRS), that guide the fusion between the nascent negative strand and the leader TRS. A minimum complementarity is needed between the leader-TRS and the cTRSs of each gene. Extension of this complementarity increases mRNA synthesis up to a certain extent, beyond a certain extension addition of 5' or 3' CS flanking sequences does not help transcription [21,48,49]; (ii) proximity of a gene to the 3'. Since the TRSs act as signals to slow down or stop the replicase complex, the smaller mRNAs should be the most abundant. Although this has been shown to be the case in the *Mononegavirales* [50] and in coronaviruses shorter mRNAs are in general more abundant, the relative abundance of coronavirus mRNAs is not strictly related to their proximity to the 3' end [21,51]; and (iii) potential interaction of proteins with the TRSs RNA, and protein–protein interactions that could regulate transcription levels. The reassociation of the nascent RNA chain with the leader TRS is probably mediated by approximation of the leader TRS through RNA–protein and protein–protein interactions.

6.1. Characteristics of the TRS

The three factors implicated in the control of mRNA abundance assume a key role for the TRS. Hence, in order to engineer vectors with high expression levels, it seems relevant to define the characteristics of the TRS, including the size of the 5' and 3' TRS sequences flanking the CS. The CS of coronaviruses belonging to groups I (hexamer 5'-CUAAAC-3') and II (heptamer 5'-UCUAAAC-3') share homology, whereas the CS of coronaviruses belonging to group III, like that of IBV, have the most divergent sequence (5'-CUUAACAA-3'). Also, arterivirus CSs have a sequence (5'-UCAACU-3') that partially resembles that of IBV. Thus, the CSs of different coronaviruses are quite similar though slightly different in length. This CS is essential for mRNA synthesis, and can be considered to be a defined domain in the TRS because it is particularly conserved within a nidovirus family, while the flanking sequences, both at the 5' (5' TRSs) and at the 3' (3' TRSs) have a unique composition for each gene even within the same virus.

The influence of the CS in transcription has been analyzed in detail in the arteriviruses [46,47]. Using an infectious cDNA clone of EAV it has been shown that sgRNA synthesis requires base pairing interaction between the leader TRS and cTRS in the viral negative strand. The construction of double mutants in which a mutant leader CS was combined with the corresponding mutant RNA7 body CS, resulting in the specific restoration of mRNA7 synthesis, suggested that the sequence of the CS *per se* is not crucial, as long as the possibility for CS base pairing is maintained. Nevertheless, it has been shown that other factors, besides leader–body base pairing, also play a role in sgRNA synthesis and that the primary sequence (or secondary structure) of TRSs may dictate strong base preferences at certain positions [46]. In addition, detailed analysis of the TRS used in the arteriviruses [47], MHV [52], BCoV [53], and TGEV [31] indicate that non-canonical CS sequences

may also be used for the switch during the discontinuous synthesis of the negative strand during transcription in the *Nidovirales*.

The promotion of transcription from a given CS is also a function of the CS flanking sequences. Data from different laboratories working with different nidoviruses have shown that CS flanking sequences can critically influence the strength of a given fusion site [21,48,49,53–55]. Although approximations to the definition of the TRS have been made, the precise length of the TRS requires further work to optimize accumulation of mRNA levels.

6.2. *Effect of CS copy number on transcription*

Studies on coronavirus transcription were performed using more than one CS to express the same mRNA. The accumulated amounts of sgRNA remained nearly the same for constructs with one to three CSs, and transcription preferentially occurred at the 3'-most TRS [29,56–58]. This observation is consistent with the coronavirus discontinuous transcription during the negative-strand synthesis model [59].

7. *Modification of coronavirus tropism and virulence*

Driving vector expression to different tissues may be highly convenient in order to preferentially induce a specific type of immune response, i.e., mucosal immunity by targeting the expression to gut-associated lymph nodes. In addition, it seems useful to change the species specificity of the vector to expand its use. Both tissue- and species-specificity have been modified using coronavirus genomes.

Group 1 coronaviruses attach to host cells through the S glycoprotein by interactions with aminopeptidase N (APN) which is the cellular receptor [60,61]. Group 2 coronaviruses use the carcinoembryonic antigen-related cell adhesion molecules (CEACAM) as receptors. Engineering the S gene can lead to changes both in the tissue- and species-specificity [7,8].

Tropism change in general leads to a change in virulence. Certainly this is the case in porcine coronavirus with a virulence directly related to its ability to grow in the enteric tract [7]. Gene expression among the non-segmented negative-stranded RNA viruses is controlled by the highly conserved order of genes relative to the single transcriptional promoter. Rearrangement of the genes of vesicular stomatitis virus eliminates clinical disease in the natural host and is considered a new strategy for vaccine development [62]. In coronavirus, genes closer to the 3' end are in general expressed more abundantly than 3' end distal ones and, in principle, gene order change can also lead to virus attenuation (P. Rottier, personal communication).

