

Development of Mouse Hepatitis Virus and SARS-CoV Infectious cDNA Constructs

R. S. Baric (✉) · A. C. Sims

Department of Epidemiology, University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599–7400, USA
rbaric@email.unc.edu

1	Introduction	230
2	The Coronavirus Genome	230
3	Systematic Approaches to Assembling Coronavirus cDNAs from a Panel of Contiguous Subclones	231
4	Assembling MHV Infectious cDNAs	235
4.1	Applications in Genomics	240
4.2	Engineering MHV Genomes	241
5	SARS-CoV Infectious Clone	245
6	Future Applications	246
	References	248

Abstract The genomes of transmissible gastroenteritis virus (TGEV) and mouse hepatitis virus (MHV) have been generated with a novel construction strategy that allows for the assembly of very large RNA and DNA genomes from a panel of contiguous cDNA subclones. Recombinant viruses generated from these methods contained the appropriate marker mutations and replicated as efficiently as wild-type virus. The MHV cloning strategy can also be used to generate recombinant viruses that contain foreign genes or mutations at virtually any given nucleotide. MHV molecular viruses were engineered to express green fluorescent protein (GFP), demonstrating the feasibility of the systematic assembly approach to create recombinant viruses expressing foreign genes. The systematic assembly approach was used to develop an infectious clone of the newly identified human coronavirus, the severe acute respiratory syndrome virus (SARS-CoV). Our cloning and assembly strategy generated an infectious clone within 2 months of identification of the causative agent of SARS, providing a critical tool to study coronavirus pathogenesis and replication. The availability of coronavirus infectious cDNAs heralds a new era in coronavirus genetics and genomic applications, especially within the replicase proteins whose functions in replication and pathogenesis are virtually unknown.

1 Introduction

Molecular analysis of the structure and function of RNA virus genomes has been profoundly advanced by the availability of full-length cDNA clones, the source of infectious RNA transcripts that replicate efficiently when introduced into permissive cell lines (Boyer and Haenni 1994). Coronaviruses contain the largest single-stranded, positive-polarity RNA genome of about 30 kb (Cavanagh et al. 1997; de Vries et al. 1997; Eleouet et al. 1995). Until recently, coronavirus genetic analysis has been limited to analysis of temperature-sensitive (ts) mutants (Fu and Baric 1992, 1994; Lai and Cavanagh 1997; Schaad and Baric 1994; Stalcup et al. 1998), defective interfering (DI) RNAs (Izeta et al. 1999; Narayanan and Makino 2001; Repass and Makino 1998; Williams et al. 1999), and recombinant viruses generated by targeted recombination (Fischer et al. 1997; Hsue and Masters 1999; Kuo et al. 2000). Among these, targeted recombination is the seminal approach developed to systematically assess the function of individual mutations in the 3'-most ~10 kb of the MHV genome. Methods to assemble an MHV full-length infectious construct have been hampered by the large size of the genome, the regions of chromosomal instability, and the inability to synthesize full-length transcripts (Almazán et al. 2000; Masters 1999; Yount et al. 2000). This is especially problematic within the group 2 coronavirus replicase, where several regions of chromosomal toxicity and instability have hampered the development of infectious cDNAs. Full-length infectious constructs will allow for the systematic dissection of the structure and function of each viral gene, the phenotypic consequences of gene rearrangement on virus replication and pathogenesis, the development of coronavirus heterologous gene expression systems, and a clearer understanding of the transcription and replication strategy of the *Coronaviridae*. In this report, we review strategies for building coronavirus infectious cDNAs by using mouse hepatitis virus strain A59 as a model.

2 The Coronavirus Genome

The coronavirus genome, a single-stranded RNA, is the largest viral RNA genome known to exist in nature (27.6–31.3 kb). Genomic RNAs have a 5' terminal cap and a 3' terminal poly (A) tail. In addition, a leader sequence of 65–98 nucleotides and a 200- to 400-base pair untranslated region are located at the 5' terminus, whereas a 200- to 500-base pair

untranslated region is located at the 3' terminus. The 5' most two-thirds of the genome encodes the replicase gene in two open reading frames (ORFs), 1a and 1b, the latter of which is expressed by ribosomal frameshifting (Almazán et al. 2000; Eleouet et al. 1995). Like many other positive-sense RNA viruses, the coronavirus replicase is translated as a large precursor polyprotein that is processed by viral proteinases, giving rise to ~15 replicase proteins. The functions of most of the coronavirus replicase proteins are unknown. However, based on nucleotide sequence homology and empirical studies, identifiable functions include two papainlike cysteine proteases, a chymotrypsin-like 3C protease, a cysteine-rich growth factor-related protein, an RNA-dependent RNA polymerase, a nucleoside triphosphate (NTP)-binding/helicase domain, and a zinc-finger nucleic acid-binding domain (Enjuanes et al. 2000a; Penzes et al. 2001; Siddell 1995). Most of the replicase gene products colocalize with replication complexes at sites of RNA synthesis on internal membranes. However, a spectrum of genetically informative mutations have not been systematically targeted to any of these replicase proteins, so we have little insight into the organization of the replicase complex and the location of functional motifs, which regulate transcription, replication, and RNA recombination. Because of the extremely rich milieu of molecular reagents that are available against the replicase proteins, the availability of a molecular clone of MHV allows for the first time a systematic genetic analysis of gene 1 function in coronavirus replication.

3

Systematic Approaches to Assembling Coronavirus cDNAs from a Panel of Contiguous Subclones

Coronavirologists have seized on several different strategies to build infectious cDNA clones. However, all were primarily designed to circumvent problems associated with the large size of the coronavirus genome, regions of chromosomal instability, and other problems associated with the production of full-length infectious transcripts (Almazán et al. 2000; Masters 1999; Yount et al. 2000). Our solution was to assemble infectious cDNAs from a panel of contiguous subclones that spanned the entire length of the TGEV and MHV genomes. Each subclone was flanked by unique restriction sites with characteristics that allow for the systematic and precise assembly of a full-length cDNA with *in vitro* ligation. For this strategy to be efficient, restricted subclone fragments had to be incapable of self-concatemer formation and not spuriously assemble with other noncontiguous subclones.

