JOURNAL OF VIROLOGY, July 1997, p. 5148–5160 Vol. 71, No. 7
0022-538X/97/$04.00+0

Analysis of a Recombinant Mouse Hepatitis Virus Expressing a Foreign Gene Reveals a Novel Aspect of Coronavirus Transcription

FRANÇOISE FISCHER,[1] CAROLA F. STEGEN,[2] CHERI A. KOETZNER,[3] AND PAUL S. MASTERS[1,3]*

Departments of Biomedical Sciences[1] and Biological Sciences,[2] State University of New York at Albany, Albany, New York 12237, and Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, New York 12201[3]

Received 13 January 1997/Accepted 18 April 1997

We have inserted heterologous genetic material into the nonessential gene 4 of the coronavirus mouse hepatitis virus (MHV) in order to test the applicability of targeted RNA recombination for site-directed mutagenesis of the MHV genome upstream of the nucleocapsid (N) gene and to develop further genetic tools for site-directed mutagenesis of structural genes other than N. Initially, a 19-nucleotide tag was inserted into the start of gene 4a of MHV strain A59 with the N gene deletion mutant Alb4 as the recipient virus. In further work, the entire gene for the green fluorescent protein (GFP) was inserted in place of gene 4, creating the currently largest known RNA virus. The expression of GFP was demonstrated by Western blot analysis of infected cell lysates; however, the level of GFP expression was not sufficient to allow detection of fluorescence of viral plaques. Northern blot analysis of transcripts of GFP recombinants showed the expected alteration of the pattern of the nested MHV subgenomic mRNAs. Surprisingly, though, GFP recombinants also produced an RNA species that was the same size as wild-type mRNA4. Analysis of the 5′ end of this species revealed that it was actually a collection of mRNAs originating from 10 different genomic fusion sites, none possessing a canonical intergenic sequence. The finding of these aberrant mRNAs suggests that long-range RNA or the ribonucleoprotein structure of the MHV genome can sometimes be the sole determinant of the site of initiation of transcription.

Coronaviruses are a family of enveloped viruses of medical and veterinary importance, causing disease in humans, livestock, and fowl (50). The most intensively studied member of this virus family is mouse hepatitis virus (MHV), which, among other distinctive features, has the largest known genome of any RNA virus. This 31.3-kb single-stranded, positive-sense RNA genome contains 10 open reading frames (ORFs) (21, 50), the products of which, depending on the strain of MHV, include five or six structural proteins: the spike or surface glycoprotein (S), the hemagglutinin-esterase glycoprotein (HE), the membrane glycoprotein (M), the envelope or small membrane protein (E), the nucleocapsid protein (N), and the product of the internal ORF of the N gene (I). The structural protein genes are clustered in the 3′ third of the genome, whereas the 5′ two-thirds of the genome is occupied by a single gene encoding a huge polyprotein (ns1) comprising viral polymerase, proteases, and associated nonstructural proteins. Three other nonstructural proteins of unknown function (ns2a, ns4, and ns5a) are encoded by genes interspersed among the structural protein genes.

The expression of most of the genes of coronaviruses involves an unusual and incompletely understood mechanism of RNA synthesis generating subgenomic mRNAs. Each of these mRNAs consists of a short leader RNA identical to the extreme 5′ end of the genome, fused upstream of a given ORF, and containing all genomic sequence distal to the point of fusion. Transcription thus results in a 3′ nested set of mRNAs (21, 22, 50, 54). The sites of fusion occur at intergenic sequences (IGSs) that contain a stretch of sequence identical or nearly identical to the 3′ end of the leader RNA. For MHV, the core consensus IGS is the 9-nucleotide (nt) motif 5′ AAU CUAAAC 3′ (5, 20).

Coronavirus RNA synthesis also results in a high rate of homologous RNA-RNA recombination, a property that has been exploited to introduce site-directed mutations into MHV. Owing to their exceptional size, coronaviruses are currently beyond the reach of the method of choice for the engineering of other positive-strand RNA viruses, the transcription of infectious RNA from a cDNA clone of the entire genome. To circumvent this limitation, targeted RNA recombination has been carried out between synthetic donor RNAs and recipient viruses, the latter almost always being thermolabile mutants that can be counterselected. This technique has been used successfully to introduce mutations into gene 1 (53), the N gene (11, 35, 36, 53), which is the most distal of the viral ORFs, and the 3′ untranslated region (3′ UTR) (19) of MHV. To date, the furthest linear reaches of targeted RNA recombination have been a point mutation constructed by van der Most et al. in gene 1, 2,297 nt from the 5′ end of the genome (53), and a pair of point mutations that we have engineered in the N gene, 1,601 nt from the 3′ end of the genome (11).

The more general utility of targeted RNA recombination, especially for the genetic characterization of MHV structural proteins other than N, will depend upon extending the range of this method. In addition, a number of genes of MHV, those for HE, I, ns2a, ns4, and ns5a, have been shown to be nonessential for MHV replication (11, 25, 46, 55, 57). In principle, any of these except I (which is embedded within the N gene) could be used as sites for the introduction of foreign sequences. In this study we inserted heterologous genetic material into gene 4 of MHV in order to: (i) test the applicability of targeted RNA

* Corresponding author. Mailing address: David Axelrod Institute, Wadsworth Center, NYSDOH, New Scotland Ave., P.O. Box 22002, Albany, NY 12201-2002. Phone: (518) 474-1283. Fax: (518) 473-1326.

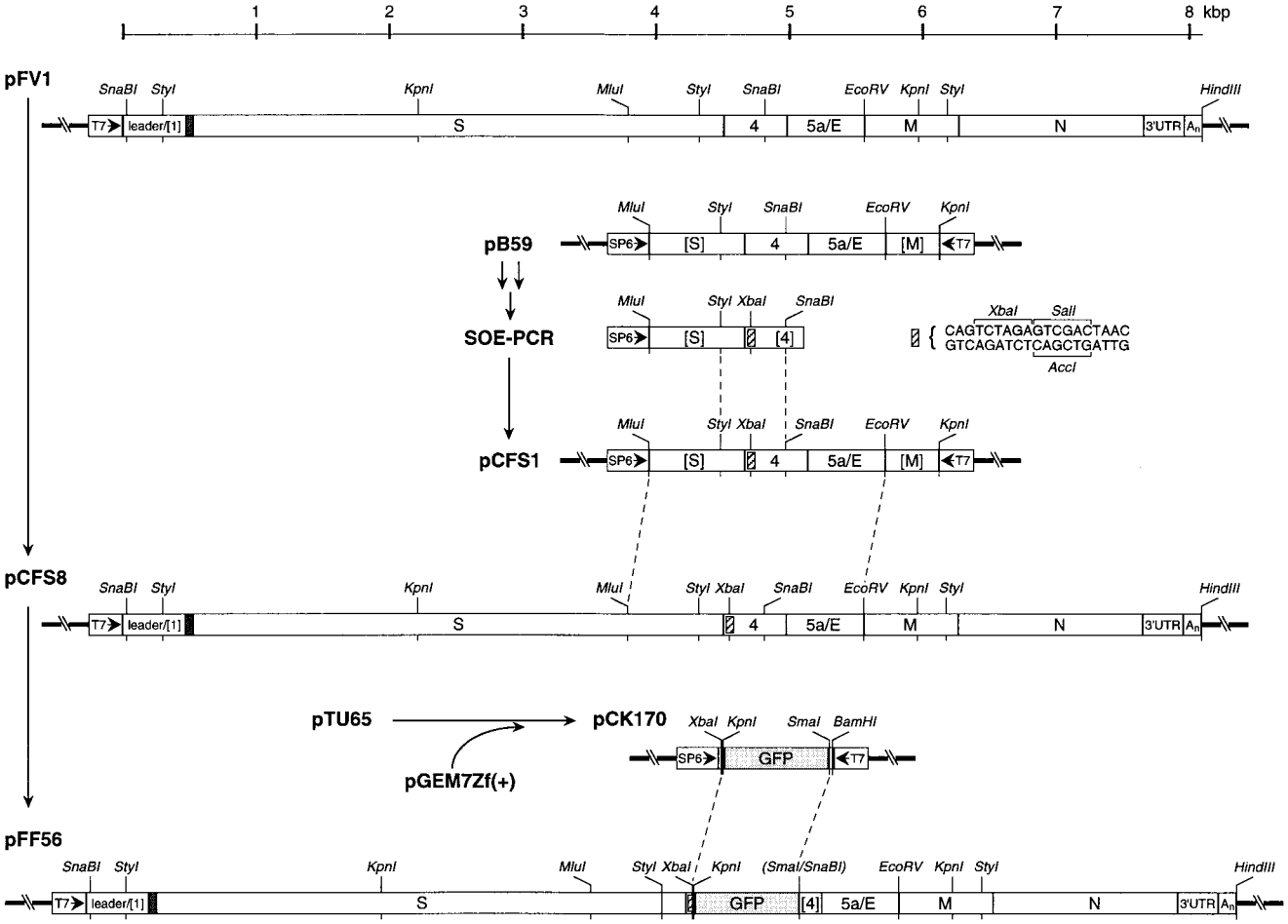

FIG. 1. Construction of transcription vectors for use as templates in synthesis of MHV DI RNAs used to incorporate heterologous genetic material into gene 4 of MHV-A59. The parent vector, pFV1, contains a 469-bp segment corresponding to the 5′ end of the MHV genome fused, via a 54-bp linker (shaded), to a 7,406-bp segment equal to the entire 3′ end of the MHV genome, beginning with the S gene. The vector pCFS8 contains a 19-bp sequence (hatched box), comprising unique *Xba*I and *Sal*I sites, inserted into gene 4 in place of the first 18 bp of ORF 4a. The final construct, pFF56, contains the entire GFP ORF plus flanking sequences (shaded) inserted into pCFS8 in place of the first two-thirds of gene 4. Restriction sites relevant to the construction of the plasmids, as detailed in Materials and Methods, are indicated. Regions shown in brackets are gene fragments rather than entire genes.

recombination for site-directed mutagenesis of the MHV genome upstream of the N gene; (ii) develop further genetic tools for mutagenesis of structural genes other than N; and (iii) examine the feasibility of using MHV for the expression of a foreign gene.

