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

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

ADVANCES IN VIRUS RESEARCH, VOL. 53

# REVERSE GENETICS OF THE LARGEST RNA VIRUSES

Paul S. Masters

Wadsworth Center for Laboratories and Research
New York State Department of Health
and Department of Biomedical Sciences
State University of New York at Albany
Albany, New York 12201

I. Introduction: Unique Aspects of Coronavirus Molecular Biology
II. Targeted RNA Recombination
A. Development of System
B. Genetic Analysis of Coronavirus Structural Proteins
C. Genetic Analysis of Coronavirus RNA Synthesis
D. Limitations of Targeted Recombination
III. Conclusions and Future Prospects
References

## I. INTRODUCTION: UNIQUE ASPECTS OF CORONAVIRUS MOLECULAR BIOLOGY

Coronaviruses are positive-sense, nonsegmented RNA viruses that infect numerous mammalian and avian species (Siddell, 1995). They form one of the two genera within the family *Coronaviridae,* which, together with the family *Arteriviridae,* make up the newly designated order *Nidovirales* (reviewed in de Vries *et al.,* 1997). The capped and polyadenylated genomes of coronaviruses, spanning some 27 to 31 kb, are the largest of all RNA virus genomes, including those of the segmented RNA viruses. Less parochially, coronavirus genomes are among the largest mature RNA molecules known to biology, so far exceeded only by Balbiani ring mRNA (Mehlin and Daneholt, 1993) and the mRNA for titin (Labeit and Kolmerer, 1995), both of which are informationally less complex, encoding highly repetitive proteins.

Like all other positive-sense RNA viruses (retroviruses excluded), coronavirus genomic RNA is infectious when transfected into cells of a permissive host. Therefore, in principle, the most direct way to perform reverse genetics on a coronavirus ought to involve the construction of a full-length genomic cDNA clone from which infectious RNA could be transcribed *in vitro.* To date this objective has not been attained. Of the many possible hindrances to the development of an error-free full-length clone (detailed in Boyer and Haenni, 1994), the most formidable in this case is genome size. For coronavirus genomes, which are nearly



Copyright © 1999 by Academic Press.
All rights of reproduction in any form reserved.
0065-3527/99 $30.00

three times the size of alphavirus genomes or four times the size of picornavirus genomes, the chances are proportionately increased that any individual full-length cDNA will harbor one or more cryptic lethal mutations. These might be generated by the viral RNA-dependent RNA polymerase itself, by the reverse transcriptase (RT) used for first-strand cDNA synthesis, or by the DNA polymerase used for second-strand cDNA synthesis or PCR (polymerase chain reaction) amplification. Identification of such mutations might also be difficult because published cDNA sequences generally have been obtained from clones subject to the same sources of error. An additional problem on the path to a full-length coronavirus clone arises from the experience of some investigators that certain large genomic subclones are unstable when propagated in *Escherichia coli.* This raises the prospect that a full-length clone will be similarly afflicted, as has been observed for clones of yellow fever virus (Rice *et al.,* 1989) and hepatitis C virus (Forns *et al.,* 1997). Although these and other potential obstacles should not be insurmountable, they have until now diminished the enthusiasm of individual research groups to invest much effort in the construction of a full-length coronavirus clone. Nevertheless, the continued development of improved techniques, particularly long RT-PCR (Thiel *et al.,* 1997), makes it likely that such a goal will eventually be achieved. Noteworthy in this regard, important milestones have been reached with the infectious full-length clones of the arteriviruses equine arteritis virus (van Dinten *et al.,* 1997) and Lelystad virus (Meulenberg *et al.,* 1998), which are 12.7 and 15.2 kb, respectively.

Because coronaviruses currently remain a class apart from all other positive-sense RNA viruses, an alternative method has had to be devised for their genetic manipulation. This method, targeted recombination, is less direct and more laborious, and so far it has been applied exclusively to site-directed mutagenesis of mouse hepatitis virus (MHV). Consequently, all discussion in this review will be limited to this viral prototype. MHV causes a number of different diseases in mice, depending on viral strain as well as host genetic makeup and immune status (Compton *et al.,* 1993). It thus provides very useful models for studies of viral pathogenesis. Because it also grows well in tissue culture, which can be problematic for other members of the coronavirus family, the molecular biology of MHV has been studied intensively.

MHV virions contain four essential structural proteins (Fig. 1), which assemble and bud into the intermediate compartment between the endoplasmic reticulum and the Golgi apparatus. The most prominent of these, the spike glycoprotein (S), forms oligomers that protrude