8. *Expression systems based on arteriviruses*

The *Arteriviridae* include four members: EAV, PRRSV, SHFV and lactate dehydrogenase-elevating virus of mice (LDHV). Defective genomes of EAV have

been isolated and used to express a reporter gene (CAT) in cell culture [35]. More interestingly, infectious cDNA clones have been obtained for EAV [14], PRRSV [14,63] and SHFV [16] creating the possibility of specifically altering their genomes for vector development and vaccine production. To insert genes in different positions a unique restriction endonuclease site has been introduced between consecutive EAV genes [63]. The viruses recovered expressed epitopes of nine amino acids from MHV within the ectodomain of the membrane (M) protein for at least three passages [35]. Foreign epitopes have also been expressed by using PRRSV vectors [64].

9. Conclusions

Both helper-dependent expression systems, based on two components, and single genomes constructed by targeted recombination, or by using infectious cDNAs, have been developed for coronaviruses. The sequences that regulate transcription have been characterized mainly using helper-dependent expression systems. These expression systems have the advantage of their large cloning capacity, in principle higher than 27 kb, produce reasonable amounts of heterologous antigens ($2\text{--}8\ \mu\text{g}/10^6$ cells), show a limited stability (synthesis of heterologous gene is maintained for around 10 passages), and elicit strong immune responses. In contrast, coronavirus vectors based on single genomes have at present a limited cloning capacity (3–3.5 kb), expression levels of heterologous genes are 10-fold over those of helper dependent systems ($> 50\ \mu\text{g}/10^6$ cells) and are very stable (> 30 passages). Furthermore, replication-competent propagation-deficient expression systems based on coronavirus genomes have been developed increasing the safety of these vectors. The possibility of expressing different genes under the control of TRSs with programmable strength, and engineering the tissue and species tropism indicate that coronavirus vectors are very flexible. Thus, coronavirus-based vectors are emerging with a high potential for vaccine development and, possibly, for gene therapy.

Acknowledgements

This work has been supported by grants from the Comisión Interministerial de Ciencia y Tecnología (CICYT), La Consejería de Educación y Cultura de la Comunidad de Madrid, and Fort Dodge Veterinaria from Spain, and the European Communities (Life Sciences Program, Key Action 2: Infectious Diseases).

Abbreviations

APN	aminopeptidase N
BAC	bacterial artificial chromosome
BCoV	bovine coronavirus
CAT	chloramphenicol acetyltransferase

cCS	sequence complementary to CS
CEACAM	carcinoembryonic antigen-related cell adhesion molecules
CMV	cytomegalovirus promoter
CS	conserved core sequence
cTRS	sequences complementary to TRS
EAV	equine arteritis virus
GFP	green fluorescent protein
GUS	β -glucuronidase
HCoV	human coronavirus
IBV	infectious bronchitis virus
LDHV	lactate dehydrogenase-elevating virus
MHV	mouse hepatitis virus
PRCV	porcine respiratory coronavirus
PRRSV	porcine respiratory and reproductive syndrome virus
sgmRNA	subgenomic mRNA
SHFV	simian hemorrhagic fever virus
SIN	vector based on Sindbis virus replicon
TGEV	transmissible gastroenteritis coronavirus
TRS	transcription-regulating sequences
UTR	untranslated region