Initially, a small insertion was incorporated into the start of gene 4, both as a trial of the larger donor RNA vector used and to provide a sequence tag detectable by hybridization. Subsequently, a much larger insertion was made. In place of most of gene 4, we incorporated the reporter gene encoding the green fluorescent protein (GFP), thereby increasing the size of the MHV genome by an additional 0.5 kb. GFP, which contains a self-generated chromophore that emits green fluorescence upon irradiation with long-wave UV light, has been utilized in a variety of experimental systems (9). We were able to demonstrate expression of the incorporated GFP gene but not at levels sufficient to allow detection of fluorescing recombinant plaques. A surprising outcome of this study was the finding of a novel aspect of the mechanism of transcription initiation in MHV. A substantial population of wild-type-size gene 4 mRNAs were produced by GFP gene-containing recombinant viruses independent of the presence of any IGS, suggesting that under certain conditions, transcription initiation can be mediated completely by long-range RNA interactions.

MATERIALS AND METHODS

Virus and cells. Mouse 17 clone 1 (17Cl1) cells were used for propagation of all wild-type, mutant, and recombinant virus stocks of MHV strain A59. Plaque assays and purifications were conducted with mouse L2 cells, which were also maintained in spinner culture for RNA transfection via electroporation, as previously described (31).

Plasmid constructs. The parent vector used in this study was pFV1 (Fig. 1), which was derived through multiple steps from the N gene-containing defective interfering (DI) vector pB36 (31). The details of the construction of pFV1 will be described elsewhere. This plasmid encodes a T7 RNA polymerase transcript of 7,929 nt followed by a poly(A) tail of approximately 115 residues. It contains the same 5′ segment of the MHV genome as pB36 (31) fused in frame, through a 54-nt linker, to the start of the S gene ORF. Thereafter, it follows the composition of the 3′ end of the MHV genome from the S gene through the poly(A) tail. To allow the use of a *Hin*dIII site immediately following the poly(A) tail for linearization of pFV1, a *Hin*dIII site near the 5′ end of the S ORF (nt 517 to 522) was eliminated by replacement of AAGCTT with AAATTA. This alteration created silent changes in the two affected codons (K173 and L174), and it generated an *Ase*I site at nt 519 to 524. With the exception of this introduced restriction site polymorphism and the linker region, the remainder of the pFV1 insert reflects the sequence of our laboratory wild-type strain of MHV-A59. This includes two base differences from previously reported sequences (4, 55): nt 278 of gene 4b was T instead of G, and nt 251 of gene 5a was G instead of C. All other

TABLE 1. Oligonucleotides used

Oligonucleotide	Sequence	Sense[a]	Purpose; location
PM233	5′ CAGTCTAGAGTCGACTAACAGGCTACATTGGCTGCTGT 3′	Pos	SOE-PCR primer for introduction of tag into beginning of gene 4 in pCFS1; replaces start of ORF 4a
PM234	5′ GTTAGTCGACTCTAGACTGAGCTATGACTGCCTCTTA 3′	Neg	SOE-PCR primer for introduction of tag into beginning of gene 4 in pCFS1; replaces start of ORF 4a
SP6	5′ ATTTAGGTGACACTATA 3′	Pos	SOE-PCR primer for introduction of tag into beginning of gene 4 in pCFS1; nt 1 to 17 of SP6 promoter
PM154	5′ CAGAATTAAGATGAGGTT 3′	Neg	SOE-PCR primer for introduction of tag into beginning of gene 4 in pCFS1 and also used in RT-PCR analysis of GFP recombinant viruses; complementary to nt 265 to 282 of ORF 4b
PM231	5′ GCATCAGCATTAACATTAAGC 3′	Neg	Sequencing across the S gene *Hin*dIII site in recombinant viruses; complementary to nt 582 to 602 of the S ORF
PM259	5′ CAATGCCTAGCATACATG 3′	Neg	Sequencing of the gene 4a tag in recombinant viruses; complementary to nt 1 to 18 downstream of the end of ORF 4a
PM155	5′ GTGGCTGGCTTGCCAGTC 3′	Neg	RT-PCR analysis of gene 4a tag recombinant viruses; complementary to nt 75 to 92 of ORF 4b
PM156	5′ GGACAGTATTGTGATACA 3′	Pos	RT-PCR analysis of gene 4a tag and GFP recombinant viruses; nt 3933 to 3950 of S ORF
PM112	5′ CCATGATCAACTTCATTC 3′	Neg	RT-PCR analysis of gene 4a tag and GFP recombinant viruses; complementary to nt 267 to 284 of 3′ UTR.
PM119	5′ TAGTACTCTACCTGGTTT 3′	Pos	RT-PCR analysis of gene 4a tag and GFP recombinant viruses; nt 1077 to 1094 of N ORF
PM277	5′ AACTCATAGACACCAGTG 3′	Neg	RT-PCR analysis of GFP recombinant viruses; complementary to nt 906 to 923 of S ORF
A59-1	5′ CGGTCATACTGCTAGGTG 3′	Pos	RT-PCR analysis of GFP recombinant viruses; nt 1163 to 1180 of HE ORF
PM143	5′ TATAAGAGTGATTGGCGTCCGTACGTAC 3′	Pos	PCR primer for cloning leader-body junctions in gene 4 and gene 4 GFP-related mRNAs; nt 1 to 28 of MHV genome
FF34	5′ CATGCTGTTTCATATGAT 3′	Neg	PCR primer for cloning leader-body junctions in gene 4 GFP-related mRNAs; complementary to nt 226 to 243 of GFP ORF
FF35	5′ TAATCTAGGGTCTTAGGC 3′	Neg	PCR primer for cloning leader-body junctions in gene 4 and gene 4 GFP-related mRNAs; complementary to nt 234 to 251 of ORF 4b
FF36	5′ CAAAACCTCTTGGAGTACC 3′	Neg	Sequencing of the GFP gene in recombinant viruses; complementary to nt 199 to 217 of ORF 4b
FF37	5′ ACCATCTAATTCAACAAG 3′	Neg	Sequencing of the GFP gene in recombinant viruses; complementary to nt 43 to 60 of the GFP ORF
DP8	5′ AGTTGGCTCCAACAGTTGGTGC 3′	Pos	PCR primer for generation of template for random-primed probe; nt 947 to 968 of N ORF
DP17	5′ GGTTAATTCTCTACTTAC 3′	Neg	PCR primer for generation of template for random-primed probe; complementary to nt 1267 to 1284 of N ORF

[a] Pos, positive; Neg, negative.

base differences between this sequence and sequences reported by other laboratories (three at the 5′ end of the genome, two in the S gene, one in the M gene, three in the N gene, and five in the 3′ UTR) have been noted in prior reports (13, 19, 31, 34, 35).

Plasmid pCFS8 (Fig. 1), which is identical to pFV1 except for a 19-bp tag replacing the first 18 bp of ORF 4a, was derived from mutagenesis of a smaller pFV1-precursor clone, pB59. The 19-bp tag was created by splicing overlap extension (SOE)-PCR (16) with oligonucleotides PM233 and PM234 as the mutagenic primers and SP6 and PM154 as the outside primers (Table 1). The SOE-PCR product was restricted with *Sty*I and *Sna*BI and inserted in place of the corresponding fragment in pB59 to generate pCFS1. The *Mlu*I-*Eco*RV fragment of pCFS1 was then inserted in place of the corresponding fragment in pFV1 to generate pCFS8.

Plasmid pFF56 (Fig. 1) was generated in two stages from pCFS8. The wild-type GFP gene, contained within the *Kpn*I-*Bam*HI fragment of pTU65 (7), was transferred into the corresponding sites of the polylinker of pGEM7Zf(+) (Promega) to produce pCK170. The GFP gene-containing *Xba*I-*Sma*I fragment of pCK170 was then used to replace the *Xba*I-*Sna*BI fragment of pCFS8 in a three-way ligation with the downstream *Sna*BI-*Hin*dIII fragment and the large *Hin*dIII-*Xba*I fragment of pCFS8. The resulting plasmid, pFF56, thus contains 777 bp of heterologous (non-MHV) sequence, including the 717-bp GFP ORF, replacing the first 217 bp of gene 4. We noted that the stop codon of the GFP ORF is TAG, not TAA, as previously reported (37).

Recombinant DNA manipulations were carried out by standard methods (40). The compositions of all constructs were verified by restriction analysis; all cloned cDNA precursors, PCR-generated regions, and newly created junctions of each plasmid were verified by DNA sequencing by the method of Sanger et al. (41) with modified T7 DNA polymerase (Sequenase; U.S. Biochemical). For pFV1, the DNA sequence of the entire transcript-encoding region was also confirmed by automated sequencing with an Applied Biosystems model 373A DNA sequencer.

Targeted recombination. Incorporation of either the 19-nt tag or the entire GFP gene into the MHV genome was carried out by targeted RNA recombination between synthetic donor RNA from *Hin*dIII-truncated pCFS8 or pFF56 and the recipient virus Alb4, a thermolabile N gene deletion mutant (19). Synthetic capped donor RNAs were produced with a T7 polymerase transcription kit (Ambion) by following the manufacturer's protocol. Conditions for electroporation of donor RNA into Alb4-infected L2 spinner cells and plating onto 17Cl1 cell monolayers were essentially as previously described (31). Candidate recombinants were selected, either directly or following heat treatment at 40°C for 24 h (19), as viruses forming wild-type-size plaques at 39°C, the nonpermissive temperature for Alb4. Prior to further analysis, candidate recombinants were plaque purified at 39°C.

Virus purification, RNA purification, and sequencing of RNA. Purification of MHV by two cycles of equilibrium centrifugation on potassium tartrate-glycerol gradients and isolation of viral genomic RNAs were exactly as described previously (11, 19). A Nonidet P-40–gentle lysis procedure was used for the isolation of total cytoplasmic RNAs from virus-infected 17Cl1 cell monolayers (19). Direct sequencing of RNAs was performed by a modified dideoxy termination method (10, 36).

Northern blot analysis. Samples (10 μg) of total cellular RNAs isolated from mock-, wild-type-, Alb148-, or Alb151-infected 17Cl1 cells were denatured with 1 M deionized glyoxal and 50% dimethyl sulfoxide in MOPS buffer [40 mM 3-(*N*-morpholino)propanesulfonic acid (pH 7.0), 10 mM sodium acetate, 2 mM EDTA] for 1 h at 50°C. Following electrophoresis through 1% agarose in MOPS buffer, RNA samples were blotted onto a nylon membrane (Zeta Probe GT; Bio-Rad) per the manufacturer's instructions and hybridized with random-primed, [α-^{32}P]dATP-labeled DNA probes (40) specific for the GFP gene or the MHV N gene. The GFP probe was prepared from a 291-bp *Rsa*I-*Rsa*I fragment of pTU65 spanning nt −13 to 278 of the GFP ORF. The N probe was prepared from a 338-bp PCR product of primers DP8 and DP17 (Table 1) on pFV1 corresponding to nt 947 to 1284 of the N ORF. Probes were hybridized for 16 to 18 h at 65°C in 250 mM sodium phosphate (pH 7.2)–7% sodium dodecyl sulfate (SDS), and blots were then washed twice for 30 min at 65°C in 20 mM sodium phosphate (pH 7.2)–5% SDS and twice for 30 min at 65°C in 20 mM sodium phosphate (pH 7.2)–1% SDS. Bound probes were visualized by autoradiography.