Conventional class II restriction enzymes, such as *EcoRI*, leave identical sticky ends that assemble with similarly cut DNA in the presence of DNA ligase (Pingoud and Jeltsch 2001; Sambrook et al. 1989). Because these enzymes leave identical compatible ends, digested fragments randomly self-assemble into large concatamers and, therefore, they are poor choices for assembling large intact genomes or chromosomes. However, a second group of class II restriction enzymes (i.e., *BglI*, *BstXI*, *SfiI*) also recognize a symmetrical sequence but leave random sticky ends 1–4 nucleotides in length, and consequently, restrict assembly cascades along specific pathways (Table 1). For example, the type II restriction enzyme, *BglI*, recognizes the symmetrical sequence GCCNNNN↓NGGC and cleaves a random DNA sequence on average every ~4,096 base pairs. Because 64 different 3-nucleotide overhangs can be generated, DNA frag-

Table 1. Selected restriction enzymes used in assembly of recombinant full-length genomes

Restriction enzyme ^a	Recognition site	No. of variable sticky end	Average cutting frequency ^b	Actual frequency of compatible ends ^b
<i>BglI</i>	GCCNNNN↓NGGC	3 nt/64 potential ends	~4,096 nt	~261,344 nt
<i>BstXI</i>	CGGN↑NNNNCCG CCANNNNN↓NTGG	4 nt/256 potential ends	~4,096 nt	~1,045,376 nt
<i>SfiI</i>	GGTN↑NNNNNACC GGCCNNNN↓NGGCC	3 nt/64 potential ends	~65,536 nt	~4,194,304 nt
<i>SapI</i>	CCGGN↑NNNNCCGG GCTCTCN↓NNNN	3 nt/64 potential ends	~16,385 nt (in either strand)	~1,048,640 nt*
<i>AarI</i>	CGAGAAGNNNNN↑ CACCTGCNNNN↓NNNN	4 nt/256 potential ends	~16,385 nt (in either strand)	~4,194,304 nt*
<i>Esp3I</i> (<i>BsmBI</i>)	GTGGACGNNNNNNN↑ CGTCTCN↓NNNN GCAGAGNNNNN↑	4 nt/256 potential ends	~4,096 nt (in either strand)	~1,048,576 nt*

^a Other enzymes leaving many different overhangs: *BsmFI*, *EclHkI*, *FokI*, *MboII*, *TthIII*, *AhdI*, *DrdI*, *BspMI*, *BsmAI*, *BcgI*, *BmRI*, *BpmI*, *BsaI*, *BseI*, *EarI*, *PfiMI*, *BstV2*, *VpaK32I*, *AbeI*, *PpiI*.

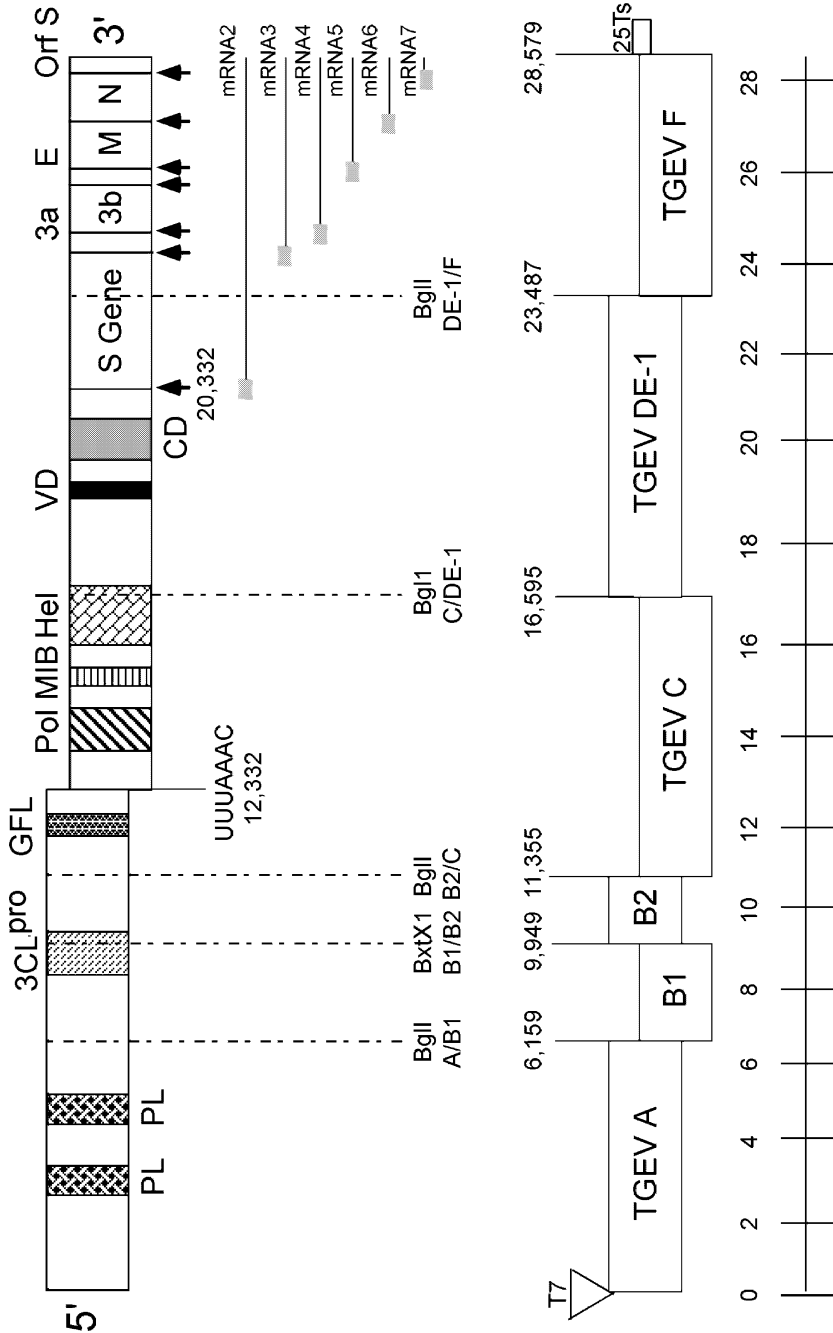
^b Assuming a totally random DNA sequence; *asymmetric cutters like *SapI*, *AarI* and *Esp3I* can have recognition sites in either strand of DNA so actual site frequency is ~1/2 of indicated values and can be engineered as “no-see-um” (Yount et al. 2002).

ments will only assemble with the appropriate 3-nucleotide complementary overhang generated at an identical *Bgl*I restriction site. As a result, identical ends are generated every ~264,000 base pairs, providing a powerful means for the construction of very large DNA and RNA genomes. Consonant with these findings, the type IIS restriction enzyme, *Esp*3I, recognizes an asymmetric sequence and makes a staggered cut 1 and 5 nucleotides downstream of the recognition sequence, leaving 256, mostly asymmetrical, 4-nucleotide overhangs (GCTCTCN[↓]NNNN). As identical *Esp*3I sites are generated every ~1,000,000 base pairs or so in a random DNA sequence, most restricted fragments usually do not self-assemble (Yount et al. 2002). Rather, specific recursive assembly pathways can be designed that hypothetically allow assembly of >1 million base pair DNA genomes (~2²⁵⁶ fragments) (Table 1). We took advantage of several unique properties inherent in type II restriction enzymes to build coronavirus infectious cDNAs.