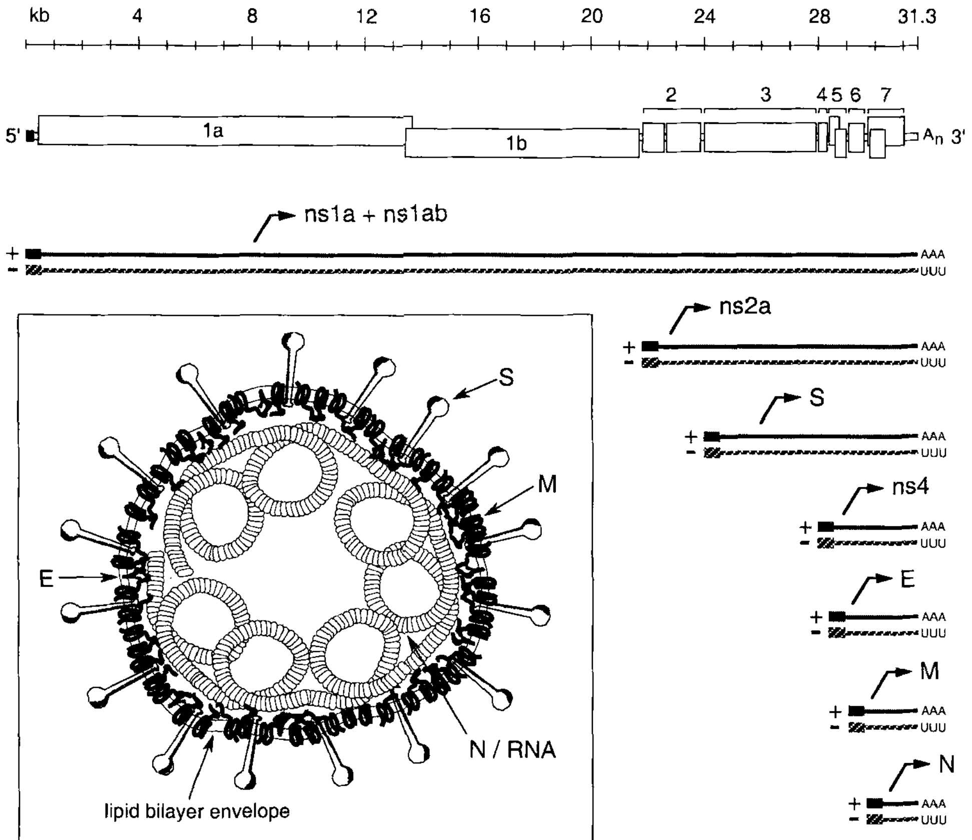


FIG 1. Coronavirus genomic organization, mRNA expression, and virion structure. The arrangement of genes in the prototype MHV (strain A59) is represented at the top. Beneath this is the set of positive- and negative-sense RNA species synthesized in infected cells. The protein products obtained from each positive-sense RNA are indicated. Two products, polyproteins, ns1a and ns1ab, are translated from genomic RNA by means of a ribosomal frameshifting mechanism. All other proteins are translated from the first open reading frame of each functionally monocistronic subgenomic RNA. The inset shows a schematic of a virion containing the minimal set of essential structural proteins.

from the virion membrane envelope and attach to species- and tissue-specific host cell receptors. The most abundant virion protein, the membrane glycoprotein (M), spans the envelope three times and has a large carboxy-terminal tail that resides in the interior of the virion. A very small membrane protein (E), present in small amounts, is the third constituent of the envelope. Finally, within the envelope, the nucleocap-

sid protein (N) wraps the RNA genome into a helically symmetric filament, which, as suggested by some work, may be further organized into an icosahedral shell (Risco *et al.,* 1996). Two additional structural proteins appear in many strains of MHV, but their absence from some strains or mutants indicates that they are nonessential. The first of these is a hemagglutinin esterase (HE), which has apparently evolved from an ancestor shared with the hemagglutinin of influenza C virus (Luytjes *et al.,* 1988). The other is the largely hydrophobic protein product (I) of the internal open reading frame of the N gene (Fischer *et al.,* 1997a).

This relative paucity of structural proteins accounts for less than one-third of the coding capacity of the MHV genome (Fig. 1). The bulk of the genome encodes a huge nonstructural polyprotein (ns1a and ns1ab), the proteolytic products of which include the RNA-dependent RNA polymerase, proteinases, a helicase, and numerous polypeptides of unknown function. These proteins constitute the machinery of viral RNA replication and transcription, and it is thought that most of them form a large complex associated with an intracellular membranous compartment (Brown and Brierley, 1995).

Coronavirus RNA synthesis at first glance appears to be a strange amalgamation of synthetic schemes employed by disparate families of RNA viruses (for reviews, see van der Most and Spaan, 1995; Lai *et al.,* 1994). Following its translation into the products of gene 1, the genomic RNA becomes the template for synthesis of at least a full-length (negative-sense) antigenome (Fig. 1). Further events produce a series of smaller RNAs of both polarities. The positive-sense subgenomic RNAs, each of which serves as the message for one of the downstream encoded proteins, have compositions equivalent to those of large genomic deletions. Each contains a 70-nt leader RNA, identical to the 5′ end of the genome, joined at a downstream intergenic region to the 3′ end of the genome. The negative-sense subgenomic RNAs, roughly one-tenth as abundant, each have the complement of this arrangement, including a 5′ oligo(U) tract and a 3′ antileader (Hofmann and Brian, 1991; Sethna *et al.,* 1991). It should be noted that, of all these RNA species, only the full-length positive-strand genome is ultimately packaged into virions. The mechanism by which the subgenomic RNAs are produced is still an open question. About the only point of consensus is that, contrary to what one might have initially assumed from their structure, the subgenomic RNAs are *not* produced by splicing of a full-length precursor. Another generally agreed-on notion is that simple base-pairing interactions alone cannot account for the mechanism of leader-to-body fusion for the subgenomic RNAs. Fusion is probably the

result of quasi-continuous synthesis across two distant portions of a looped-out template that are brought together via protein–protein interactions (Zhang *et al.,* 1994). Two major unresolved issues remain, however. The first is whether the leader-to-body fusion event occurs during positive-strand or negative-strand RNA synthesis. The second is whether the subgenomic negative-strand RNA species are active templates for the amplification of positive-strand subgenomic RNAs or, conversely, whether they represent dead-end products.

An additional distinctive feature of coronavirus RNA synthesis, crucial to the subject of this review, is the high rate of homologous RNA–RNA recombination that occurs during the course of infection (Lai, 1996). This is thought to result from a template-switching mechanism similar to that shown to operate for poliovirus (Kirkegaard and Baltimore, 1986). On a fine scale, MHV recombination seems to be random (Banner and Lai, 1991), although selective pressures can create local biases (Banner *et al.,* 1990). Over the whole genome, the rate of MHV recombination has been estimated to be 1% per 1.3 kb (Baric *et al.,* 1990).

## II. Targeted RNA Recombination

### A. *Development of System*

Given that coinfecting coronavirus genomes can exchange genetic information, it seemed likely that this property could be exploited to introduce specific mutations into MHV. Toward this end, Koetzner *et al.* (1992) initially reported the engineering of a 5-nt insertion into the 3′ untranslated region (3′ UTR) of MHV via targeted recombination with an *in vitro* synthesized RNA. This experiment was facilitated by the finding of an N gene mutant, designated Alb4, that was both temperature sensitive and thermolabile owing to the deletion of a 29-amino-acid linker between two domains of the N protein. Although this mutation is not completely lethal at the nonpermissive temperature (39°C), Alb4 forms tiny plaques at this temperature that are easily distinguishable from wild-type plaques. Moreover, incubation of Alb4 virions at 40°C for 24 hours results in a 100-fold greater loss of titer than for wild-type virions (Koetzner *et al.,* 1992). These phenotypic traits allowed the selection of recombinant viruses 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 (Fig. 2). Progeny of the cotransfection were subjected to heat treatment, and recombinants were chosen as large-plaque survi-