Western blot analysis. Confluent 20-cm^2 monolayers of 17Cl1 cells were mock infected or infected with wild-type virus, Alb148, or Alb151 at a multiplicity of 5 PFU/cell. At 16 h postinfection, monolayers were washed twice with phosphate-buffered saline, lysed with 200 μl of 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 0.25% Nonidet P-40, 0.2 mg of phenylmethylsulfonyl fluoride per ml, and 0.01% sodium azide for 15 min at 4°C, and clarified by centrifugation. A control sample of [^{35}S]methionine-labeled GFP was produced by in vitro translation in a rabbit reticulocyte lysate of T3 polymerase transcripts from *Bam*HI-truncated pTU65 with standard reaction conditions (30). A 15-μl sample of each cell lysate and 5-μl control samples were separated by SDS-polyacrylamide gel electrophoresis (PAGE) on 14% polyacrylamide, transferred to a polyvinylidine membrane (Millipore), and probed with an anti-GFP polyclonal antibody immunoglobulin G fraction (Clontech) at a dilution of 1/200 in phosphate-buffered saline containing 10% Tween 20. Bound antibody was visualized with an enhanced-chemiluminescence detection system (Amersham).

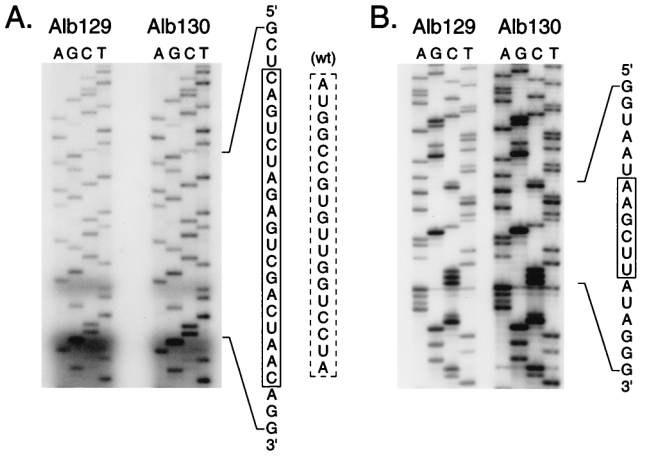

FIG. 2. Portions of genomic RNA sequences from purified virions of gene 4-tagged recombinants Alb129 and Alb130. The indicated segments of sequence are given as positive-sense RNA. (A) Sequence of the region where the 19-nt heterologous tag (solid-line box) was introduced into gene 4. The corresponding 18 nt of the wild-type (wt) virus (dashed-line box) is given to the right of this. (B) Sequence of the region of the S gene containing a unique *Hin*dIII site (box).

Analysis of candidate recombinants and of gene 4-related subgenomic mRNAs. Total RNA isolated from cells infected with candidate recombinant viruses, wild-type virus, Alb148, or Alb151 was reverse transcribed under standard conditions (40) with a random primer, p(dN)6 (Boehringer Mannheim). For characterization of candidate recombinants, cDNA was amplified by PCR with the primer pairs indicated below, and products were examined by agarose gel electrophoresis, in some cases with prior restriction digestion. PCRs were run for 30 cycles of 1 min at 94°C, 1 min at 45°C, and 2 min at 72°C. For the analysis of gene 4-related mRNAs in the GFP recombinants, cDNA was PCR amplified with the leader-specific primer PM143 (Table 1) paired with either a primer specific to the GFP gene (FF34) or a primer specific to gene 4b (FF35). PCR products were cloned into the TA cloning vector pCR2.1 (Invitrogen), and the inserts of multiple clones were sequenced.

RESULTS

Generation of a gene 4a-tagged recombinant virus by targeted RNA recombination. In MHV-A59, gene 4 is disrupted by an early frameshift generating a very short, 19-amino-acid ORF, 4a, followed by a longer 106-amino-acid ORF, 4b (55). The ORF 4a protein is not homologous to the amino terminus of the product encoded by MHV-JHM gene 4 (51), suggesting that it cannot even partially fulfill the same function, if any, possessed by that protein. In another MHV strain, MHV-S, transcription of mRNA 4 cannot be detected (57). Together, these observations argue with near certainty that MHV ns4 is not essential for the replication of MHV and they suggest that gene 4 is a fortuitous location for the insertion of foreign sequences into the MHV genome.

In previous work, we have used a synthetic DI RNA, containing a small portion of the 5′ end of the genome fused to the entire N gene and 3′ UTR, as the donor for targeted RNA recombination to map and to construct N gene mutants (31, 36). A slightly enlarged modification of this donor RNA containing half of the adjacent M gene was subsequently employed to place two point mutations very near the 5′ end of the N gene (11). Our strategy, then, for mutagenesis upstream of the N gene was to provide donor RNA containing all of the essential structural genes of MHV. For this purpose, we constructed a transcription vector, pFV1 (Fig. 1), that served as the template for synthesis of an 8-kb RNA consisting of the 5′ genomic segment fused to the genes for S, ns4, ns5a, E, M, N, and the 3′ UTR. One alteration with respect to the wild-type MHV A59 sequence was made in pFV1: silent changes were intro-

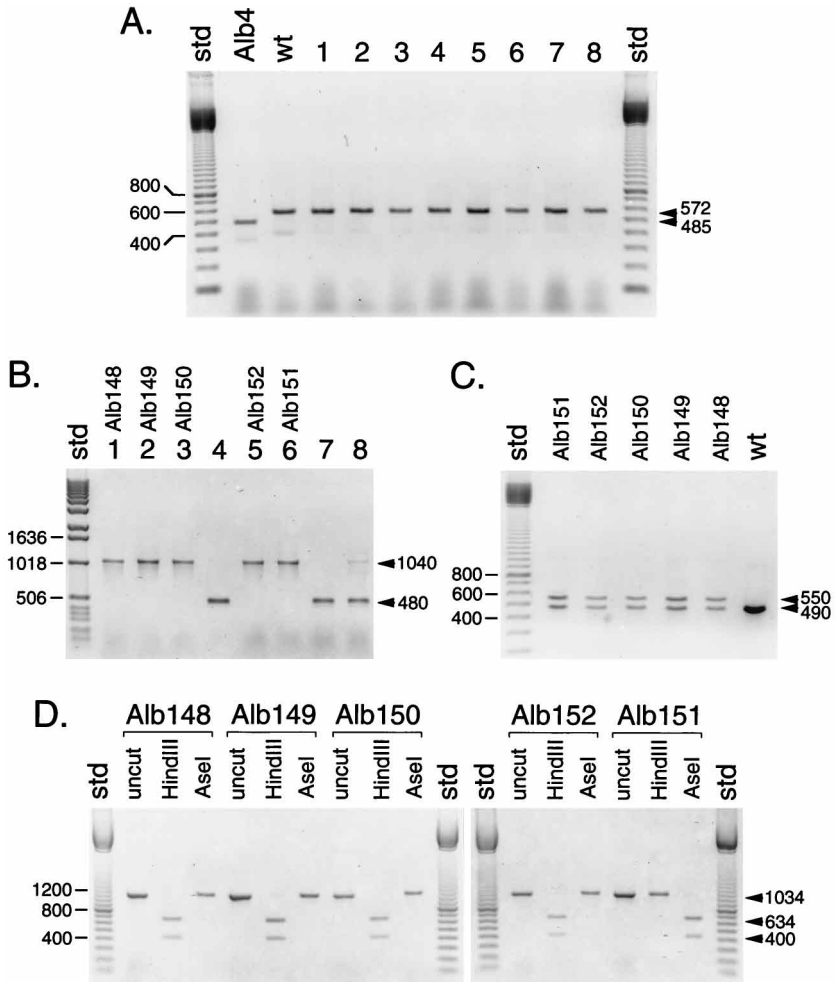

FIG. 3. PCR analysis of GFP recombinants. RT-PCR was used to amplify regions of RNA isolated from cells infected with each of eight plaque-purified candidate recombinant viruses or wild-type (wt) or Alb4 controls. PCR products were analyzed by electrophoresis on 1% agarose stained with ethidium bromide. Sizes of product fragments (in base pairs) are indicated by arrowheads to the right of each gel; sizes of relevant marker DNA fragments are indicated to the left of each gel. (A) RT-PCR products obtained with the primer pair PM112 and PM119 (Table 1), which flank the Alb4 deletion region in the N gene. (B) RT-PCR products obtained with primer pair PM154 and PM156 (Table 1), which flank the site of the GFP gene insertion. The five candidates found to contain the GFP gene were given the designations Alb148 to Alb152. (C) *Dra*I digests of the 1,040-bp RT-PCR products from panel B corresponding to the GFP-containing recombinant viruses and the 480-bp product from a wild-type control. (D) RT-PCR products obtained with primer pair A59-1 and PM277 (Table 1), which flank the region of the unique *Hin*dIII site in the wild-type MHV S gene (nt 517 to 522 of the S ORF), were analyzed without digestion or following digestion with *Hin*dIII or *Ase*I. std, standard markers.

duced into two codons at nt 517 to 522 of the S ORF in order to eliminate a *Hin*dIII site and simultaneously replace it with an *Ase*I site (see Materials and Methods). It had been previously demonstrated that the original synthetic donor RNA (which included only the N gene) was an authentic DI RNA, i.e., it was replication competent in the presence of helper virus (31). However, it should be noted that it is not clear, nor is it relevant to the present work, whether the larger donor RNA species (from pFV1 and its descendants) are actually DI RNAs.