Initially, we isolated five cDNA subclones spanning the entire TGEV genome (designated TGEV A, B, C, D/E, and F) by RT-PCR using primers that introduced unique *Bgl*I restriction sites at the 5' and 3' ends of each fragment without altering the amino acid coding sequences of the virus (Table 2). The TGEV A, C, DE, and F clones were stable in plasmid DNAs in *Escherichia coli*. The B fragment, however, was unstable, containing deletions or insertions in the wild-type sequence at a region of instability in the TGEV genome noted by other investigators (Almazán et al. 2000; Eleouet et al. 1995). To prevent fragment instability, we used primer-mediated mutagenesis to bisect the B fragment at the unstable site with an adjoining *Bst*XI (CCATTCAC[↓]TTGG) site, resulting in TGEV B1 and TGEV B2 amplicons (Fig. 1; Table 2). It is likely that sequences

Table 2. Design of TGEV junction sequences

Restriction site junction	Location	Junction
5'-GCCTGTT [↓] TGGC-3'	<i>Bgl</i> I, nt 6,159	A-B1
3'-CGGA ₇ CAAACCG-5'		
5'-CCATTCAC [↓] TTGG-3'	<i>Bst</i> XI, nt 9,949	B1-B2
3'-GGTA ₇ AGTGAACC-5'		
5'-GCCGCAT [↓] TGGC-3'	<i>Bgl</i> I, nt 11,355	B2-C
3'-CGGC ₇ GTAGCCG-5'		
5'-GCCTTCT [↓] TGGC-3'	<i>Bgl</i> I, nt 16,595	C-D/E1
3'-CGGA ₇ AGAACCG-5'		
5'-GCCGTGC [↓] AGGC-3'	<i>Bgl</i> I, nt 23,487	D/E1-F
3'-CGGC ₇ ACGTCCG-5'		



(9600–9950) in and around the TGEV 3C like protease (3CL^{pro}) motif are either bactericidal or unstable in microbial vectors (Almazán et al. 2000; Yount et al. 2000). The resulting 6 fragments, TGEV A, B1, B2, C, D/E, and F, were ligated in vitro to generate a full-length cDNA of the TGEV genome (Fig. 1). Molecularly cloned viruses were indistinguishable from wild type and contained the marker mutations and unique *Bgl*I and *Bst*XI junction sequences used in the assembly of the infectious construct (Yount et al. 2000).

4 Assembling MHV Infectious cDNAs

One potential problem with the original approach was that several “silent” mutations were inserted to introduce the unique *Bgl*I sites into the TGEV component clones. To circumvent this problem, a variation of the systematic assembly approach was used to build the group II coronavirus, mouse hepatitis virus (MHV) infectious cDNA (Yount et al. 2002). The enzyme *Esp*3I recognizes an asymmetrical site and cleaves external to the recognition sequence, allowing for traditional and “no-see-um” cloning applications (Fig. 2, Table 1). With traditional approaches, *Esp*3I sites can be oriented to reform the recognition site after ligation of two MHV cDNAs, leaving the restriction site within the genomes of recombinant viruses. However, the *Esp*3I recognition site is asymmetrical, so a simple reverse orientation allows for the insertion of an *Esp*3I recognition sequence on the ends of two adjacent clones with the cleavage site derived from virtually any 4-nucleotide sequence combination dictated by the virus sequence. On cleavage and ligation with the adjoining fragment, the *Esp*3I sites are lost from the final ligation products, leaving a



Fig. 1. Strategy for the systematic assembly of TGEV full-length cDNA. The TGEV genome is a positive-sense, single-stranded RNA of about 28.5 kb. Six independent subclones (*A*, *B1*, *B2*, *C*, *DE*, and *F*) that span the entire length of the genome were isolated by RT-PCR using primer pairs that introduced unique *Not*I, *Bgl*I, and/or *Bst*XI restriction sites at each end. On ligation, the intact viral genome is generated as a cDNA. A unique T7 start site and a 25 poly(T) tail allow for in vitro transcription of full-length, capped, polyadenylated transcripts (Yount et al. 2000). *PL*, papainlike protease; *3CL^{pro}*, 3CL protease; *GFL*, growth factor like; *pol*, polymerase motif; *MIB*, metal binding motif; *hel*, helicase motif; *VD/CD*, variable or conserved domains