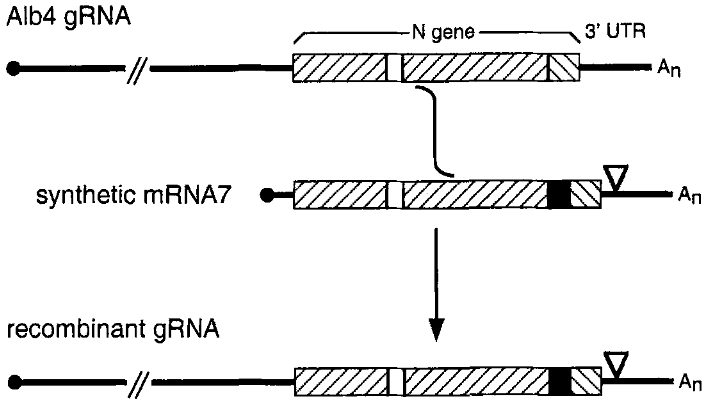


FIG 2. Basic scheme of targeted recombination in MHV. Rectangles indicate different segments of the N gene; the 87-nt region that is deleted in the Alb4 mutant is shown in black. The triangle represents a 5-nt insertion created as a marker in the 3′ UTR.

vors at 39°C. Analysis of the purified genomes of these viruses, by RT-PCR and by direct RNA sequencing, showed that recombination had restored the material that had been deleted in Alb4 and that the unselected 5-nt insertion was present in the 3′ UTR. These results were thus consistent with the generation of recombinants by a single crossover event between the recipient genomic RNA and the donor synthetic RNA occurring at any point within the N gene upstream of the Alb4 deletion (Fig. 2).

Shortly thereafter, van der Most *et al.* (1992) reported the isolation of site-specific mutants of MHV obtained with the use of derivatives of a cloned defective interfering (DI) RNA as the donor species. Coronavirus DI RNAs are extensively deleted versions of the genomic RNA that have retained signals sufficient for their own replication when provided with the necessary enzymatic activities by coinfecting helper virus (Brian and Spaan, 1997). The DI RNA employed in this study, MIDI-C, contains a fusion of three segments of the MHV genome: 3.8 kb of the 5′ end of the genome, which includes part of gene 1a; a 0.8-kb fragment of gene 1b; and 0.8 kb of the 3′ end of the genome, which includes the distal third of the N gene (Fig. 3). A derivative of MIDI-C was used to restore the material that is deleted in Alb4, simultaneously marking it with a coding-silent point mutation, and recombinants were selected by a method similar to that described above. There was a notable and important distinction from the other work, however. Whereas Koetzner *et al.* (1992) had estimated a fre-

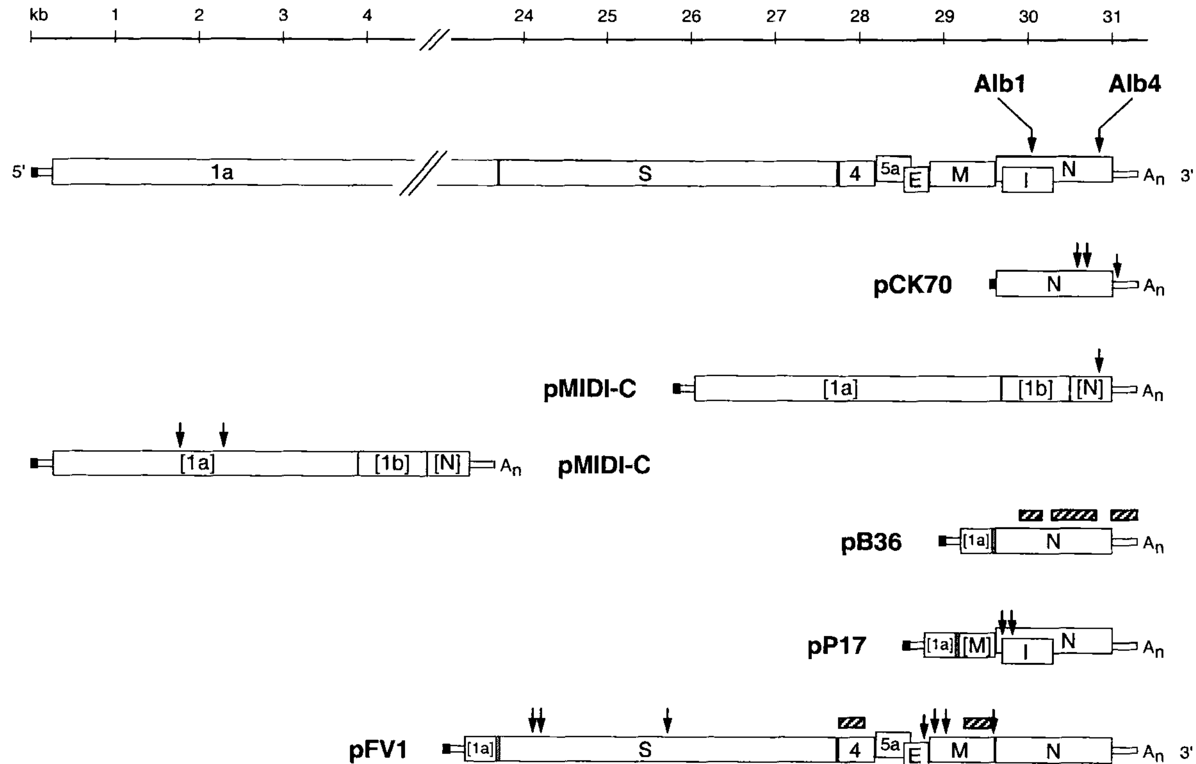


FIG 3. Targeted recombination vectors and recipient viruses. At the top, the composition and scale of the complete MHV genome are represented and the loci of the mutations in recipient viruses Alb1 and Alb4 are denoted. Shown beneath this are the names of parent plasmids used as transcription vectors and the compositions of the donor RNAs they encode aligned with the MHV genome. Brackets denote fragments of full-length genes. Arrows and hatched rectangles above the donor RNAs indicate the loci of point mutations and large substitutions that have been transferred to isolated recombinant viruses.

quency of recombination on the order of $10^{-5}$ with subgenomic RNA7 as the donor, the use of the DI donor RNA yielded recombinants at a rate some three orders of magnitude higher. This higher efficiency made it feasible to screen for recombinants in the absence of selection. In this manner, van der Most *et al.* (1992) demonstrated the transfer of coding-silent mutations in gene 1a of MIDI-C to wild-type MHV at a frequency of about 1%. To date, these remain the only engineered mutations in gene 1 of a coronavirus.