References

1. Enjuanes, L. (2000) In: van Regenmortel, M.H.V. *Virus Taxonomy. Classification and Nomenclature of Viruses*, Academic Press, New York, pp. 835–849.
2. Enjuanes, L., Spaan, W., Snijder, E., and Cavanagh, D. (2000) In: van Regenmortel, M.H.V. (eds.) *Virus Taxonomy. Classification and Nomenclature of Viruses*, Academic Press, New York, pp. 827–834.
3. Cowley, J.A., Dimmock, C.M., Spann, K.M., and Walker, P.J. (2000) *J. Gen. Virol.* 81, 1473–1484.
4. Mayo, M.A. (2002) *Arch. Virol.* 147, 1655–1656.
5. Lai, M.M.C. and Cavanagh, D. (1997) *Adv. Virus Res.* 48, 1–100.
6. Masters, P.S. (1999) *Adv. Virus Res.* 53, 245–264.
7. Sánchez, C.M., Izeta, A., Sánchez-Morgado, J.M., Alonso, S., Sola, I., Balasch, M., Plana-Durán, J., and Enjuanes, L. (1999) *J. Virol.* 73, 7607–7618.
8. Kuo, L., Godeke, G.-J., Raamsman, M.J.B., Masters, P.S., and Rottier, P.J.M. (2000) *J. Virol.* 74, 1393–1406.
9. Almazán, F., González, J.M., Pénczes, Z., Izeta, A., Calvo, E., Plana-Durán, J., and Enjuanes, L. (2000) *Proc. Natl. Acad. Sci. USA* 97, 5516–5521.
10. Yount, B., Curtis, K.M., and Baric, R.S. (2000) *J. Virol.* 74, 10600–10611.
11. Thiel, V., Herold, J., Schelle, B., and Siddell, S. (2001) *J. Gen. Virol.* 82, 1273–1281.
12. Yount, B., Denison, M.R., Weiss, S.R., and Baric, R.S. (2002) *J. Virol.* 76, 11065–11078.
13. Casais, R., Thiel, V., Siddell, S.G., Cavanagh, D., and Britton, P. (2001) *J. Virol.* 75, 12359–12369.
14. van Dinten, L.C., den Boon, J.A., Wassenaar, A.L.M., Spaan, W.J.M., and Snijder, E.J. (1997) *Proc. Natl. Acad. Sci. USA* 94, 991–996.
15. Meulenberg, J.J.M., Bos-de-Ruijter, J.N.A., Wensvoort, G., and Moormann, R.J.M. (1998) *Adv. Exp. Med. Biol.* 440, 199–206.
16. Godeny, E.K., Zeng, L., Smith, S.L., and Brinton, M.A. (1995) *J. Virol.* 69, 2679–2683.
17. Escors, D., Ortego, J., Laude, H., and Enjuanes, L. (2001) *J. Virol.* 75, 1312–1324.