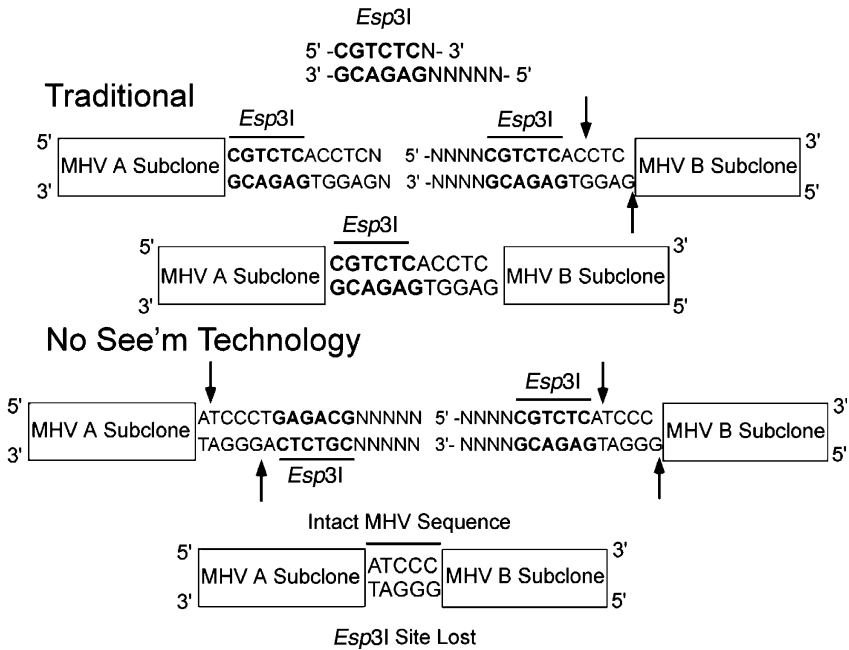


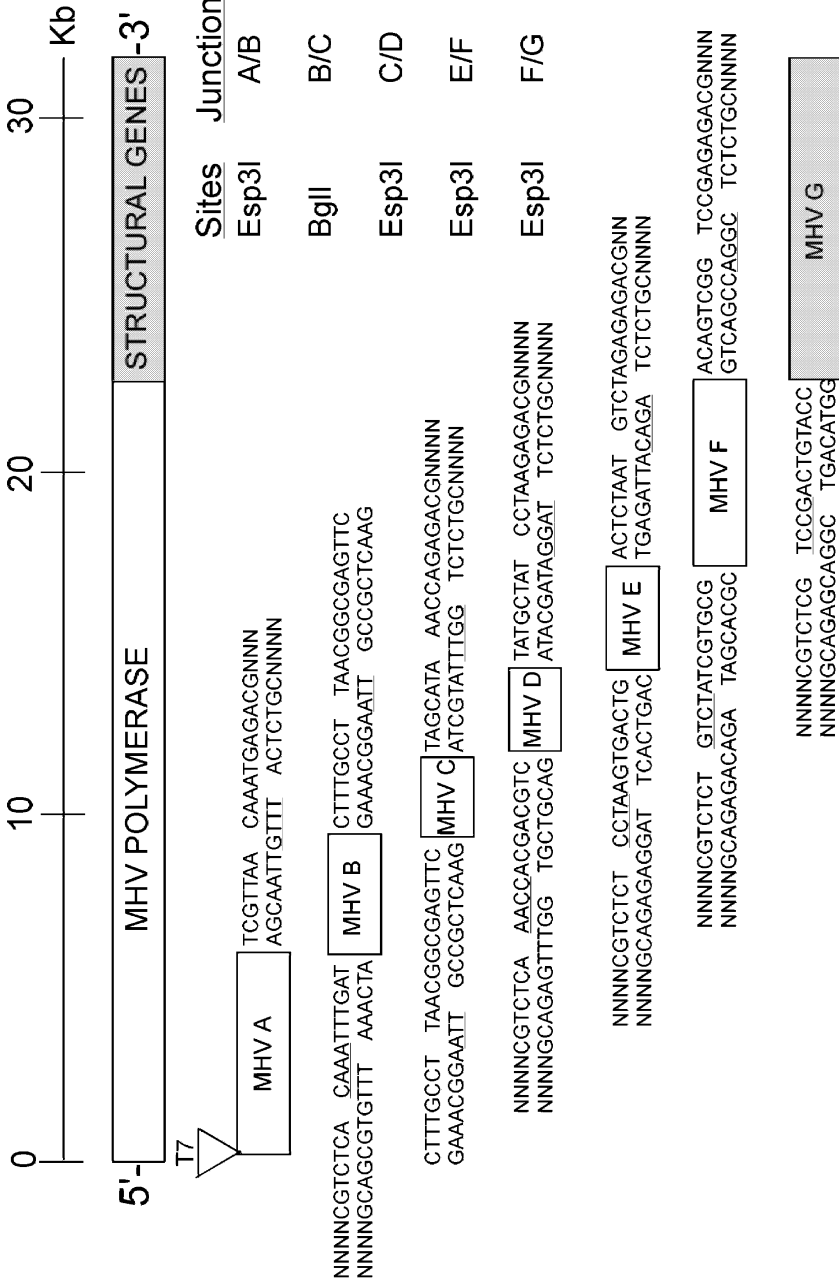
Fig. 2. Use of *Esp31* in the traditional and “no-see-um” approaches. The traditional approach to the use of *Esp31* involves the ligation of two fragments containing identical *Esp31* restriction sites, resulting in a ligation product with an intact *Esp31* site remaining. In the “no-see-um” approach, a simple reverse orientation of the restriction sites allows for the specific removal of the *Esp31* site from the two fragments, resulting in a ligation product lacking the engineered restriction site. The use of the “no-see-um” technology allows for the assembly of large DNAs from smaller subclones without the incorporation of unique restriction sites into the genome. (Yount et al. 2002)

seamless junction compiled from the exact MHV-A59 sequence. Because of this property, unique junctions can be inserted at virtually any position between two component clones without mutating the viral genome sequence. Additionally, a large number of other restriction enzymes share this property (e.g., *SapI*, *AarI*), expanding the utility of the “no-see-um” technology (Table 1).

During the isolation of the MHV component clones, it was also necessary to remove three preexisting *Esp31* sites located throughout the MHV ORF1 sequence (Bonilla et al. 1994). Mutations inserted to ablate these sites were used as marker mutations to distinguish molecularly

cloned and wild-type virus. We then isolated seven consensus cDNAs that spanned the entire length of the MHV-A59 genome in the same manner as the TGEV infectious construct (Fig. 3). This was necessary because the MHV-A59 genome contains several major regions of sequence toxicity in microbial cloning vectors, most of which map between ~10 and 15 kb in the MHV ORF 1a/ORF 1b polyprotein and an unstable region mapping ~5.0 kb in ORF 1a. As described for the TGEV B fragment, cDNAs were isolated after intersecting the toxic domains and separating them into independent subclones. However, many subclones were still unstable in traditional PUC-based cloning vectors (e.g., pGem, TopoII) even when maintained at low temperature. Consequently, we used pSMART cloning vectors (Lucigen), which lack a promoter and indicator gene and contain transcriptional and translational terminators surrounding the cloning site. Instability appears to be associated with expression, as this entire domain (nucleotides 9,555–15,754) is also stable in yeast vectors (pYES2.1 Topo TA Cloning Kit from Invitrogen) that maintain tight regulation over foreign gene expression (Yount et al., unpublished results). Full-length MHV-A59 cDNA was systematically assembled through the simultaneous *in vitro* ligation of a series of seven subgenomic cDNAs (Yount et al. 2002). In the future, it may be possible to construct larger subgenomic fragments spanning the entire genome by using the pSMART cloning vectors, thereby simplifying the assembly strategy, although we have not tested this directly.

The TGEV and MHV A fragments contain a T7 promoter, whereas the TGEV F and MHV G fragments terminate in a poly(T) tract at the 3' end, allowing for *in vitro* T7 transcription of infectious capped, polyadenylated transcripts. The poly(A) tails generated from these transcripts are 25 nucleotides in length, which appears sufficient for transcript infectivity. At this time, we do not know the minimal number of 3' poly(A) residues necessary for transcript infectivity or whether a 5' methylated cap is essential. Electroporation of the genomic-length RNAs resulted in the production of recombinant MHV virus with growth characteristics identical to those of the wild-type viruses (Yount et al. 2000, 2002). Importantly, the molecularly cloned viruses contained marker mutations engineered into the component clones. Inclusion of nucleocapsid(N)-encoding transcripts enhanced the infectivity of full-length MHV and TGEV transcripts. In MHV, N transcripts enhanced the infectivity of full-length MHV-A59 transcripts by 10- to 15-fold as evidenced by increased viral antigen expression and virus titers at 25 h postinfection (Yount et al. 2002). It is unclear whether MHV N transcripts, N protein, or both are essential for increased virus yields after electroporation, or



whether this effect would be observed with transcripts encoding unrelated genes. Coronaviruses have been demonstrated to package low concentrations of subgenomic mRNAs, especially N transcripts, and several studies have suggested that N may function in transcription and replication and are tightly associated with the replication complex. With IBV, but not TGEV or HCoV-229E, N transcripts are absolutely essential for full-length transcript infectivity (Casais et al. 2001). With HCoV-229E, other groups have shown that the N gene is not required for subgenomic transcription (Thiel et al. 2001). Clearly, additional studies are needed to evaluate the role of N protein in RNA transcript infectivity.