Subsequently, Masters *et al.* (1994) directly compared the rates of targeted recombination with RNA7 versus DI donor RNAs and noted frequencies with the latter comparable to those that had been observed with MIDI-C RNA. For this purpose a vector template for synthesis of a pseudo-DI RNA, pB36 (Fig. 3), was constructed in imitation of a naturally occurring DI RNA of bovine coronavirus (BCV) (Chang *et al.*, 1994). The frequency of recombination obtained with pB36 RNA was sufficiently high to allow simplification of the protocol for constructing MHV mutants from the Alb4 recipient virus. An infection–transfection procedure replaced cotransfections, which had required purification of substantial quantities of Alb4 genomic RNA, and in many cases large plaque-forming recombinants could be detected directly among progeny without a heat treatment to counterselect the Alb4 parent.

### *B. Genetic Analysis of Coronavirus Structural Proteins*

In more recent work, ever-larger pseudo-DI RNA vectors have been constructed to extend the reach of the technique to encompass all of the structural protein genes of MHV (Fig. 3). Although pB36 RNA was shown to be an authentically replicating DI RNA (Masters *et al.*, 1994), it has never been determined if the larger constructs are replication-competent because they serve well as donor RNAs. Moreover, one particular variant smaller construct, pBL34 RNA (Hsue and Masters, 1997), was found to *not* replicate despite being an efficient donor RNA. It was originally thought that the ability of the donor RNA to replicate was the crucial feature that led to an enhanced efficiency of recombination (Masters *et al.*, 1994), but this may not be the case. Another observation, possibly related, is that many targeted recombinants were isolated that were found to have been generated by multiple crossovers (Hsue and Masters, 1997; Peng *et al.*, 1995b; de Haan *et al.*, 1998; L. Kuo and P. S. Masters, unpublished results, 1998). These recombinants appear to occur more frequently than would be expected if each crossover were an independent event; that is, if the probability of a single

crossover is $p$, then double crossovers should occur as $p^2$ and triple crossovers as $p^3$. This suggests that the alignment of two templates is the rate-limiting event in recombination, and once this has been achieved, the barrier to multiple crossovers may be only marginally higher than that for single crossovers. DI RNAs possibly possess elements, absent from subgenomic RNAs, that mediate the colocalization and alignment of RNA synthesis templates.

The choice of the other component for targeted recombination, the recipient virus, generally has depended on the fortuitousness of classical genetics. Generally, one must first find a mutant in order to construct a mutant. For the purposes of selection, any conditional lethal mutant with a sufficiently low reversion rate ought to work well as a recipient virus. In practice, the N gene deletion mutant Alb4 has been used almost exclusively. In a few instances another N mutant, Alb1, also has proved a suitable recipient (Fischer *et al.,* 1997a; Peng *et al.,* 1995b). Alb1 is a point mutant that, like Alb4, is both temperature sensitive and thermolabile (Masters *et al.,* 1994). In contrast, an S gene mutant, Alb18 (Ricard *et al.,* 1995), thus far has not been a useful recipient, perhaps because there is significant selective pressure favoring accumulation of revertants even at the supposedly permissive temperature for this virus. As noted above, there is no absolute requirement that the recipient virus be capable of being selected against. In one of the original demonstrations of the method, van der Most *et al.* (1992) used wild-type MHV as the recipient and screened individual plaques to identify recombinants bearing markers in gene 1.

To date, the N gene has been the gene most thoroughly explored by targeted recombination, probably because it falls in the region of the genome that was the earliest accessible to the method. The first mutants in which specified coding changes were incorporated into MHV were for second-site point mutations that were found to revert the phenotype of Alb4. To obtain an understanding of the nature of these mutations in isolation, targeted recombination was used to introduce two of them back into the N gene in the absence of the Alb4 deletion (Peng *et al.,* 1995a). The resulting recombinants were phenotypically indistinguishable from the wild type, showing that the reverting mutations in isolation were not grossly deleterious to the virus, unlike the deletion mutation for which they could compensate. However, the N proteins of these mutants were found to bind more tightly to RNA than wild-type N protein did, and a more detailed analysis of one of them showed that it was at a competitive disadvantage with respect to the wild type. More extensive mutagenesis to probe important features of the N protein structure was carried out through construction of MHV mutants con-

taining chimeric N genes in which portions of the BCV N gene were substituted for their corresponding MHV sequences (Peng *et al.*, 1995b). This work identified segments of the two N proteins that, despite an evolutionary divergence of 30%, remained functionally equivalent. Notably, this included most of the RNA binding domain of the protein. Equally important, it was possible to tentatively map noninterchangeable regions of the protein that participate in species-specific protein–RNA and protein–protein interactions, thus providing foci for future genetic study. This work demonstrated that targeted recombination could be used to create extensive substitutions in the coronavirus genome, generating recombinants that could not be made otherwise between two viruses separated by a species barrier. A surprising finding was that some of the BCV-MHV chimeric N protein mutants exhibited altered pathogenicity compared to wild-type MHV. These mutant viruses, following intracerebral inoculation, were unable to exit the brain and reach the bloodstream, and this property was found to map to a region of 80 amino acids in the carboxy-terminal third of the N protein (Lavi *et al.*, 1998). Further work is underway to localize this pathogenic determinant more precisely to a subset of the 17 residue differences between BCV and MHV in this region.