To create a marker in gene 4 of MHV, a mutational insertion was generated in pFV1. In the resulting vector, pCFS8, the first 18 bp of gene 4a were replaced by 19 bp of heterologous sequence encompassing the diagnostic restriction sites *Xba*I, *Sal*I, and *Acc*I (Fig. 1). Donor RNA transcripts from pCFS8 were transfected into cells that had been infected with Alb4, a temperature-sensitive and thermolabile N gene deletion mutant (19). Following heat treatment to counterselect the recipient virus (19), candidate recombinants were isolated as progeny able to form wild-type-size plaques at the nonpermissive temperature (39°C). Dot blot hybridization of infected cellular RNAs from 30 candidates showed that 26 of them had recovered the region that is deleted in the Alb4 mutant and thus had recombined with donor RNA at least in the N gene. Six of these were then examined by reverse transcription followed by PCR (RT-PCR), and restriction digestion analysis of the products showed that five of the six candidates contained the inserted tag (data not shown). Final proof of incorporation of the tag into gene 4 came from direct RNA sequencing of genomic RNA isolated from purified virions of two independently obtained recombinants, Alb129 and Alb130 (Fig. 2A). We also sequenced the region of the *Hin*dIII site in the S gene and found that this was still intact in Alb129 and Alb130 (Fig. 2B), which thus had not received the silent codon changes that had been incorporated into the donor RNA. The simplest interpretation of these data is that these two recombinants had been generated by a single crossover event occurring somewhere in the 3.5 kb between the S gene *Hin*dIII site and the start of gene 4. We cannot rule out the possibility of an even number of additional crossover events, but these would have

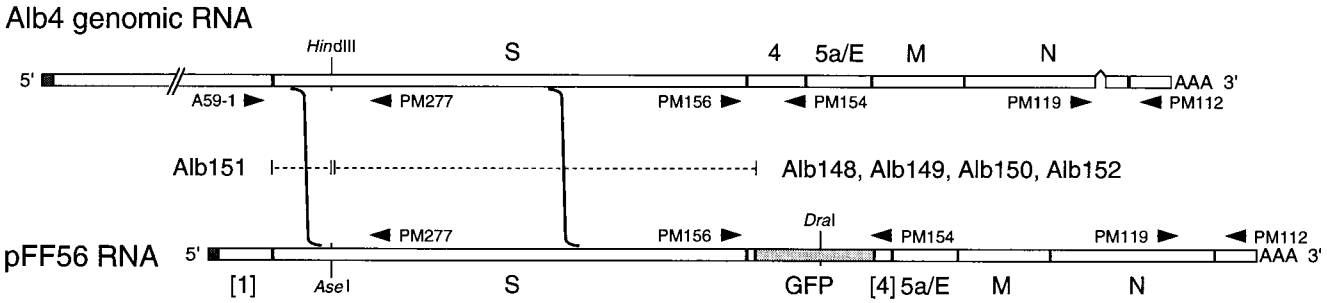

FIG. 4. Targeted recombination between Alb4 genomic RNA and pFF56 RNA. The locations of the primers and the *Dra*I, *Hin*dIII, and *Ase*I sites used in the analysis in Fig. 3 are shown. The delimited broken line to the left indicates the limits within which a single crossover event could have generated recombinant virus Alb151. The delimited broken line to the right indicates the limits within which a single crossover event could have generated recombinant viruses Alb148, Alb149, Alb150, and Alb152.

been undetectable since there were no other sequence differences between the donor RNA and the recipient virus. These results showed that an extension of the previously used targeted RNA recombination technique could be applied to a gene upstream of N and that further disruption of gene 4 had no observable phenotypic consequence.

Generation of recombinant MHV containing the GFP gene. We next addressed the possibility of inserting an entire functional gene into the same locus of the MHV genome. This was accomplished through the incorporation of the GFP reporter into the 4a-tagged plasmid pCFS8 to produce the vector pFF56 (Fig. 1). In pFF56, the 717-bp GFP ORF, as well as some flanking sequence, replaced the first two-thirds of wild-type MHV gene 4. Synthetic donor RNA transcribed from pFF56 was transfected into Alb4-infected cells, and released progeny virus was harvested. After an initial plaque assay screen at 37°C failed to detect any green fluorescing plaques under UV light, we employed our standard selection procedures to identify MHV potentially harboring the GFP gene. Plaques of progeny virus, obtained either directly or with prior heat treatment, were assayed at the temperature that is nonpermissive for the parental Alb4 mutant. Eight candidate recombinants were thus isolated as viruses able to form wild-type-size plaques at 39°C, and these were plaque purified.

RT-PCR analysis was performed with total RNAs purified from cells individually infected by each of the eight recombinant virus candidates. At first, the region that is absent in the

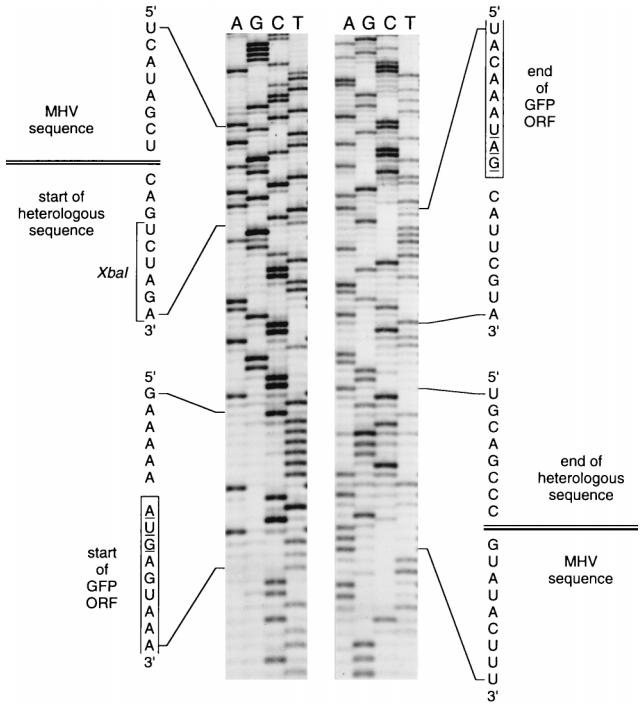

FIG. 5. Portions of genomic RNA sequence from purified virions of GFP recombinant Alb148. The indicated segments of sequence, given as positive-sense RNA, show the limits of the GFP ORF and the junctions between authentic MHV sequence and inserted heterologous sequence.

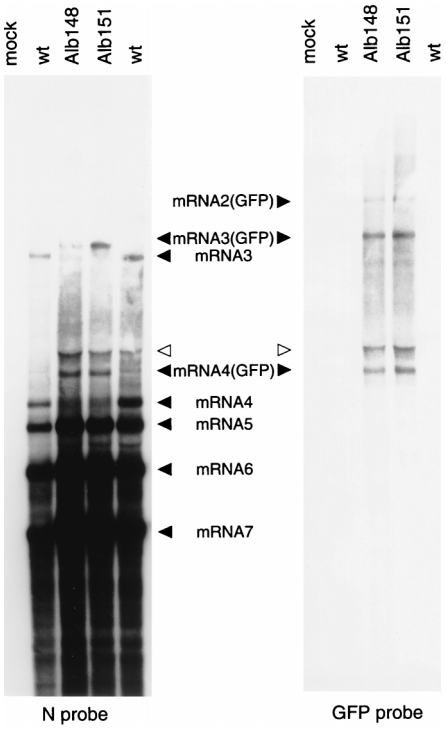

FIG. 6. Northern blot analysis of RNA synthesized by GFP recombinants. Total cellular RNA prepared from mock-, wild-type (wt)-, Alb148-, or Alb151-infected 17Cl1 cells was glyoxal denatured and separated by electrophoresis through 1% agarose. RNA was transferred to a blot and hybridized with a probe specific for the MHV N gene (left panel) or a probe specific for the 5′ end of the GFP gene (right panel). Subgenomic mRNAs are indicated by filled arrowheads. The open arrowheads mark the position of 28S rRNA. The lower portion of the autoradiogram was overexposed in order to more clearly show the less-abundant higher-molecular-weight viral mRNAs.

Alb4 N gene was amplified with a primer pair flanking the locus of the deletion in that mutant (PM112 and PM119) (Table 1). In an Alb4 RNA control, this procedure produced a PCR product of 485 bp, whereas the corresponding product from wild-type-infected cellular RNA was 572 bp, as has been observed previously (19). As shown in Fig. 3A, all eight candidates produced a wild-type-size PCR product and therefore had recombined with donor RNA, at least to the extent necessary to repair the Alb4 deletion.

A second RT-PCR amplification of the same set of eight infected cellular RNA samples was then conducted with a primer pair flanking the insertion site of the GFP gene (PM154 and PM156) (Table 1). For recombinants containing the GFP gene insertion, this was expected to produce a fragment of 1,040 bp, while recombinants lacking the GFP gene were expected to give rise to a wild-type-size fragment of 480 bp. This analysis showed that five of the eight recombinants, which were assigned the names Alb148, Alb149, Alb150, Alb151, and Alb152, contained the inserted GFP gene (Fig. 3B). The PCR products from these viruses were digested with *Dra*I, which has a unique site in the amplified region containing the GFP gene (Fig. 4). As expected, *Dra*I digestion produced two fragments of 490 and 550 bp for all five GFP recombinants, whereas a control (480-bp) wild-type fragment produced with the same primers was not cut by *Dra*I (Fig. 3C).

To determine whether any of the five GFP recombinants also contained the restriction site polymorphism that had been introduced into the 5′ end of the S gene of the donor RNA, primers producing a product crossing this region (A59-1 and PM277) (Table 1) were used in an RT-PCR of infected cellular RNA. Since in the donor RNA the wild-type S gene *Hin*dIII site had been replaced by an *Ase*I site, portions of the five PCR products were separately digested with these two enzymes. Four of the GFP recombinants (Alb148, Alb149, Alb150, and Alb152) were found to have retained the wild-type *Hin*dIII site. However, the remaining recombinant (Alb151) had lost the *Hin*dIII site and had gained the *Ase*I site (Fig. 3D). The last result showed that at least two sets of siblings had been isolated from the same infection-transfection experiment. Assuming that recombinants were generated by a single crossover, this crossover must have occurred between the *Hin*dIII locus and the 3′ end of the S gene for Alb148, Alb149, Alb150, and Alb152 (Fig. 4). For Alb151, on the other hand, the putative crossover could be more precisely localized to within the first 519 nucleotides of the S ORF (Fig. 4). It is noteworthy that this crossover was at least 6.3 kb upstream of the selected marker.

To provide final confirmation of the presence of the GFP gene, for one of the recombinants, Alb148, genomic RNA was isolated from highly purified virions and the junctions of the inserted gene were directly sequenced (Fig. 5). This result, together with that of the RT-PCR analysis, established unambiguously that a substantial amount of heterologous RNA could be incorporated into the genome of MHV without producing any obvious deleterious effect on the virus.