The MHV cDNA cassettes can be ligated systematically as described for TGEV or simultaneously. Although numerous incomplete assembly intermediates were evident, our demonstration that simultaneous ligation of seven cDNAs will result in full-length cDNA will simplify the complexity of the assembly strategy. At this time, there is no evidence to indicate that this approach might introduce spurious mutations or genome rearrangements from aberrant assembly cascades. However, it is possible that such variants might arise after RNA transfection, as a consequence of high-frequency MHV RNA recombination between incomplete and genome-length transcripts. It is likely that such variants would be replication impaired and rapidly out-competed by wild-type virus. A second limitation is that the yield of full-length cDNA product is reduced, resulting in less robust transfection efficiencies compared with the more traditional systematic assembly method. At this time, the MHV approach suffers from the large number of component clones (seven), which increase the complexity of the system and reduce the yield of full-length cDNA product after *in vitro* ligation. If the large number of toxic domains in the MHV genome is duplicated in other group II coronaviruses, this will likely interfere with the development of other infec-



Fig. 3. Systematic assembly strategy for the construction of MHV-A59 full-length cDNA. The MHV genome is a positive-sense, single-stranded RNA of ~31.5 kb. Seven independent subclones (A, B, C, D, E, F, and G) that span the entire MHV genome were isolated by RT-PCR. Unique *Bgl*I and *Esp*3I restriction sites, located at the 5' and 3' ends of each subclone, were used to assemble a full-length cDNA. A unique T7 start site was inserted at the 5' end of the MHV A fragment and a 25 poly(T) tail was inserted at the 3' end of the MHV F fragment, allowing for *in vitro* transcription of full-length, capped, poly-adenylated transcripts. Note: *Esp*3I sites are lost in the assembly process. (Yount et al. 2002)

tious cDNAs as well. Topics of future research include: (1) Can group II coronavirus cDNAs be stabilized as full-length constructs in bacterial artificial chromosomes or poxvirus vectors as has been reported with TGEV, IBV, and HCoV 229E? (2) How does N function to enhance infectivity of full-length transcripts? (3) How can we enhance yields or the infectivity of coronavirus infectious cDNAs and transcripts and allow for critical review of the consequences of lethal mutations? (4) Can we reduce the number of component clones needed to assemble group II coronavirus infectious cDNAs?

4.1

Applications in Genomics

Our assembly strategy for coronavirus infectious constructs is simple and straightforward, although the synthesis of full-length transcripts is technically challenging. In contrast to infectious clones of other positive-strand viruses, our TGEV and MHV constructs must be assembled *de novo* and do not exist intact in bacterial or viral vectors. This does not restrict the method's applicability for reverse genetic applications. Rather, it allows for rapid genetic manipulation of independent sub-clones, which minimizes the introduction of spurious mutations elsewhere in the genome during recombinant DNA manipulation. Theoretical limits of our method may exceed several million base pairs of DNA and will likely surmount the cloning capacity of bacterial (BAC) and eukaryotic artificial chromosome vectors (Grimes and Cooke 1998). Our systematic assembly method should also be appropriate for constructing full-length infectious clones of other large RNA viruses, including coronaviruses (27–32 kb), toroviruses (24–27 kb), and filoviruses like Marburg (19 kb) (de Vries et al. 1997; Peters et al. 1996). Viral genomes that are unstable in prokaryotic vectors can also be cloned by these methods (Boyer and Haenni 1994; Rice et al. 1989). Moreover, the technique should allow the systematic assembly of full-length infectious dsDNA genomes of adenoviruses, herpesviruses, and perhaps other large DNA viruses that promise to be powerful tools in vaccination, gene transfer, and gene therapy (Smith and Enquist 2000; van Zijl et al. 1988). Recently, genome sequences from a large number of prokaryotic and eukaryotic organisms have been obtained, providing significant insight into gene organization, structure, and function (Cho et al. 1999; Hutchison et al. 1999) (TIGR homepage <http://www.tigr.org>). Using this strategy, it may be possible to reconstruct a minimal microbial genome from the bottom up. However, problems associated with isolating large DNA fragments

and the introduction of large DNA genomes into environments that permit replication will likely be significant hurdles. Nevertheless, our assembly strategy may provide a means to analyze the function of large blocks of DNA, such as pathogenesis islands, or to engineer chromosomes that contain large gene cassettes of interest (Cho et al. 1999).

4.2 Engineering MHV Genomes

Coronaviruses provide a unique system for the incorporation and expression of one or more foreign genes (Enjuanes and Van der Zeijst 1995). Coronavirus genes rarely overlap, simplifying the design and expression of foreign genes from downstream intergenic sequences (IS) start sites. Integration of the coronavirus RNA genome into the host cell chromosome is unlikely (Lai and Cavanagh 1997). Additionally, recombinant viruses or replicon particles could be readily targeted to other mucosal surfaces in swine or to other species by simple replacements in the S glycoprotein gene, which has been shown to determine tissue- and species tropism (Ballesteros et al. 1997; Delmas et al. 1992; Kuo et al. 2000; Leparc-Goffart et al. 1998; Sánchez et al. 1999; Tresnan et al. 1996). Furthermore, coronaviruses infect a number of different species, including human, porcine, bovine, canine, and feline, and are available for the development of expression systems (Sánchez et al. 1992). Additionally, the coronavirus helical ribonucleocapsid structure may further relax the packaging constraints of the virus, as compared to icosahedral structures (Enjuanes and Van der Zeijst 1995; Lai and Cavanagh 1997; Risco et al. 1996). Selected questions that remain unanswered include: (1) What is the coding capacity of coronavirus based expression systems? (2) What is the minimal genome required for efficient replication? (3) Can high-titer coronavirus replicon particles be obtained for vaccine applications? (4) What are the minimal sequence requirements for subgenomic transcription? (5) How many foreign genes can be coordinately regulated without impeding virus replication or immunogenicity? (6) What are the efficacy, stability, and safety of the recombinant coronaviruses in natural settings? Clearly, these vaccine-related topics will provide fruitful avenues of investigation over the next decade and will greatly enhance our understanding of the mechanics of coronavirus transcription, replication, assembly and release, and pathogenesis.

The future development of vaccines and expression vectors are particularly intriguing applications of our TGEV and MHV infectious clones. Importantly, at least two TGEV downstream ORFs encode luxury func-