Another N gene mutant was made to assess the significance of the internal open reading frame (I gene) that is embedded entirely within the first half of the N gene in the +1 reading frame (Fischer *et al.*, 1997a). The I gene encodes a largely hydrophobic protein that is expressed via a leaky-scanning translational strategy (Senanayake and Brian, 1997). To disrupt this gene its start codon was replaced, and a stop codon was introduced shortly thereafter. Both of these changes were chosen to create only coding-silent mutations in the N reading frame. These mutations were demonstrated to abolish I protein expression while not detectably affecting N protein expression; in the same experiments it was discovered that the I protein is a previously unrecognized structural protein of MHV. The resulting I− mutant was found to be viable at all temperatures tested and formed plaques only slightly smaller than those of its wild-type counterpart. The mutant also showed no gross differences with respect to the wild type in its ability to infect mice. An important consequence of the nonlethality of disruption of the I gene is that this eliminated a potential source of ambiguity in the analysis of any N gene mutants isolated or constructed in the region where N and I overlap.

Upstream of the N gene, targeted recombination has just begun to be used to gain insight into the workings of the proteins of the viral envelope. One of the most exciting developments in the study of corona-

viruses was the surprising discovery by Vennema *et al.* (1996) and Bos *et al.* (1996) that intracellular expression of just the E and M proteins resulted in the formation and extracellular release of virus-like particles (VLPs) that were morphologically identical to the envelopes of whole virions. Thus, coronavirus assembly can take place through interactions largely independent of the participation of the nucleocapsid or the spike glycoprotein. What role the E protein may be playing in this process is particularly intriguing because this component is very small and is present in minute amounts in virions or VLPs. To approach this question genetically, mutations were created in the E gene (Fischer *et al.*, 1998) by clustered charged-to-alanine mutagenesis, a method that has a high reported rate of success in producing temperature-sensitive and other conditional-lethal phenotypes (Wertman *et al.*, 1992). Of the four possible mutants to be made by this algorithm, one was apparently lethal and one had a wild-type phenotype. The other two mutants were partially temperature sensitive and were much more thermolabile than the wild type when grown at the permissive temperature. The more defective of the two mutants, when examined by negative-stained electron microscopy, exhibited morphological aberrations rarely seen in wild-type virions, including narrow, elongated shapes pinched at many points. These results provided striking corroboration for the pivotal role of E protein in coronavirus assembly established by the VLP expression studies. An additional noteworthy aspect of this work was that the two temperature-sensitive E mutants were obtained by PCR-based screening of pools of candidate recombinants rather than by selection.

In a similar fashion, some forays have been made into the M gene to probe its role in assembling the virus. As with the E gene, targeted mutagenesis of the M gene has been done mostly in the wake of meticulous studies in the VLP system, which identified the carboxyl terminus of the M molecule as crucial to assembly (de Haan *et al.*, 1998). In the expression system, deletion of as little as the final carboxy-terminal residue of M protein almost completely abolished formation of VLPs, and many amino acid substitutions at this position had a similar effect. In contrast, the incorporation of each of the same mutations into the whole virion had little or no effect on the viral phenotype. Deletion of two carboxy-terminal residues, however, was lethal to both VLPs and virions. These experiments supported a critical role for this region of the M protein in MHV assembly and indicated that additional intermolecular interactions in virions lead to an increased structural stabilization not present in VLPs. This work has also underscored the value of establishing a dialogue between the VLP expression system and the targeted recombination system. It is anticipated that future interplay

of the two methodologies will result in evidence of the functional roles in assembly of the components of the virion envelope and lead to a map of protein–protein interactions.

The remaining essential structural protein, the spike glycoprotein S, is the object of intense scrutiny by many researchers seeking to understand its central roles in receptor binding and initiation of infection through fusion with the host cell membrane. Most aspects of coronavirus pathogenesis are thought to stem from the variable selectivity of these processes or from the host immune response provoked by the S protein (Compton *et al.,* 1993; Holmes and Dveksler, 1994). Both the size of the S gene and its relative distance from the 3′ end of the genome (Figs. 1 and 3) forestalled its earlier accessibility to targeted recombination, but this boundary has now been crossed. In the course of constructing mutations in the nonstructural (and nonessential) gene 4, a recombinant was isolated that had inherited a coding-silent marker near the 5′ end of the S gene from a pFV1-derived donor RNA (Fig. 3) (Fischer *et al.,* 1997b). Separately, Zhang *et al.* (1997) reported the detection and analysis by RT-PCR of a number of different chimeric viruses in which portions of the S gene of the enterotropic MHV strain RI were recombined into the S gene of the respiratory MHV strain A59. Significantly, the donor vector contained the MHV-RI S gene inserted ectopically into a MIDI-C-related DI RNA, so that recombinant viruses had to be formed by two crossover events. This approach holds promise if a method of isolating these recombinants can be devised.

The first isolated MHV recombinants harboring engineered coding mutations in the S gene have now been obtained by Weiss and coworkers. One of these, containing a point mutation less than 0.5 kb from the 5′ end of the S gene (Q159L), was constructed to assess the effects of this single amino acid change, segregated from any other changes, on the pathogenic properties of the virus. Comparison of the Q159L mutant to its isogenic wild-type counterpart confirmed that this mutation plays an important role in the loss of hepatotropism by the virus (Leparc-Goffart *et al.,* 1998). A second mutant was constructed with a single alteration (H716D) at the proteolytic cleavage site near the center of the S protein in order to examine the role played by cleavage of S protein in cell-to-cell fusion and *in vivo* virulence. As predicted, the H716D mutation was found to prevent S protein cleavage and interfered with syncytia formation by infected cells in tissue culture, but the mutant replicated as efficiently as wild-type virus in the brains and livers of infected mice. Both of these mutants were generated with pFV1-derived donor RNAs and with Alb4 as the recipient virus. Remarkably, linkages of 10–20% were observed between the selected

N gene marker and the (unselected) S gene mutations. This bodes well for future genetic manipulation of this centrally important virion protein.