Transcription of new subgenomic mRNAs containing the GFP gene. To analyze whether the introduced GFP gene altered the transcription pattern of MHV subgenomic mRNAs upon viral replication, Northern blot analysis was performed (Fig. 6). Total cellular RNAs prepared from mock-, wild-type-, Alb148-, and Alb151-infected cells were electrophoresed, blotted to a membrane, and hybridized with either an N gene probe or a probe derived from the 5′ end of the GFP gene. The N gene probe detected subgenomic mRNAs 3, 4, 5, 6, and 7 for both the wild-type and recombinant Alb148 and Alb151 RNA preparations (Fig. 6, left panel). No N gene-hybridizing material was detected in mock-infected cellular RNA. mRNA5, -6, and -7 of wild-type and recombinant viruses were identical in size and relative molar ratios, as was expected since these mRNAs are transcribed from IGSs downstream from the site of the GFP insertion. By contrast, mRNA4 and -3 of Alb148 and Alb151 migrated at higher molecular weights than the corresponding mRNAs from wild-type virus. This difference in size was compatible with the size of the inserted GFP gene in Alb148 and Alb151. The GFP probe detected only those mRNAs containing the GFP gene, mRNA4(GFP) and mRNA3(GFP) of Alb148 and Alb151, and no RNA species in wild-type- or mock-infected cells (Fig. 6, right panel). The genomic mRNA1 was not detected with either probe, and subgenomic mRNA2(GFP) was only weakly detected by the GFP probe, most likely because of inefficient transfer of these very large RNA species during blotting. These results showed that, in cells infected by the recombinant viruses, the GFP gene was transcribed as a part of a new, larger version of subgenomic mRNA4 and of all mRNAs larger than mRNA4. There was one anomaly in this pattern, however. Surprisingly, the N probe, but not the GFP probe, also detected a wild-type-size mRNA4 among the RNA species from both Alb148- and Alb151-infected cells.

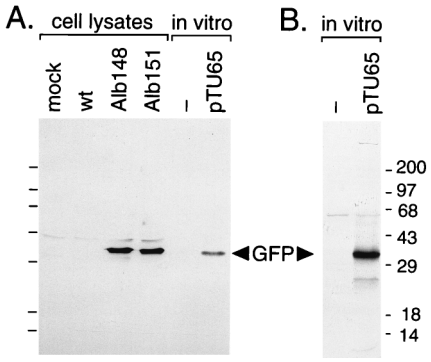

FIG. 7. Analysis of GFP expressed in vivo. (A) Western blot of lysates prepared from mock-, wild-type (wt)-, Alb148-, or Alb151-infected cells and in vitro-translated protein from rabbit reticulocyte lysates programmed with H_2O (lane −) or RNA transcribed from pTU65. Samples were separated by SDS-PAGE in a 14% polyacrylamide gel and probed in a Western blot with anti-GFP antibody. The blot was visualized by detection of chemiluminescence. (B) Autoradiogram (48-h exposure) of a portion of the same blot shown in panel A following decay of the chemiluminescent signal. Molecular masses (in kilodaltons) of marker proteins are indicated.

Expression of GFP during recombinant virus replication. Because the mRNA4(GFP) detected in Alb148- and Alb151-derived RNAs should have been a functionally monocistronic message for GFP, we sought evidence for GFP expression in cells infected by these recombinants. Western blot analysis of infected cell lysates revealed that anti-GFP antibody recognized a protein in Alb148- and Alb151-infected cells that was absent from wild-type- and mock-infected cells (Fig. 7A). This protein had the same mobility as determined by SDS-PAGE as authentic GFP translated in vitro from synthetic pTU65-derived mRNA (Fig. 7). Although the calculated molecular mass of GFP is 27 kDa, we observed that GFP migrated as a protein of at least 31 kDa. However, GFP has been previously described as having an unusual electrophoretic mobility in SDS-PAGE analysis and molecular masses of 30 to 31 kDa have been reported by other investigators (17, 38). GFP, which has a rigid β-barrel native structure (33), may exhibit various mobilities in SDS-PAGE analysis owing to incomplete denaturation (1). Clearly, our results show that GFP was expressed in

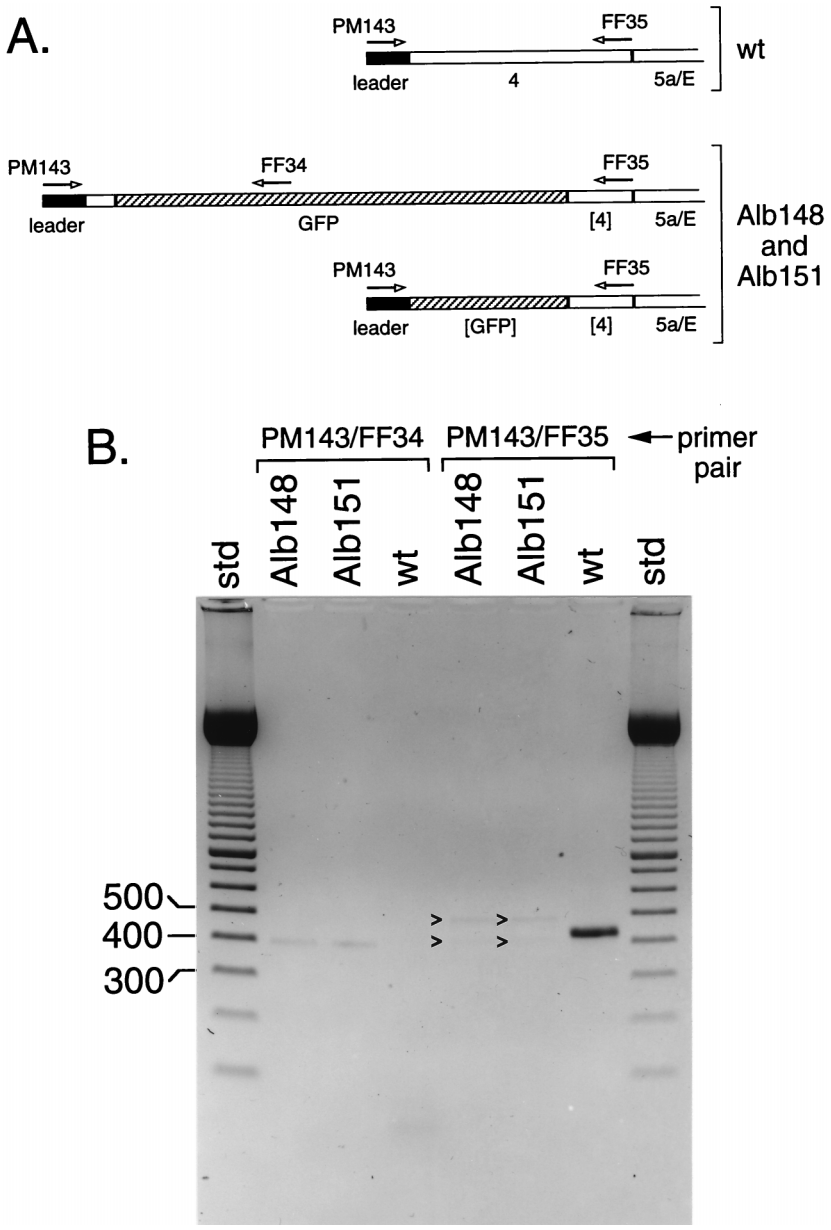

FIG. 8. PCR analysis of gene 4-related mRNAs. (A) Schematic of primers and templates used in PCRs. The locations and the orientations of the primers are indicated by arrows. (B) RT-PCR products obtained from total RNAs from cells infected with wild-type (wt), Alb148, or Alb151 were analyzed by electrophoresis on 1% agarose stained with ethidium bromide. The RT-PCR products obtained with Alb148 and Alb151 with primer pair PM143 and FF35 are indicated by carets. Sizes (in base pairs) of relevant marker DNA fragments are indicated to the left of the gel. std, standard markers.

MHV-infected cells even though we were unable to directly detect its fluorescence.

Determination of aberrant leader-body junctions of gene 4-related mRNAs produced by GFP recombinant viruses. As noted above, the N gene probe detected both mRNA4(GFP) and an unexpected RNA that was the same size as wild-type mRNA4 in Northern blot analysis of total cellular RNAs from Alb148- and Alb151-infected cells (Fig. 6, left panel). These two mRNAs were present in approximately equivalent amounts, and each appeared to be less abundant than wild-type mRNA4. When duplicate samples were hybridized with the GFP probe, only the mRNA4(GFP) and larger subgenomic mRNAs were observed, but not the wild-type-size mRNA4-like species (Fig. 6, right panel). Since the GFP probe was derived from the 5′ end of the GFP gene, both the size and the hybridization pattern of the unexpected RNA species suggested that it originated in the 3′ half of the GFP gene. Although it is conceivable that the wild-type-size mRNA4-like species originated from contaminating wild-type virus, this possibility was definitively excluded by the PCR analysis of the region flanking the GFP insertion in Alb148 and Alb151, which yielded only a GFP gene-size fragment (Fig. 3B), and by the unique reading obtained from direct RNA sequencing of the genome of purified Alb148 (Fig. 5). Furthermore, we did not detect an RNA of the same size as wild-type mRNA3 in Northern blot analysis of Alb148 and Alb151 with either probe, which would have been expected if there had been contaminating wild-type RNA in the RNAs from these two recombi-

TABLE 2. Leader-body junctions of gene 4 and gene 4 GFP-related mRNAs

Type	No. of clones/ total no. of clones	Sequence of mRNA leader-body junction[a]
Wild-type mRNA4	8/8	5′ . . .uaaaacucuuguaguuuaaaucu**AAUCUAAAC**AAUUUAUAGCA. . . 3′
Alb148 and Alb151 mRNA4(GFP)	17/17	5′ . . .uaaaacucuuguaguuuaaaucu**AAUCUAAAC**AAUUUAUAGCA. . . 3′
Alb148 and Alb151 mRNA4(GFP)-related aberrant transcripts		
a	23/52	5′ . . .uaaaacucuuguaguuuaaaucuaa**UCU**UGGACACAAAUUGGAAUACAAC. . . 3′
b	13/52	5′ . . .uaaaacucuuguaguuuaaaucu**AAUC**AAAGUUAACUUCAAAAUUAGACACAAC. . . 3′
c	1/52	5′ . . .uaaaacucuuguaguuua**AAUC**AAAGUUAACUUCAAAAUUAGACACAAC. . . 3′
d	3/52	5′ . . .uaaaacucuuguaguuuaaaucuaauc**U**UAACUUCAAAAUUAGACACA. . . 3′
e	2/52	5′ . . .uaaaacucuuguaguuuaaaucuaaucua**AACUU**CAAAAUUAGACACAAC. . . 3′
f	1/52	5′ . . .uaaaacucuuguaguuuaaaucuaaucu**AA**UUAGACACAACAUUGA. . . 3′
g	2/52	5′ . . .uaaaacucuuguaguuuaaaucuaauc**UA**GACACAACAUUGAAGAUGGAAGCGU. . . 3′
h	1/52	5′ . . .uaaaacucuuguaguuuaaaucu**AA**CUAGCAGACCAUUAUCAACAAAA. . . 3′
i	2/52	5′ . . .uaaaacucuuguaguuuaa**A**CCAUUAUCAACAAAAUACUCCAAUUGGCG. . . 3′
j	2/52	5′ . . .uaaaacucuuguaguuuaaaucua**AUC**AACAAAAUACUCCAAUUGGCGA. . . 3′
k	2/52	5′ . . .uaaaacucuuguaguuuaaaucu**AAUCU**GCCCUUUCGAAAGAUCCCAAC. . . 3′

[a] For each junction, the 5′ segment shown in lowercase type, which begins at nt 41 of the MHV-A59 leader, consists of those nucleotides in the mRNA that are unique to the leader. The central segment, underlined and in boldface type, comprises nucleotides common to both the leader and the mRNA body. The 3′ segment following this section marks the beginning of those nucleotides that are unique to the body of the mRNA.

nants. The possibility of wild-type virus contamination was also ruled out by the experiments described below.