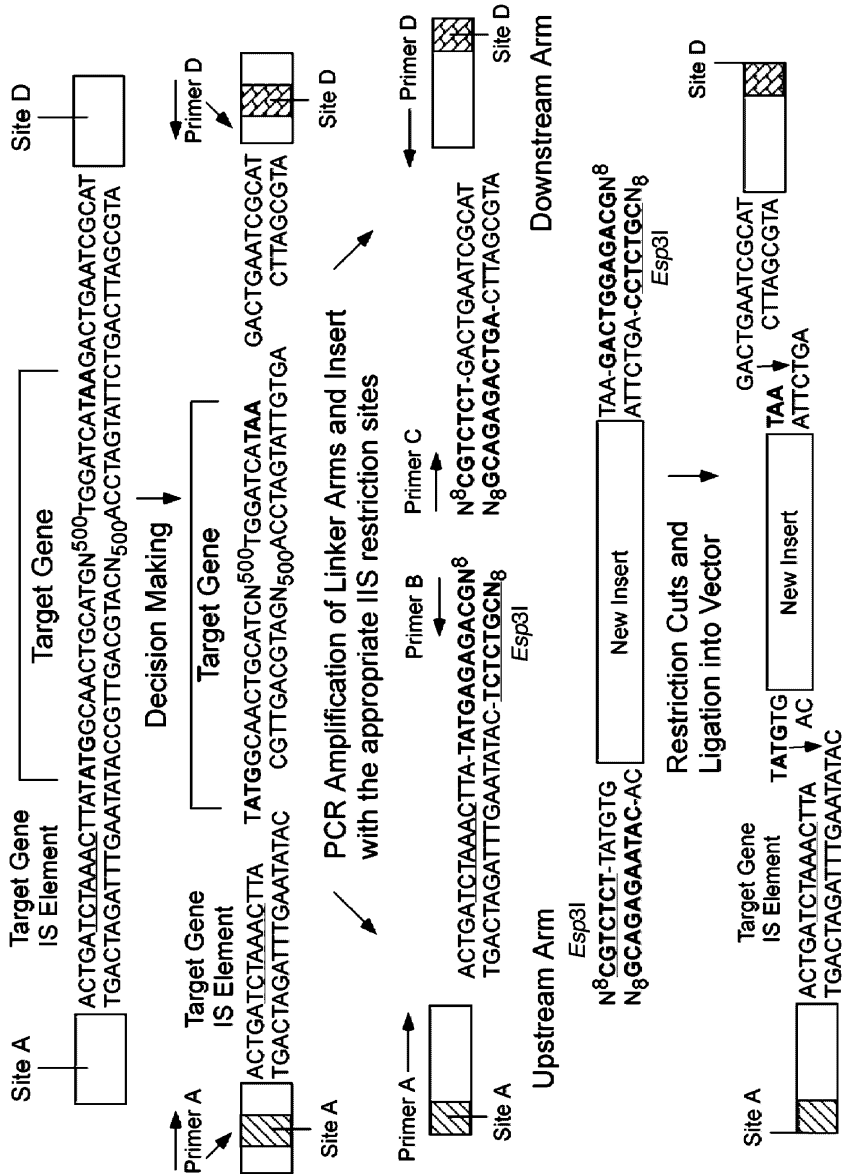


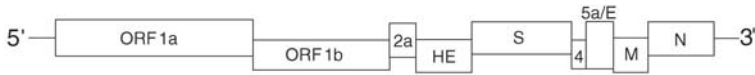
Fig. 4. Rapid mutagenesis of the MHV infectious cDNA with Class IIS restriction endonucleases. Seamless insertion of foreign genes into the coronavirus genome can be accomplished with Class IIS restriction enzymes. In this case, a target gene is systematically removed and replaced by a new gene (*new insert*). Using a primer with overlaps a unique upstream (*Site A*) restriction site, the upstream arm amplicon is

tions (ORF 3a and 3b) that may be deleted from the viral genome without impacting infectivity (Curtis et al. 2002; Laude et al. 1990; McGoldrick et al. 1999; Wesley et al. 1991). We have developed a rapid approach that allows seamless insertion of foreign sequences into virtually any nucleotide position in the MHV genome, based on class IIS restriction endonucleases (Fig. 4). In this approach, flanking sequences around the target domain are amplified as separate arms linked by unique class IIS restriction site oriented as described in Fig. 3. A third amplicon encoding the payload sequence of interest is isolated and flanked by similar class IIS sites. After restriction digestion and ligation, the foreign sequences are inserted into the backbone sequence at any given nucleotide, leaving no evidence of the restriction sites that were used to “sew” the new sequences into the MHV backbone. We have successfully expressed GFP from the ORF 3a locus of TGEV (Curtis et al. 2002) and ORF 4 of MHV (Fig. 5) (manuscript in preparation), demonstrating the feasibility of the method and the use of TGEV and MHV as expression vectors. In the case with TGEV, GFP expression was stable for at least 10 passages. In addition, we have removed the ORF 3a and replaced it with GP5 of PRRSV to create icTGEV PRRSV GP5 recombinant viruses (Curtis KM and Baric RS, unpublished data). Recombinant viruses expressed the PRRSV GP5 glycoprotein as evidenced by indirect immunofluorescence assay (IFA) and RT-PCR using primer pairs within the TGEV leader and PRRSV GP5 gene (data not shown). Recently, expression of the reporter gene β -glucuronidase (GUS) and PRRSV ORF 5 from a TGEV-derived minigenome was demonstrated (Alonso et al. 2002). Importantly, strong humoral immune responses against GUS and PRRSV ORF5 were generated in swine with these vectors, demonstrating the feasibility of coronavirus-based vectors for future vaccine development.



amplified with a second primer (*Site B*) containing a *Esp3I* recognition at the 5' end of the nonsense strand of DNA by PCR. A similar approach is used to amplify the downstream arm (*Site C* and *D* primers). The insert DNA is amplified with primer pairs containing compatible C and D *Esp3I* sites. After PCR amplification and restriction digestion, the new insert can be inserted into the viral genome without evidence of the restriction sites used in the assembly cascade. A large number of class IIS restriction enzymes greatly enhances the plasticity of the approach

MHV-A59



icMHV-A59 GFP

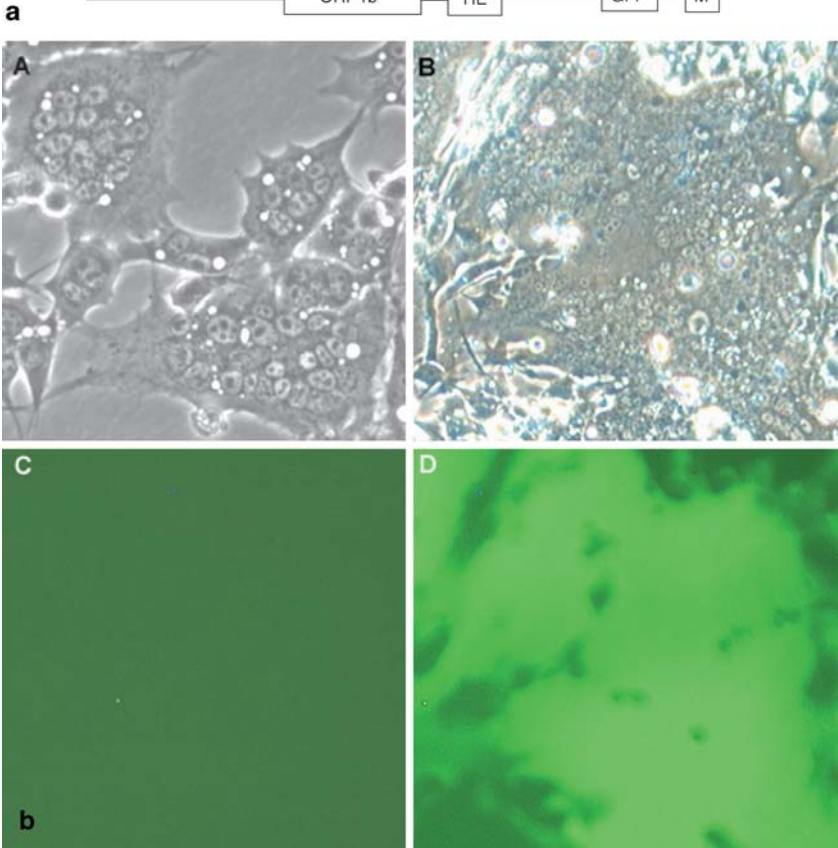
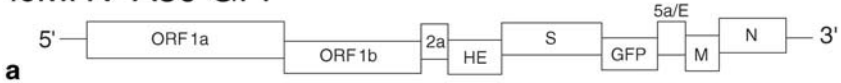