### *C. Genetic Analysis of Coronavirus RNA Synthesis*

Many questions about coronavirus RNA synthesis have been successfully approached by the use of DI RNAs. Indeed, DI RNA systems are the best models for the dissection of *cis*-acting elements whose alteration would be lethal in the context of the whole genome (van der Most and Spaan, 1995; Lai *et al.,* 1994; Brian and Spaan, 1997). There is also an entire class of questions that will require the development of a robust *in vitro* viral RNA synthesis system capable of initiating both transcription and replication. Whole-genome mutagenesis would not be expected to be useful in addressing these issues, either. There are, however, areas where viral genetics can complement DI RNA studies or reveal properties of viral RNA synthesis that are lost from DI RNAs because of their extensive deletions. An example of the latter was found serendipitously when the gene for green fluorescent protein (GFP) was inserted into MHV in place of the 5′ two-thirds of gene 4 (Fischer *et al.,* 1997b). (This, incidentally, resulted in the creation of the largest known RNA viral genome.) As anticipated, this recombinant virus produced a larger version of subgenomic RNA4 originating from the intergenic sequence upstream of the inserted GFP gene. Quite surprisingly, however, a set of 10 additional RNA species were also produced, all originating from a cluster of sites in the middle of the GFP gene centered about the point where, in distance from the 3′ end, the intergenic sequence had been prior to the insertion (Fig. 4). This suggested that long-range RNA interactions can contribute to the determination of the sites of leader-to-body fusion, even in the complete absence of a canonical intergenic sequence.

A second discovery facilitated by targeted recombination was the finding of a conserved RNA secondary structure in the 3′ UTR of MHV that is essential for viral replication (Hsue and Masters, 1997). The existence of this bulged stem–loop immediately following the N gene stop codon (Fig. 5) became apparent from analysis of allowed and unallowed substitutions of segments of the BCV 3′ UTR in place of the corresponding sequences in MHV, and it was confirmed through a combination of viral genetic constructs and DI RNA replication experiments. Genetic and structural methods are currently being used to further investigate the role of this structure in viral RNA synthesis.

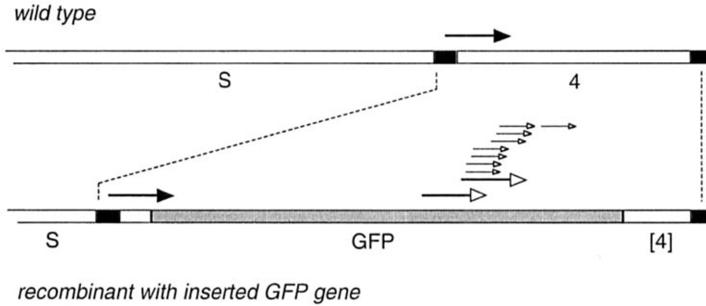


FIG 4. Aberrant transcription by an MHV mutant containing an inserted foreign gene. Portions of the wild-type and insertion mutant genomes are aligned at the 3′ end of gene 4. Intergenic sequences are represented by black rectangles. Corresponding intergenic sequences in the two viruses are connected by broken lines. Arrows with closed heads mark the points of leader-to-body fusion for normal gene 4 or gene 4(GFP) subgenomic RNAs. Arrows with open heads mark the points of leader-to-body fusion for major (large arrows) or minor (small arrows) aberrant subgenomic RNAs synthesized in the absence of a canonical intergenic sequence.

### *D. Limitations of Targeted Recombination*

In the foregoing examples, an attempt has been made to indicate the usefulness of targeted recombination as a means of carrying out reverse genetics on coronaviruses, but at least three current drawbacks or limitations of the method must be pointed out. The most obvious of these is the requirement to design mutations that are informative but are not lethal. Indeed, if a recombinant is to be obtained by selection, then it cannot be less fit than the recipient virus that is being selected against. This means that one must choose one's battles wisely. Although it can be unequivocally established that a mutation is *not* lethal or significantly deleterious to the virus, the converse is not true. As the technique is presently constituted, it cannot be determined without ambiguity whether failure to construct a given mutant is due to the lethality of the mutation in question or to other unknown factors in the experiment. A fairly convincing argument can be made that a particular mutation is lethal if one consistently obtains recombinants that inherit markers both upstream and downstream of the mutation of interest but exclude that mutation via multiple crossover events (de Haan *et al.,* 1998; L. Kuo and P. S. Masters, unpublished results, 1998). Nevertheless, an infectious full-length clone would afford a more straightforward means of determining whether a mutation was absolutely lethal.

A second limitation of targeted recombination is the current restriction of the method to the 3′ end of the genome. As noted above, only

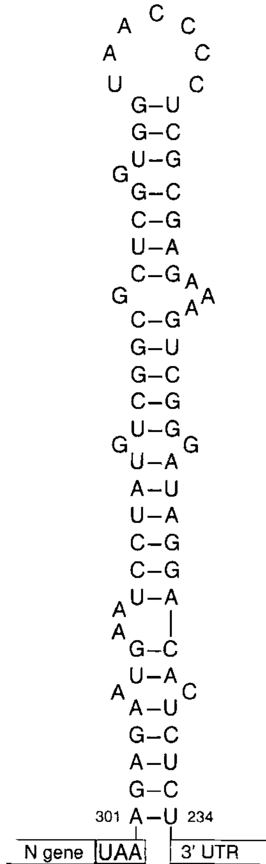


FIG 5. A bulged stem–loop structure in the 3′ UTR of MHV essential for viral replication. Nucleotide numbering begins at the first non-poly(A) base at the 3′ end of the genome. The stop codon for the N gene is boxed.

two silent mutations have been created thus far in gene 1 (van der Most *et al.,* 1992). Although strategies analogous to those used for the structural protein genes can be imagined, they will probably be most applicable only to the 5′ proximal region of the genome. A comprehensive genetic study of the highly complex gene for the RNA polymerase and all of its associated activities will likely await either the construction of an infectious full-length clone or the development of an innovative scheme for mutant selection.

A third limitation of targeted recombination, it is to be hoped, is a temporary one. So far, the application of this technique has been con-

fined to MHV. However, there are numerous cousins of MHV, particularly BCV, transmissible gastroenteritis virus (TGEV), infectious bronchitis virus (IBV), and feline infectious peritonitis virus (FIPV), that cause economically important diseases of livestock, poultry, and domestic animals. The development of targeted recombination for these viruses may aid the design of vaccines or other means of intervention. In addition, many fundamental advances in understanding the molecular biology and pathogenesis of coronaviruses have come from the study of these viruses, and a broadening of the genetic tools available for any one of them will undoubtedly benefit the entire field. Toward this end, DI RNAs have been identified, cloned, and characterized for BCV (Chang *et al.,* 1994), TGEV (Mendez *et al.,* 1996), and IBV (Penzes *et al.,* 1994). It has also been demonstrated that RNA recombination is not a unique property of MHV (Kottier *et al.,* 1995). It therefore seems likely that practical selection or screening strategies will be devised for site-directed mutagenesis and chimera construction for other coronaviruses.