To investigate the origin of the unexpected mRNA, RT-PCR was conducted with primer pairs designed to cross the junctions between the leader and body portions of all mRNA4- and mRNA4(GFP)-related species. For PCR amplification of cDNA, the primer PM143, falling at the start of the leader, was paired with either primer FF34, located within the GFP ORF, or primer FF35, located near the 3′ end of gene 4b (Fig. 8A and Table 1). The primer pair PM143 and FF34, when used with RNA from Alb148- or Alb151-infected cells, gave the predicted RT-PCR product of 403 bp, corresponding to mRNA4(GFP) (Fig. 8B). As expected, wild-type RNA yielded no product with this GFP-specific primer pair. The primer pair PM143 and FF35, in conjunction with RNA from wild-type-infected cells, gave the predicted RT-PCR product of 438 bp, corresponding to wild-type mRNA4 (Fig. 8B). The same primer pair, when used with RNA from Alb148- or Alb151-infected cells, did not give a wild-type-size PCR fragment, but at least two other PCR products in the 400- to 500-bp range were produced (Fig. 8B). We did not detect the larger, 998-bp fragment that was predicted for the full-length mRNA4(GFP) with this primer pair, but most likely the smaller 400- to 500-bp species were preferentially amplified at the expense of this potential product.

The products of each PCR were cloned in bulk, and the corresponding leader-body junctions of multiple clones were sequenced (Table 2). The leader-body junction that was observed for wild-type mRNA4 (8 of 8 clones) was identical to that previously determined for this junction in MHV-A59 (20). Moreover, this same leader-body junction, containing 9 nt common to the leader and the mRNA body, was also seen for the full-length mRNA4(GFP) produced by both Alb148 and Alb151 (17 of 17 clones, distributed almost equally between the two recombinants). Surprisingly, 11 different types of leader-body junctions (Table 2) were revealed upon sequencing of a total of 52 clones corresponding to the smaller (wild-type-size) mRNA4(GFP)s produced by Alb148 and Alb151. Of these 11 aberrant junctions, two were predominant (types a and b, 23 of 52 clones and 13 of 52 clones, respectively) and were distributed equally between Alb148 and Alb151. The two major populations found by cloning appear to correspond to the two major PCR product bands observed by gel electrophoresis (Fig. 8B). The remaining aberrant junctions (16 of 52 clones) were each represented by 1 to 3 clones. For each of the aberrant junctions, the number of nucleotides common to the leader and the mRNA body was as few as 1 and was, at most, 5 (Table 2), contrasting markedly with the common regions of wild-type MHV leader-body junctions, which range between 7 and 18 nt (20, 54).

The genomic localization of the normal and aberrant mRNA fusion sites found in the gene 4-related mRNAs is summarized in Fig. 9. We identified 10 aberrant sites in Alb148 and Alb151, corresponding to the 11 aberrant leader-body junctions described above. None of these junctions falls within a consensus IGS or a recognizable cryptic IGS. Since the multiple aberrant mRNA fusion sites were contained within a narrow range equivalent to only 5% of the total size of wild-type mRNA4 (3.7 kb), this explained why the resulting mRNAs appeared to migrate as a single species in the Northern blot (Fig. 6). Remarkably, all of the aberrant fusion sites were located within a 205-nt window (between nt 408 and 612 of the GFP ORF), and the two predominant sites defined an even smaller interval of 73 nt, which encompassed the position that would correspond to wild-type-size mRNA4. This suggests that the synthesis of the aberrant mRNA4(GFP) species was IGS independent and was determined by MHV RNA sequences or structures (of either RNA or ribonucleoprotein) outside the GFP gene, distant from the actual fusion sites.

DISCUSSION

In this study we inserted heterologous sequences into the genome of MHV, replacing portions of the nonessential gene 4 by targeted RNA recombination. We first inserted a 19-nt tag and, subsequently, an entire reporter gene and flanking sequences (777 nt) into the gene 4 region. Both the 4a-tagged and GFP recombinants grew to high titers and were pheno-

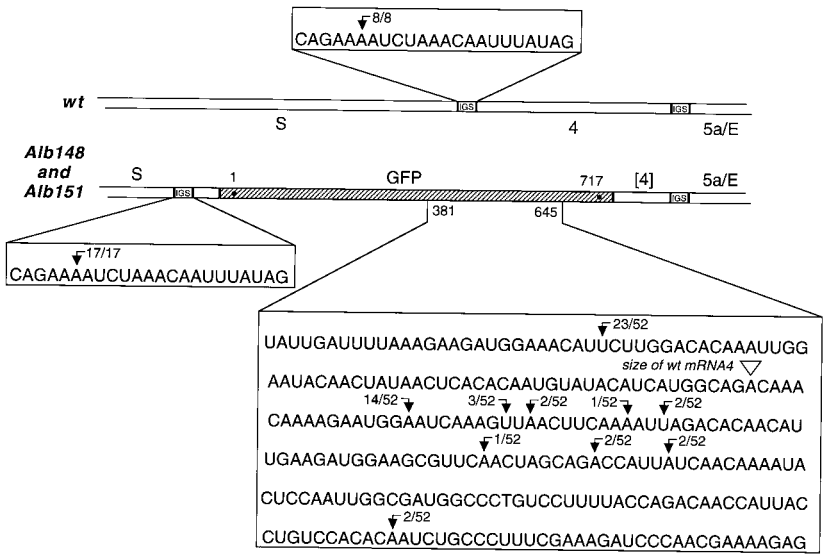

FIG. 9. Summary of normal and aberrant mRNA fusion sites found in gene 4-related mRNAs produced by wild-type (wt) and GFP recombinant viruses Alb148 and Alb151. The schematic representation of portions of the mutant and wild-type viral genomes are aligned at their 3′ ends, and consensus IGSs are indicated. Numbers above and below the GFP insert (hatched rectangle) indicate nucleotide residues in the GFP ORF. Expanded regions show genomic RNA sequence at initiation sites. Filled arrows point to each mRNA fusion site, defined as the first nucleotide in the genome common to the leader and the body of the mRNA (Table 2). Numbers above these arrows indicate the ratio of PCR clones exhibiting this fusion site to the total number of clones sequenced for each class of transcript. The open arrowhead indicates the position in Alb148 and Alb151 that would correspond to the fusion site of wild-type-size mRNA4.

typically indistinguishable from the wild type with respect to growth in tissue culture. To date, these mutants mark the furthest distance from either terminus of the MHV genome to which the targeted recombination method has been applied. The 4a-tagged and GFP recombinants described here were generated by crossover events that occurred at least 3.3 kb from the 3′ end of the genome, well interior to the sites of previous mutations targeted 1.6 kb from the 3′ end of the genome (11) or 2.3 kb from the 5′ end of the genome (53). Furthermore, one of the GFP recombinants (Alb151) was the product of a crossover event more than 6.8 kb from the 3′ end of the genome, as shown by the incorporation of a restriction site polymorphism close to the start of the S gene (Fig. 3). It is striking that this crossover event was also more than 6.3 kb away from the Alb4 deletion in the N gene, the phenotypic correction of which was used as the selected marker.

Our results show that targeted recombination with a synthetic donor RNA can occur significantly beyond the N gene, and this suggests that in the future we may be able to generate mutations of interest in the S gene as well as in the other structural genes, M and E. The tag inserted into gene 4 can potentially serve as a marker for an RNA hybridization-based screening procedure in future work to obtain recombinant viruses with mutations in genes located in the vicinity of the tag. By this method, we may be able to screen for low-frequency and temperature-sensitive recombinants, which cannot be selected by the Alb4 rescue-based procedure used in most previous targeted recombination. Preliminary results indicate that Alb129 can, indeed, be specifically identified against a background of wild-type virus either by plaque hybridization or by RNA dot blot analysis (data not shown). Furthermore, viable Alb129 could be recovered from replica filters of those containing hybridization-positive plaques (data not shown). Efforts to use these techniques for mutant isolation are under way.

The study reported here demonstrates for the first time that a foreign gene can be introduced into a coronavirus genome and be expressed during viral replication. An incidental consequence of this work was the expansion of the genome of MHV (31,327 nt [2]) to produce what is currently the largest known RNA virus genome (31,887 nt). MHV thus joins a catalog of other RNA viruses that have been engineered to express foreign genes (reviewed in references 3 and 44): positive-strand RNA viruses (Sindbis virus [39], Semliki Forest virus [24], and poliovirus [12]), negative-strand RNA viruses (influenza virus [26], rabies virus [32], vesicular stomatitis virus [45], and respiratory syncytial virus [6]), and retroviruses (27). The value of MHV as a recombinant gene expression vector remains to be established, but some of the properties of this virus may prove useful. Four MHV genes have been identified as nonessential for viral replication (25, 46, 55, 57) and, therefore, can serve as insertion sites. In addition, the coronavirus helical nucleocapsid structure may impose fewer packaging constraints, thereby allowing the incorporation of heterologous sequences longer than those that can be accommodated by icosahedrally symmetric nucleocapsids.