Fig. 5a, b. Recombinant MHV-A59 expressing GFP. With standard molecular techniques, ORF 4 was removed and the gene encoding GFP inserted downstream of the ORF 4 IS (a). DBT cells were infected with wild-type MHV-A59 (A and C) or icMHV-A59 GFP (B and D) and subsequently analyzed for CPE by light microscopy (A and B) and GFP expression by fluorescent microscopy (C and D) (b)

5 SARS-CoV Infectious Clone

Rapid response and control of exigent emerging pathogens require an approach to quickly generate full-length cDNAs from which molecularly cloned viruses are rescued, allowing for genetic manipulation of the genome. Identification of the first human coronavirus to cause considerable morbidity and mortality worldwide provided the first template to test the rapidity of our systematic assembly strategy (Drosten et al. 2003; Ksiazek et al. 2003). Development of novel vaccine candidates and therapeutics requires a better understanding of viral pathogenesis, a process greatly facilitated by the availability of an infectious clone. A systematic assembly strategy based on the TGEV infectious clone was employed to create an infectious construct of the SARS-CoV, within ~2 months of the identification and isolation of genomic SARS-CoV RNA (Yount et al. 2003). Consensus clones were assembled from sibling clones of each SARS-CoV fragment by taking advantage of the special properties of asymmetric type IIS restriction enzymes. Within 9 weeks, infectious clone SARS-CoV was isolated that was phenotypically indistinguishable from wild-type SARS-CoV strains.

The SARS-CoV genome was cloned as six contiguous subclones that could be systematically linked by unique *Bgl*I restriction endonuclease sites (Fig. 6). Two *Bgl*I junctions were derived from sites encoded within the SARS-CoV genome at nt 4,373 (A/B junction) and nt 12,065 (C/D junction). A third *Bgl*I site at nt 1,577 was removed, and new *Bgl*I sites were inserted by the introduction of silent mutations into the SARS-CoV sequence at nt 8,700 (B/C junction), nt 18,916 (D/E junction) and nt 24,040 (E/F junction). The resulting cDNAs include SARS A (nt 1–4,436), SARS B (nt 4,344–8,712), SARS C (nt 8,695–12,070), SARS D (nt 12,055–18,924), SARS E (nt 18,907–24,051), and SARS F (nt 24,030–29,736) subclones. The SARS A subclone also contains a T7 promoter, and the SARS F subclone terminates in 21Ts, allowing synthesis of capped, polyadenylated transcripts. SARS-CoV infectious clone virus was assembled, transcribed and transfected as described previously, and recombinant viruses contained the marker mutations inserted into the infectious clone. Recombinant viruses produced a mild pneumonia on x-ray in macaques similar to wild-type viruses and replicated to similar titers in the mouse model (unpublished observation). These data suggest that recombinant viruses recapitulated the pathogenesis of wild type in animal models, allowing for the identification of pathogenesis determinants and developing attenuated viruses as candidate live and killed vaccines.

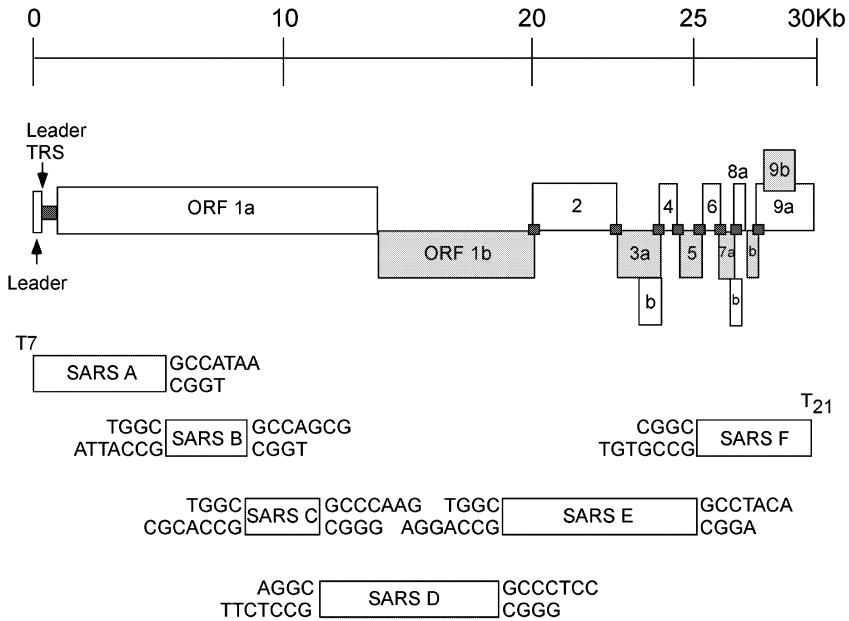


Fig. 6. Systematic assembly strategy for the SARS-CoV infectious clone. The SARS-CoV genome is about 30 kb in length and contains ~14 open reading frames (ORFs). The predicted functions of the group specific ORFs (*ORF 3a/b*, *ORF 6*, *ORF 7a/b*, *ORF 8a/b*, *ORF 9b*) are unknown. *Dark gray squares* indicate highly conserved consensus sequence sites that function in subgenomic RNA synthesis. Six independent subclones (*A*, *B*, *C*, *D*, *E*, and *F*) that span the entire SARS-CoV genome were isolated by RT-PCR (genome fragments are not shown to scale). The *A* fragment spans nt 1–4436, the *B* fragment nt 4344–8712, the *C* fragment nt 8695–12,070, the *D* fragment nt 12,055–18,924, the *E* fragment 18,907–24,051, and the *F* fragment nt 24,030–29,736. Unique *Bgl*I restriction sites located at the 5' and 3' ends of each subclone were used to assemble a full-length cDNA. A unique T7 start site was inserted at the 5' end of the SARS-CoV *A* fragment, and a 21 poly(T) tail was inserted at the 3' end of the SARS-CoV *G* fragment, allowing for in vitro transcription of full-length, capped, polyadenylated transcripts

6 Future Applications

The availability of infectious cDNA clones will undoubtedly have a profound effect on the field of coronavirology. These new tools will facilitate basic studies and allow for more precise analyses of the molecular mechanisms of viral replication, including the definition of RNA elements important for RNA replication, subgenomic RNA transcription, and ge-