## III. Conclusions and Future Prospects

Despite the limitations noted, the targeted RNA recombination system has given us access to a subset of the possibilities for genetic manipulation that will be allowed by the development of an infectious full-length clone of a coronavirus. Even when the latter goal is realized, targeted recombination may still be the method of choice for the construction of many genomic mutations. The sheer size of a full-length clone and its scarcity of unique restriction sites will likely necessitate that particular site-directed mutations be introduced via two stages of subcloning. Thus, at least for structural gene mutations that are not expected to be severely deleterious, targeted recombination may remain the less complicated alternative for creation of MHV mutants.

One area that may be explored further is the development of MHV as an expression vector. The failure of earlier attempts to insert a new transcription unit into the 3′ UTR of the virus (Masters *et al.,* 1995) can now be understood as having disrupted the essential bulged stem–loop structure in this region (Hsue and Masters, 1997). More judicious placement of an extra intergenic sequence in the 3′ UTR has been found to result in the synthesis of a new subgenomic transcript. However, there is still a barrier, not yet understood, to the length of the sequence that can be inserted downstream of the intergenic motif (Hsue and Masters, 1999). An alternative site of foreign gene insertion would be that which

has already been tried in gene 4 (Fischer *et al.*, 1997b). It may be possible to overcome the aberrant transcription observed here previously by altering the gene 4 intergenic sequence to resemble that of one of the more abundantly transcribed genes of MHV. It is conceivable, though, that we may then encounter other cryptic RNA sequences or structures that inherently limit the usefulness of coronaviruses as expression vectors. A novel approach to this problem has been pioneered by Liao *et al.* (1995), who engineered an MHV DI RNA to serve as an expression vector. This was used to express HE protein in cells infected with a strain of MHV that does not express its own HE protein, thereby producing pseudorecombinant virions that efficiently incorporated the expressed protein. Dissection of the DI HE gene led to the mapping of regions of the ectodomain of HE protein critical for its incorporation into virus particles.

An area of considerable promise for the future application of targeted recombination is the investigation of the pathogenesis of coronaviruses. Many previous studies in this field have, of necessity, relied on comparisons between viral strains with numerous major differences, not all of which are well characterized. Targeted recombination, in many cases, will allow construction of isogenic viruses differing at as little as a single codon. In addition, the mutations sought are generally not severely impaired in tissue culture and should thus be amenable to isolation by selection rather than by more cumbersome screening procedures. The two recently constructed S gene mutants discussed above illustrate these points well.

## Acknowledgments

I am grateful to Susan Weiss for the communication of unpublished results and to Lili Kuo for critically reading the manuscript. I am also indebted to numerous colleagues, especially Kathryn Holmes and David Brian, who have freely and generously shared materials important to many aspects of the work discussed. This work was supported in part by Public Health Service Grant AI 39544 from the National Institutes of Health.

## References

Banner, L. R., and Lai, M. M. C. (1991). Random nature of coronavirus RNA recombination in the absence of selective pressure. *Virology* **185,** 441–445.

Banner, L. R., Keck, J. G., and Lai, M. M. C. (1990). A clustering of RNA recombination sites adjacent to a hypervariable region of the peplomer gene of murine coronavirus. *Virology* **175,** 548–555.

Baric, R .S., Fu, K., Schaad, M. C., and Stohlman, S. A. (1990). Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. *Virology* **177,** 646–656.

Bos, E. C. W., Luytjes, W., van der Meulen, H., Koerten, H. K., and Spaan, W. J. M. (1996). The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. *Virology* **218,** 52–60.

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

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

Brown, T. D. K., and Brierley, I. (1995). The coronavirus nonstructural proteins. *In* "The Coronaviridae" (S. G. Siddell, ed.), pp. 191–217. Plenum, New York.

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

Compton, S. R., Barthold, S. W., and Smith, A. L. (1993). The cellular and molecular pathogenesis of coronaviruses. *Lab. Anim. Sci.* **43,** 15–28.

de Haan, C. A. M., Kuo, L., Masters, P. S., Vennema, H., and Rottier, P. J. M. (1998). Coronavirus particle assembly: Primary structure requirements of the membrane (M) protein. *J. Virol.* **72,** 6838–6850.

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.* **8,** 33–47.

Fischer, F., Peng, D., Hingley, S. T., Weiss, S. R., and Masters, P. S. (1997a). The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J. Virol.* **71,** 996–1003.

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

Fischer, F., Stegen, C. F., Masters, P. S., and Samsonoff, W.A. (1998). Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. *J. Virol.* **72,** 7885–7894.

Forns, X., Bukh, J., Purcell, R. H., and Emerson, S. U. (1997). How *Escherichia coli* can bias the results of molecular cloning: Preferential selection of defective genomes of hepatitis C virus during the cloning procedure. *Proc. Natl. Acad. Sci. U.S.A* **94,** 13909–13914.

Hofmann, M. A., and Brian, D. A. (1991). The 5′ end of coronavirus minus-strand RNAs contain a short poly(U) tract. *J. Virol.* **65,** 6331–6333.

Holmes, K. V., and Dveksler, G. S. (1994). Specificity of coronavirus/receptor interactions. *In* "Cellular Receptors for Animal Viruses" (E. Wimmer, ed.), pp. 403–443. Cold Spring Harbor Press, Plainview, NY.

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

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

Kirkegaard, K., and Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. *Cell (Cambridge, Mass.)* **47,** 433–443.

Koetzner, C. A., Parker, M. M., Ricard, C. S., Sturman, L. S., and Masters, P. S. (1992). Repair and mutagenesis of the genome of a deletion mutant of the coronavirus mouse hepatitis virus by targeted RNA recombination. *J. Virol.* **66,** 1841–1848.

Kottier, S.A., Cavanagh, D., and Britton, P. (1995). Experimental evidence of recombination in coronavirus infectious bronchitis virus. *Virology* **213,** 569–580.