Our motivation for choosing the GFP gene for insertion into MHV was the possibility of using its activity for the identification of recombinant viruses. The usefulness of GFP as a marker protein has been shown in numerous systems (9). Since its chromophore emits an intense green fluorescence upon irradiation with long-wave UV light, we hoped to observe GFP recombinants directly in plaque assays. However, despite repeated attempts, we were unable to detect specific fluorescence in Alb148- or Alb151-infected cells either by plaque assay or in culture, using a plate reader or a fluorescence-activated cell sorter. Nevertheless, we were able to demonstrate expression of GFP by immunological assay (Fig. 7). A number of factors probably contributed to the failure to achieve a level of expression necessary for fluorescence. First, we used the wild-type GFP gene, which has a much lower fluorescence intensity than the S65T mutant of GFP (9) and does not contain the optimal codon distribution for efficient translation in higher eukaryotes (60). Second, mRNA4 is a

relatively minor species among the subgenomic mRNAs of MHV (Fig. 6). For this reason, gene 4 may not be the optimum insertion site for the expression of foreign genes. Third, mRNA4(GFP) transcription in Alb148 and Alb151 was apparently further reduced by the downstream synthesis of aberrant mRNA species.

This latter phenomenon has important implications for the mechanism of transcription in MHV. The incorporation of the GFP gene into Alb148 and Alb151 altered the spacing of the gene 4 region with respect to that of the wild-type MHV genome (Fig. 9). Although these recombinants carried out synthesis of the expectedly larger mRNA4(GFP), they also retained a propensity to synthesize a roughly equimolar amount of mRNA of the same size as wild-type mRNA4. The leader-body fusion in the larger mRNA4(GFP) was homogeneous and was identical to that of wild-type mRNA4. By contrast, the leader-body junctions of the aberrant (wild-type-size) mRNA4(GFP) species were heterogeneous and apparently random, i.e., independent of any IGS-like motifs.

A number of models have been put forth to explain the mechanism of formation of the leader-body junctions observed in coronavirus transcripts (for recent reviews, see references 22 and 54). These models can be divided according to two criteria. First, leader-body fusion may be brought about by annealing of a partially synthesized strand to a new part of the template, followed by resumption of synthesis. Alternatively, transcription may be quasi-continuous, with the polymerase copying two regions of template that have been brought together by protein-RNA and protein-protein interactions (59). Second, the fusion event may occur either during positive-strand synthesis or during negative-strand synthesis. Each model envisions a role for the IGSs that are found at the 3′ end of the leader RNA and prior to each transcription unit on the genome and that are then shared by the resulting mRNAs. Depending on the model under consideration, the IGSs (or their complements) are either templates for priming by the first part of the discontinuous subgenomic RNA or else they are polymerase launch sites (or polymerase landing sites).

The most durable earlier model of coronavirus transcription postulated that (positive-strand) subgenomic mRNAs were generated by pausing of the polymerase near the end of the leader sequence followed by detachment and reattachment to an internal portion of negative-strand template (21). The latter event was thought to be mediated by base pairing between the 3′ end of the leader and the complement of an IGS, and consequently, it was predicted that the relative abundance of the subgenomic mRNAs was determined by the amount of leader-IGS base pairing at each junction. A number of subsequent findings have led to revisions of this picture. Studies with DI RNAs or other RNA expression constructs containing authentic and mutated IGS elements have led many investigators to conclude that, beyond a minimum threshold, the extent of potential base pairing between leader and IGS complement cannot account for the efficiency of mRNA synthesis (14, 28, 52). Moreover, a recent study with bovine coronavirus has questioned whether free leader RNA actually exists (8). An additional puzzle has resulted from the discovery that all subgenomic mRNAs have negative-strand, anti-leader-containing counterparts, leading to the realization that the initial leader-body fusion event may equally well occur during negative-strand, rather than positive-strand, RNA synthesis (42, 47, 48).

A common feature of all the models is the critical role played by the IGSs, either as templates or as the sites recognized by protein factors proposed to mediate the joining of discontinuous regions of the template (59). In MHV, the consensus core IGS sequence is 5′ AAUCUAAAC 3′. With few exceptions (43, 49, 54), this entire motif is found in leader-body junctions, and all mRNAs produced at a given IGS are identical or vary only in the number of repeats that they contain of a multiple 5′ UCUAA 3′ element that immediately precedes and partially overlaps the consensus motif in the leader (29). A similar situation pertains in bovine caronavirus, without the complexity of the 5′ UCUAA 3′ repeats (15). The most intensively studied atypical IGS in MHV is that which produces mRNA2-1 (HE mRNA) in MHV-JHM, 5′ UAAACUU 3′ (49). The utilization of this IGS has been found to correlate with the number of 5′ UCUAA 3′ repeats in the leader RNA: strain variants of MHV-JHM with two repeats synthesize large amounts of mRNA2-1, while strains with three repeats express small amounts (56). Curiously, this relationship can be inverted for the utilization of other cryptic IGSs in other strains of MHV (23). Leader repeat copy number, however, does not seem to bear on the aberrant GFP mRNAs that we have discovered, since none of the junctions of these had accumulated one or more extra copies of the 5′ UCUAA 3′ element (Table 2).

By comparison with previous work, the fusion site heterogeneity reported here perhaps most resembles that seen by Zhang and Lai (58), who examined mRNAs produced by a DI RNA containing either an orthodox IGS or the atypical IGS of mRNA2-1 of MHV-JHM placed upstream of a reporter gene. A broad spectrum of fusion sites was found to result from certain combinations of helper virus and DI leader elements, the latter varying in the number of 5′ UCUAA 3′ repeats and in the presence or absence of a 9-nt sequence element immediately downstream of the leader. In the most extreme cases, fusion junctions contained base substitutions, partial duplications of leader, small deletions, or insertions of either templated or apparently nontemplated sequences. However, the anomalous fusion sites observed by these investigators, too, were correlated with variant leader RNA composition and all fusions occurred within, or adjacent to, authentic IGS motifs.

The aberrant leader-body fusion sites that we have analyzed in the gene 4-GFP recombinants differ critically from those in previous reports in at least four respects (Fig. 9). First, there were 10 distinct sites of fusion within a relatively narrow window. Second, these were quite distant (at least 495 nt downstream or at least 238 nt upstream) from the nearest authentic IGS. Third, the junctions in mRNAs produced from the authentic IGS upstream of the GFP gene were found to be canonical and homogeneous. Fourth, there were no identifiable cryptic IGSs at or near the aberrant sites of fusion in the GFP gene, and we contend that transcription from this region was IGS independent. Since cryptic IGSs are always defined post hoc (23, 43, 49, 56), it is not possible to completely disprove their existence at the aberrant fusion sites. However, for the two predominant aberrant mRNAs, there was little convincing homology between the leader RNA and either genomic fusion site (Fig. 10B) compared with that of the wild-type mRNA4 (Fig. 10A). Moreover, there were multiple loci in the GFP gene that did not serve as fusion sites but had homology to the leader sequence that was at least as good as that of the two predominant aberrant sites. Four examples of these are shown in Fig. 10C.

We are left to conclude, then, that long-range RNA or ribonucleoprotein interactions, which must be mediated by higher orders of structure, contribute strongly to the formation of the template for discontinuous transcription and that at least in the case shown here, they can be the major or sole determinant of the sites of leader-body fusion. A recent study has proposed that sequences flanking the IGS can influence the efficiency of MHV DI RNA transcription (18), but the

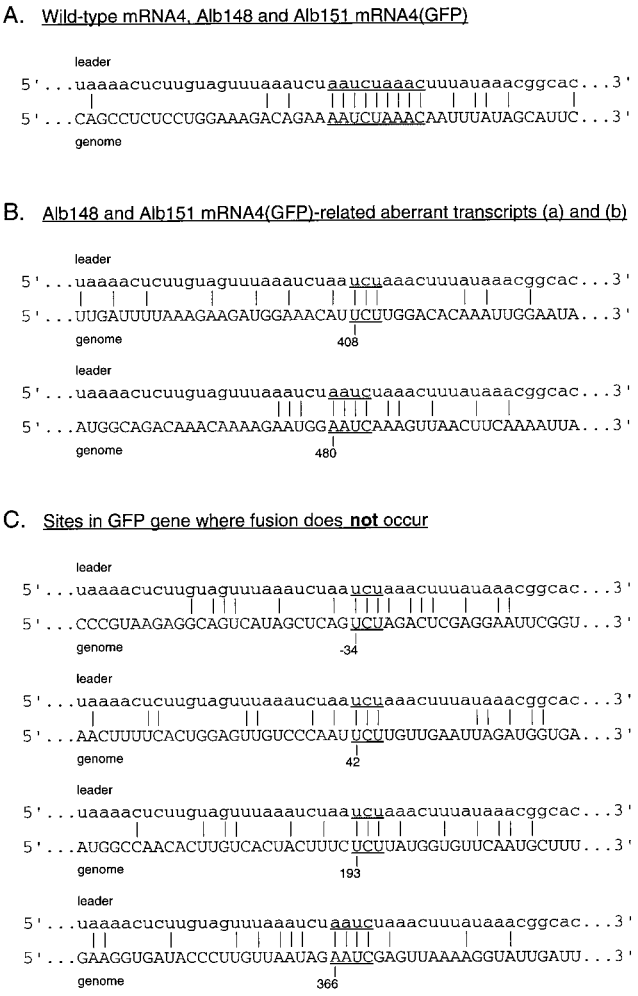

FIG. 10. Homology between the leader RNA and various sites in the genomes of wild-type or GFP recombinants. (A) Homology at the wild-type gene 4 or recombinant gene 4 (GFP) IGS. (B) Homology at the two sites within the GFP gene that correspond to the two predominant aberrant mRNA species. (C) Homology at a sampling of sites in the GFP gene that do not give rise to aberrant mRNAs. Real (A and B) or hypothetical (C) leader-body common regions are underlined. Numbers beneath genomic sequences indicate positions relative to the start of the GFP ORF.

effects reported were all suppressing and occurred in DI RNAs with extensive rearrangements relative to the genomic context of the same elements. By contrast, our results document a novel aspect of transcription that must occur over a greater distance and which has a dominant activating effect. It remains to be seen whether this property is a peculiarity of gene 4 of MHV or if it extends to all coronavirus transcriptional units.

ACKNOWLEDGMENTS

We are grateful to James Gombold for providing a clone of the MHV-A59 S gene and to Monica Parker for one of the precursor clones incorporated into pFV1. We thank Matthew Shudt and Tim Moran of the Molecular Genetics Core Facility of the Wadsworth Center for the synthesis of oligonucleotides and for automated DNA sequencing.

This work was supported in part by Public Health Service grants AI 31622 and AI 39544 from the National Institutes of Health. F.F. received support from the Fondation Mérieux.