18. Escors, D., Camafeita, E., Ortego, J., Laude, H., and Enjuanes, L. (2001) *J. Virol.* 75, 12228–12240.
19. Gorbalenya, A.E. (2001) In: Lavi, E., Weiss, S. and Hingley, S.T. (eds.) *The Nidoviruses (Coronaviruses and Arteriviruses)*, Vol. 494, Kluwer Academic/Plenum Publishers, New York, pp. 1–17.
20. Izeta, A., Smerdou, C., Alonso, S., Penzes, Z., Méndez, A., Plana-Durán, J., and Enjuanes, L. (1999) *J. Virol.* 73, 1535–1545.
21. Alonso, S., Izeta, A., Sola, I., and Enjuanes, L. (2002) *J. Virol.* 76, 1293–1308.
22. Alonso, S., Sola, I., Teifke, J., Reimann, I., Izeta, A., Balach, M., Plana-Durán, J., Moormann, R.J.M., and Enjuanes, L. (2002) *J. Gen. Virol.* 83, 567–579.
23. Thiel, V., Siddell, S.G., and Herold, J. (1998) *Adv. Exp. Med. Biol.* 440, 109–114.
24. Liao, C.L., Zhang, X., and Lai, M.M.C. (1995) *Virology* 208, 319–327.
25. Zhang, X., Hinton, D.R., Cua, D.J., Stohlman, S.A., and Lai, M.M.C. (1997) *Virology* 233, 327–338.
26. Liao, C.-L. and Lai, M.M.C. (1995) *Virology* 209, 428–436.
27. Zhang, X., Hinton, D.R., Park, S., Parra, B., Liao, C.-L., and Lai, M.M.C. (1998) *Virology* 242, 170–183.
28. Lai, M.M.C., Zhang, X., Hinton, D., and Stohlman, S. (1997) *J. Neurovirol.* 3, S33–S34.
29. Stirrups, K., Shaw, K., Evans, S., Dalton, K., Casais, R., Cavanagh, D., and Britton, P. (2000) *J. Gen. Virol.* 81, 1687–1698.
30. González, J.M., Penzes, Z., Almazán, F., Calvo, E., and Enjuanes, L. (2002) *J. Virol.* 76, 4655–4661.
31. Sola, I., Alonso, S., Plana-Durán, J., and Enjuanes, L. (2003) *J. Virol.* 77, 4357–4369.
32. Frolov, I., Hoffman, T.A., Prágai, B.M., Dryga, S.A., Huang, H.V., Schlesinger, S., and Rice, C.M. (1996) *Proc. Natl. Acad. Sci. USA* 93, 11371–11377.
33. Curtis, K.M., Yount, B., and Baric, R.S. (2002) *J. Virol.* 76, 1422–1434.
34. Thiel, V., Herold, J., Schelle, B., and Siddell, S.G. (2001) *J. Virol.* 75, 6676–6681.
35. Molenkamp, R., van Tol, H., Rozier, B.C., van der Meer, Y., Spaan, W.J., and Snijder, E.J. (2000) *J. Gen. Virol.* 81, 2491–2496.
36. Curtis, K.M., Yount, B., and Baric, R.S. (2002) *J. Virol.* 76, 1422–1434.
37. Koetzner, C.A., Parker, M.M., Ricard, C.S., Sturman, L.S., and Masters, P.S. (1992) *J. Virol.* 66, 1841–1848.
38. van der Most, R.G., Heijnen, L., Spaan, W.J.M., and de Groot, R.J. (1992) *Nucleic Acids Res.* 20, 3375–3381.
39. Leparco-Goffart, I., Hingley, S.T., Chua, M.M., Phillips, J., Lavi, E., and Weiss, S.R. (1998) *J. Virol.* 72, 9628–9636.
40. de Haan, C.A.M., Kuo, L., Masters, P.S., Vennema, H., and Rottier, P.J.M. (1998) *J. Virol.* 72, 6838–6850.
41. Fischer, F., Stegen, C.F., Koetzner, C.A., and Masters, P.S. (1997) *J. Virol.* 71, 5148–5160.
42. Ortego, J., Escors, D., Laude, H., and Enjuanes, L. (2002) *J. Virol.* 76, 11518–11529.
43. Ortego, J., Sola, I., Almazán, F., Ceriani, J.E., Riquelme, C., Balach, M., Plana-Durán, J., and Enjuanes, L. (2003) *Virology*, 308, 13–22.
44. Brian, D.A. (2001) In: Lavi, E., Weiss, S. and Hingley, S.T. (eds.) *Nidoviruses*, Vol. 494, Plenum Press, New York, pp. 415–428.
45. Sawicki, D.L., Wang, T., and Sawicki, S.G. (2001) *J. Gen. Virol.* 82, 386–396.
46. Pasternak, A.O., van den Born, E., Spaan, W.J.M., and Snijder, E.J. (2001) *EMBO J.* 20, 7220–7228.
47. van Marle, G., Dobbe, J.C., Gultyaev, A.P., Luytjes, W., Spaan, W.J.M., and Snijder, E.J. (1999) *Proc. Natl. Acad. Sci. USA* 96, 12056–12061.
48. Makino, S. and Joo, M. (1993) *J. Virol.* 67, 3304–3311.
49. van der Most, R.G., De Groot, R.J., and Spaan, W.J.M. (1994) *J. Virol.* 68, 3656–3666.
50. Wertz, G.W., Perepelitsa, V.P., and Ball, L.A. (1998) *Proc. Natl. Acad. Sci. USA* 95, 3501–3506.
51. Penzes, Z., González, J.M., Calvo, E., Izeta, A., Smerdou, C., Méndez, A., Sánchez, C.M., Sola, I., Almazán, F., Enjuanes, L. (2001) *Virus Genes* 23, 105–118.
52. Zhang, X. and Liu, R. (2000) *Virology* 278, 75–85.
53. Ozdarendeli, A., Ku, S., Rochat, S., Senanayake, S.D., and Brian, D.A. (2001) *J. Virol.* 75, 7362–7374.
54. An, S. and Makino, S. (1998) *Virology* 243, 198–207.

55. Hsue, B. and Masters, P.S. (1999) *J. Virol.* 73, 6128–6135.
56. Joo, M. and Makino, S. (1995) *J. Virol.* 69, 272–280.
57. Krishnan, R., Chang, R.Y., and Brian, D.A. (1996) *Virology* 218, 400–405.
58. van Marle, G., Luytjes, W., Van der Most, R.G., van der Straaten, T., and Spaan, W.J.M. (1995) *J. Virol.* 69, 7851–7856.
59. Sawicki, S.G. and Sawicki, D.L. (1990) *J. Virol.* 64, 1050–1056.
60. Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L.K., Norén, O., and Laude, H. (1992) *Nature* 357, 417–420.
61. Yeager, C.L., Ashmun, R.A., Williams, R.K., Cardellichio, C.B., Shapiro, L.H., Look, A.T., and Holmes, K.V. (1992) *Nature* 357, 420–422.
62. Flanagan, E.B., Zamparo, J.M., Ball, L.A., Rodriguez, L., and Wertz, G.W. (2001) *J. Virol.* 75, 6107–6114.
63. de Vries, A.A.F., Glaser, A.L., Raamsman, M.J.B., de Haan, C.A.M., Sarnataro, S., Godeke, G.J., and Rottier, P.J.M. (2000) *Virology* 270, 84–97.
64. Groot Bramel-Verheije, M.H., Rottier, P.J.M., and Meulenber, J.J.M. (2000) *Virology* 278, 380–389.