Labeit, S., and Kolmerer, B. (1995). Titins: Giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270,** 293–296.

Lai, M. M. C. (1996). Recombination in large RNA viruses: Coronaviruses. *Semin. Virol.* **7,** 381–388.

Lai, M. M. C., Liao, C.-L., Lin, Y.-J., and Zhang., X. (1994). Coronavirus: How a large RNA viral genome is replicated and transcribed. *Infect. Agents Dis.* **3,** 98–105.

Lavi, E., Haluskey, J. A., and Masters, P. S. (1998). The pathogenesis of MHV nucleocapsid protein gene chimeric viruses. *Adv. Exp. Med. Biol.* **440,** 537–541.

Leparc-Goffart, I., Hingley, S. T., Chua, M. M., Phillips, J., Lavi, E., and Weiss, S. R. (1998). Targeted recombination within the spike gene of murine coronavirus mouse hepatitis virus-A59: Q159 is a determinant of hepatotropism. *J. Virol.* **72,** 9628–9636.

Liao, C.-L., Zhang, X., and Lai, M. M. C. (1995). Coronavirus defective-interfering RNA as an expression vector: The generation of a pseudo-recombinant mouse hepatitis virus expressing hemagglutinin-esterase. *Virology* **208,** 319–327.

Luytjes, W., Bredenbeek, P. J., Noten, A. F. H., Horzinek, M. C., and Spaan, W. J. M. (1988). Sequence of mouse hepatitis virus A59 mRNA2: Indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **166,** 415–422.

Masters, P. S., Koetzner, C. A., Kerr, C. A., and Heo, Y. (1994). Optimization of targeted RNA recombination and mapping of a novel nucleocapsid gene mutation in the coronavirus mouse hepatitis virus. *J. Virol.* **68,** 328–337.

Masters, P. S., Peng, D., and Fischer, F. (1995). Mutagenesis of the genome of mouse hepatitis virus by targeted RNA recombination. *Adv. Exp. Med. Biol.* **380,** 543–549.

Mehlin, H., and Daneholt, B. (1993). The Balbiani ring particle: A model for the assembly and export of RNPs from the nucleus? *Trends Cell Biol.* **3,** 443–447.

Mendez, A., Smerdou, C., Izeta, A., Gebauer, F., and Enjuanes, L. (1996). Molecular characterization of transmissible gastroenteritis coronavirus defective interfering genomes: Packaging and heterogeneity. *Virology* **217,** 495–507.

Meulenberg, J. J. M., Bos-de Ruijter, J. N. A., van de Graaf, R., Wensvoort, G., and Moormann, R. J. M. (1998). Infectious transcripts from cloned genome-length cDNA of porcine reproductive and respiratory syndrome virus. *J. Virol.* **72,** 380–387.

Peng, D., Koetzner, C. A., and Masters, P. S. (1995a). Analysis of second-site revertants of a murine coronavirus nucleocapsid protein deletion mutant and construction of nucleocapsid protein mutants by targeted RNA recombination. *J. Virol.* **69,** 3449–3457.

Peng, D., Koetzner, C. A., McMahon, T., Zhu, Y., and Masters, P. S. (1995b). Construction of murine coronavirus mutants containing interspecies chimeric nucleocapsid proteins. *J. Virol.* **69,** 5475–5484.

Penzes, Z., Tibbles, K., Shaw, K., Britton, P., Brown, T. D., and Cavanagh, D. (1994). Characterization of a replicating and packaged defective RNA of avian coronavirus infectious bronchitis virus. *Virology* **203,** 286–293.

Ricard, C. S., Koetzner, C. A., Sturman, L. S., and Masters, P. S. (1995). A conditional-lethal murine coronavirus mutant that fails to incorporate the spike glycoprotein into assembled virions. *Virus Res.* **39,** 261–276.

Rice, C. M., Grakoui, A., Galler, R., and Chambers, T. J. (1989). Transcription of infectious yellow fever RNA from full-length cDNA templates produced by in vitro ligation. *New Biol.* **1,** 285–296.

Risco, C., Anton, I. M., Enjuanes, L., and Carrascosa, J. L. (1996). The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. *J. Virol.* **70,** 4773–4777.

Senanayake, S. D., and Brian, D. A. (1997). Bovine coronavirus I protein synthesis follows ribosomal scanning on the bicistronic N mRNA. *Virus Res.* **48,** 101–105.

Sethna, P. B., Hofmann, M. A., and Brian, D. A. (1991). Minus-strand copies of replicating coronavirus mRNAs contain antileaders. *J. Virol.* **65,** 320–325.

Siddell, S. G. (1995). The Coronaviridae: An introduction. *In* "The Coronaviridae" (S. G. Siddell, ed.), pp. 1–10. Plenum, New York.

Thiel, V., Rashtchian, A., Herold, J., Schuster, D. M., Guan, N., and Siddell, S. G. (1997). Effective amplification of 20-kb DNA by reverse transcription PCR. *Anal. Biochem.* **252,** 62–70.

van der Most, R. G., and Spaan, W. J. M. (1995). Coronavirus replication, transcription, and RNA recombination. *In* "The Coronaviridae" (S. G. Siddell, ed.), pp. 11–31. Plenum, New York.

van der Most, R. G., Heijnen, L., Spaan, W. J. M., and de Groot, R. J. (1992). Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. *Nucleic Acids Res.* **20,** 3375–3381.

van 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 that abolishes discontinuous mRNA transcription. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 991–996.

Vennema, H., Godeke, G.-J., Rossen, J. W. A., Voorhout, W. F., Horzinek, M. C., Opstelten, D.-J. E., and Rottier, P. J. M. (1996). Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* **15,** 2020–2028.

Wertman, K. F., Drubin, D. G., and Botstein, D. (1992). Systematic mutational analysis of the yeast ACT1 gene. *Genetics* **132,** 337–350.

Zhang, L., Homberger, F., Spaan, W., and Luytjes, W. (1997). Recombinant genomic RNA of coronavirus MHV-A59 after coreplication with a DI RNA containing the MHV-RI spike gene. *Virology* **230,** 93–102.

Zhang, X., Liao, C.-L., and Lai, M. M. C. (1994). Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription both in *trans* and in *cis*. *J. Virol.* **68,** 4738–4746.