REFERENCES

1. **Aoki, T., Y. Takahashi, K. S. Koch, H. L. Leffert, and H. Watabe.** 1996. Construction of a fusion protein between protein A and green fluorescent protein and its application to western blotting. FEBS Lett. **384:**193–197.
2. **Bonilla, P. J., A. E. Gorbalenya, and S. R. Weiss.** 1994. Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: heterogeneity among MHV strains. Virology **198:**736–740.
3. **Bredenbeek, P. J., and C. M. Rice.** 1992. Animal RNA virus expression systems. Semin. Virol. **3:**297–310.
4. **Budzilowicz, C. J., and S. R. Weiss.** 1987. In vitro synthesis of two polypeptides from a nonstructural gene of coronavirus mouse hepatitis virus strain A59. Virology **157:**509–515.
5. **Budzilowicz, C. J., S. P. Wilczynski, and S. R. Weiss.** 1985. Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3′ end of the viral mRNA leader sequence. J. Virol. **53:**834–840.
6. **Bukreyev, A., E. Camargo, and P. L. Collins.** 1996. Recovery of infectious respiratory syncytial virus expressing an additional, foreign gene. J. Virol. **70:**6634–6641.
7. **Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher.** 1994. Green fluorescent protein as a marker for gene expression. Science **263:**802–805.
8. **Chang, R.-Y., R. Krishnan, and D. A. Brian.** 1996. The UCUAAAC promoter motif is not required for high-frequency leader recombination in bovine coronavirus defective interfering RNA. J. Virol. **70:**2720–2729.
9. **Cubitt, A. B., R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross, and R. Y. Tsien.** 1995. Understanding, improving and using green fluorescent proteins. Trends Biochem. Sci. **20:**448–455.
10. **Fichot, O., and M. Girard.** 1990. An improved method for sequencing of RNA templates. Nucleic Acids Res. **18:**6162.
11. **Fischer, F., D. Peng, S. T. Hingley, S. R. Weiss, and P. S. Masters.** 1997. 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.
12. **Girard, M., R. Altmeyer, S. van der Werf, C. Wychowski, and A. Martin.** 1995. The use of picornaviruses as vectors for the engineering of live recombinant vaccines. Biologicals **23:**165–169.
13. **Gombold, J. L., S. T. Hingley, and S. R. Weiss.** 1993. Fusion-defective mutants of mouse hepatitis virus A59 contain a mutation in the spike protein cleavage signal. J. Virol. **67:**4504–4512.
14. **Hiscox, J. A., K. L. Mawditt, D. Cavanagh, and P. Britton.** 1995. Investigation of the control of coronavirus subgenomic mRNA transcription by using T7-generated negative-sense RNA transcripts. J. Virol. **69:**6219–6227.
15. **Hofmann, M. A., R.-Y. Chang, S. Ku, and D. A. Brian.** 1993. Leader-mRNA junction sequences are unique for each subgenomic mRNA species in the bovine coronavirus and remain so throughout persistent infection. Virology **196:**163–171.
16. **Horton, R. M., and L. R. Pease.** 1991. Recombination and mutagenesis of DNA sequences using PCR. *In* M. J. McPherson (ed.), Directed mutagenesis, a practical approach. IRL Press, New York, N.Y.
17. **Inouye, S., and F. I. Tsuji.** 1994. *Aequorea* green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. FEBS Lett. **341:**277–280.
18. **Jeong, Y. S., J. F. Repass, Y.-N. Kim, S.-M. Hwang, and S. Makino.** 1996. Coronavirus transcription mediated by sequences flanking the transcription consensus sequence. Virology **217:**311–322.
19. **Koetzner, C. A., M. M. Parker, C. S. Ricard, L. S. Sturman, and P. S. Masters.** 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.
20. **Konings, D. A. M., P. J. Bredenbeek, J. F. H. Noten, P. Hogeweg, and W. J. M. Spaan.** 1988. Differential premature termination of transcription as a proposed mechanism for the regulation of coronavirus gene expression. Nucleic Acids Res. **16:**10849–10860.
21. **Lai, M. M. C.** 1990. Coronavirus: organization, replication and expression of genome. Annu. Rev. Microbiol. **44:**303–333.
22. **Lai, M. M. C., C.-L. Liao, Y.-J. Lin, and X. Zhang.** 1994. Coronavirus: how a large RNA viral genome is replicated and transcribed. Infect. Agents Dis. **3:**98–105.
23. **La Monica, N., K. Yokomori, and M. M. C. Lai.** 1992. Coronavirus mRNA synthesis: identification of novel transcription initiation signals which are differentially regulated by different leader sequences. Virology **188:**402–407.
24. **Liljeström, P., S. Lusa, D. Huylebroeck, and H. Garoff.** 1991. In vitro mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. J. Virol. **65:**4107–4113.
25. **Luytjes, W., P. J. Bredenbeek, A. F. H. Noten, M. C. Horzinek, and W. J. M. Spaan.** 1988. Sequence of mouse hepatitis virus A59 mRNA2: indications for RNA recombination between coronaviruses and influenza C virus. Virology **166:**415–422.
26. **Luytjes, W., M. Krystal, M. Enami, J. D. Parvin, and P. Palese.** 1989.

Amplification, expression, and packaging of a foreign gene by influenza virus. Cell **59:**1107–1113.
27. **Majors, J. E.** 1992. Retroviral vectors—strategies and applications. Semin. Virol. **3:**285–295.
28. **Makino, S., and M. Joo.** 1993. Effect of intergenic consensus sequence flanking sequences on coronavirus transcription. J. Virol. **67:**3304–3311.
29. **Makino, S., L. H. Soe, C.-K. Shieh, and M. M. C. Lai.** 1988. Discontinuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs. J. Virol. **62:**3870–3873.
30. **Masters, P. S.** 1992. Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus. Arch. Virol. **125:**141–160.
31. **Masters, P. S., C. A. Koetzner, C. A. Kerr, and Y. Heo.** 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.
32. **Mebatsion, T., M. J. Schnell, J. H. Cox, S. Finke, and K.-K. Conzelmann.** 1996. Highly stable expression of a foreign gene from rabies virus vectors. Proc. Natl. Acad. Sci. USA **93:**7310–7314.
33. **Ormö, M., A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, and S. J. Remington.** 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. Science **273:**1392–1395.
34. **Parker, M. M., and P. S. Masters.** 1990. Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. Virology **179:**463–468.
35. **Peng, D., C. A. Koetzner, and P. S. Masters.** 1995. 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.
36. **Peng, D., C. A. Koetzner, T. McMahon, Y. Zhu, and P. S. Masters.** 1995. Construction of murine coronavirus mutants containing interspecies chimeric nucleocapsid proteins. J. Virol. **69:**5475–5484.
37. **Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier.** 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene **111:**229–233.
38. **Prendergast, F. G., and K. G. Mann.** 1978. Chemical and physical properties of aequorin and the green fluorescent protein isolated from *Aequorea forskålea*. Biochemistry **17:**3448–3453.
39. **Rice, C. M., R. Levis, J. H. Strauss, and H. V. Huang.** 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutations. J. Virol. **61:**3809–3819.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
42. **Sawicki, S. G., and D. L. Sawicki.** 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. J. Virol. **64:**1050–1056.
43. **Schaad, M. C., and R. S. Baric.** 1993. Evidence for new transcriptional units encoded at the 3′ end of the mouse hepatitis virus genome. Virology **196:** 190–198.
44. **Schlesinger, S.** 1995. RNA viruses as vectors for the expression of heterologous proteins. Mol. Biotechnol. **3:**155–165.
45. **Schnell, M. J., L. Buonocore, M. A. Whitt, and J. K. Rose.** 1996. The minimal conserved transcription stop-start signal promotes stable expression of a foreign gene in vesicular stomatitis virus. J. Virol. **70:**2318–2323.
46. **Schwarz, B., E. Routledge, and S. G. Siddell.** 1990. Murine nonstructural protein ns2 is not essential for virus replication in transformed cells. J. Virol. **64:**4784–4791.
47. **Sethna, P. B., M. A. Hofmann, and D. A. Brian.** 1991. Minus-strand copies of replicating coronavirus mRNAs contain antileaders. J. Virol. **65:**320–325.
48. **Sethna, P. B., S.-L. Hung, and D. A. Brian.** 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. Proc. Natl. Acad. Sci. USA **86:**5626–5630.
49. **Shieh, C.-K., H.-J. Lee, K. Yokomori, N. La Monica, S. Makino, and M. M. C. Lai.** 1989. Identification of a new transcription initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome. J. Virol. **63:**3729–3736.
50. **Siddell, S. G.** 1995. The Coronaviridae: an introduction. *In* S. G. Siddell (ed.), The Coronaviridae. Plenum Press, New York, N.Y.
51. **Skinner, M. A., and S. G. Siddell.** 1985. Coding sequence of coronavirus MHV-JHM mRNA4. J. Gen. Virol. **66:**593–596.
52. **Van der Most, R. G., R. J. de Groot, and W. J. M. Spaan.** 1994. Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus: a study of coronavirus transcription initiation. J. Virol. **68:** 3656–3666.
53. **Van der Most, R. G., L. Heijnen, W. J. M. Spaan, and R. J. de Groot.** 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.
54. **Van der Most, R. G., and W. J. M. Spaan.** 1995. Coronavirus replication, transcription, and RNA recombination, p. 11–31. *In* S. G. Siddell (ed.), The Coronaviridae. Plenum Press, New York, N.Y.
55. **Weiss, S. R., P. W. Zoltick, and J. L. Leibowitz.** 1993. The ns 4 gene of mouse hepatitis virus (MHV), strain A59 contains two ORFs and thus differs from ns 4 of the JHM and S strains. Arch. Virol. **129:**301–309.
56. **Yokomori, K., L. R. Banner, and M. M. C. Lai.** 1991. Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. Virology **183:**647–657.
57. **Yokomori, K., and M. M. C. Lai.** 1991. Mouse hepatitis virus S RNA sequence reveals that nonstructural proteins ns4 and ns5a are not essential for murine coronavirus replication. J. Virol. **65:**5605–5608.
58. **Zhang, X., and M. M. C. Lai.** 1994. Unusual heterogeneity of leader-mRNA fusion in a murine coronavirus: implications for the mechanism of RNA transcription and recombination. J. Virol. **68:**6626–6633.
59. **Zhang, X., C.-L. Liao, and M. M. C. Lai.** 1994. Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription both in *trans* and in *cis*. J. Virol. **68:**4738–4746.
60. **Zolotukhin, S., M. Potter, W. W. Hauswirth, J. Guy, and N. Muzyczka.** 1996. A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. J. Virol. **70:**4646–4654.