nomic RNA packaging. In addition, studies of gene function will be enhanced by the availability of infectious cDNA clones by allowing for the construction of recombinant viruses and/or replicons containing mutations and the analysis of their effects on viral replication and assembly. MHV has long been used as a premiere model to study coronavirus assembly and release, replication, transcription, entry, and pathogenesis. The availability of MHV and SARS-CoV infectious cDNA clones will complement the existing targeted recombination approaches by providing a tool for the mutagenesis of the replicase gene, which encode a large number of cleavage products that have not been fully characterized. The structure and function of the ~20-kb MHV replicase domain will likely remain a fertile area of research for the next decade and reveal novel protein functions that participate and regulate discontinuous transcription and high-frequency RNA recombination. Although large panels of reagents are available for analyzing replicase protein expression, processing, and subcellular localization, a spectrum of genetically informative mutations have not been systematically targeted to any of these replicase proteins. Given the complexity and size of the coronavirus replicase gene, the number of potential mutants that can be generated is enormous and will likely require bioinformatic approaches for building and testing specific hypotheses. For example, the ORF1a C-terminal MHV p15 protein is highly conserved among group I through III coronaviruses and contains a large number of conserved cysteine residues and predicted phosphorylation, myristylation, and glycosylation sites (prosite, unpublished) (Fig. 7). The original sequence report of p15 also suggested possible similarities to growth factor-like proteins (Lee et al. 1991). Recent studies with an IBV homolog suggest that p15 exists as a dimer and accumulates on stimulation with epidermal growth factor, providing some evidence that the protein might be involved in the growth factor signaling pathway (Ng and Liu 2002). A single amino acid mutation has been identified in p15 of the temperature sensitive mutant, LA6, an MHV-A59 mutant with a defect in RNA synthesis at nonpermissive temperature (Siddell et al. 2001). The availability of infectious cDNAs allows, for the first time, a systematic mutagenesis approach for studying the function of specific structural features within this and other replicase proteins.

Coupled with the capacity to isolate large panels of mutants in each of the replicase proteins, selected questions include: (1) Are each of the PL1^{Pro}, PL2^{Pro}, and 3CL^{Pro} cleavage sites necessary for MHV replication? (2) Are the PL1^{Pro}, PL2^{Pro}, or 3CL^{Pro} proteases essential for replication? (3) Are any replicase proteins nonessential? (4) Is replicase gene order

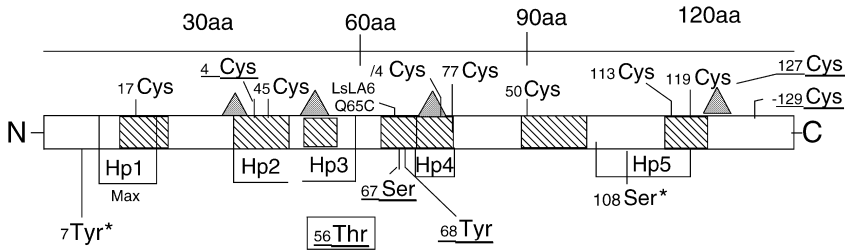


Fig. 7. Potential sites of mutagenesis within the C-terminal Orf1a p15 replicase protein. The MHV p15 replicase protein is highly conserved among all *Coronaviridae* (hatched domains), contains several hydrophobic domains (Hp1–5) and several potential sites for myristylation (gray triangles), and 10 highly conserved cysteine residues (Cys). Several sites for phosphorylation and glycosylation are predicted with prosite analysis, although it is unclear whether p15 is phosphorylated or glycosylated

critical? (5) Are replicase proteins interchangeable between the group 1 and/or group 2 coronaviruses? (6) How do replication complexes form on membranes? (7) What replicase complexes regulate discontinuous transcription and synthesis of genome-length and subgenomic-length mRNAs and negative-strand RNAs? (8) What are the *cis*-acting sequence elements required for genomic RNA packaging and replication? (9) What are the structure-function relationships within and between various replicase proteins and/or RNAs? (10) What are the functions of the group-specific ORFs, and how do they influence pathogenesis? The next decade of research may well be defined as the golden age of coronavirus genetics.

References

- Almazán F, González JM, Péntes Z, Izeta A, Calvo E, Plana-Durán J, Enjuanes L (2000) Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc Natl Acad Sci USA* 97:5516–5521
- Alonso S, Sola I, Teifke J, Reimann I, Izeta A, Balach M, Plana-Durán J, Moormann RJM, Enjuanes L (2002) In vitro and in vivo expression of foreign genes by transmissible gastroenteritis coronavirus-derived minigenomes. *J Gen Virol* 83:567–579
- Ballesteros ML, Sánchez CM, Enjuanes L (1997) Two amino acid changes at the N-terminus of transmissible gastroenteritis coronavirus spike protein result in the loss of enteric tropism. *Virology* 227:378–388

- Bonilla PJ, Gorbalenya AE, Weiss SR (1994) Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: heterogeneity among MHV strains. *Virology* 198:736–740
- Boyer JC, Haenni AL (1994) Infectious transcripts and cDNA clones of RNA viruses. *Virology* 198:415–426
- Casais R, Thiel V, Siddell SG, Cavanagh D, Britton P (2001) Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J Virol* 75:12359–12369
- Cavanagh D, Brian DA, Britton P, Enjuanes L, Horzinek MC, Lai MMC, Laude H, Plegemann PGW, Siddell S, Spaan W, Talbot PJ (1997) *Nidovirales*: a new order comprising *Coronaviridae* and *Arteriviridae*. *Arch Virol* 142:629–635
- Cho MK, Magnus D, Caplan AL, McGee D, Ethics of Genomics Group (1999) GENETICS: Ethical Considerations in Synthesizing a Minimal Genome. *Science* 286:2087–2090
- Curtis KM, Yount B, Baric RS (2002) Heterologous gene expression from transmissible gastroenteritis virus replicon particles. *J Virol* 76:1422–1434
- Delmas B, Gelfi J, L'Haridon R, Vogel LK, Norén O, Laude H (1992) Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV. *Nature* 357:417–420
- de Vries AAF, Horzinek MC, Rottier PJM, de Groot RJ (1997) The genome organization of the *Nidovirales*: similarities and differences between arteri-, toro-, and coronaviruses. *Semin Virol* 8:33–47
- Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RAM, Berger A, Burguiere A-M, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra J-C, Muller S, Rickerts W, Stürmer MV, S., Klenk H-D, Osterhaus ADME (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 348:1967–1976
- Eleouet JF, Rasschaert D, Lambert P, Levy L, Vende P, Laude H (1995) Complete sequence (20 kilobases not fully characterized). The structure and function of the ~20-kb MHV replicase domain will likely remain a fertile area of research for the next decade and reveal novel protein functions that participate and regulate discontinuous transcription and high-frequency RNA recombination. Although large panels of reagents are available for analyzing replicase protein expression, processing, and subcellular localization, a spectrum of genetically informative mutations have not been systematically targeted to any of these replicase proteins. Given the complexity and size of the coronavirus replicase gene, the number of potential mutants that can be generated is enormous and will likely require bioinformatic approaches for building and testing specific hypotheses. For example, the ORF1a C-terminal MHV p15 protein is highly conserved among group I through III coronaviruses and contains a large number of conserved cysteine residues and predicted phosphorylation, myristylation, and glycosylation sites (prosite, spect of coronavirus transcription). *J Virol* 71:5148–5160
- Fu K, Baric RS (1994) Map locations of mouse hepatitis virus temperature-sensitive mutants: confirmation of variable rates of recombination. *J Virol* 68:7458–7466
- Fu KS, Baric RS (1992) Evidence for variable rates of recombination in the MHV genome. *Virology* 189:88–102

- Grimes B, Cooke H (1998) Engineering mammalian chromosomes. *Hum Mol Genet* 7:1635–1640
- Hsue B, Masters PS (1999) Insertion of a new transcriptional unit into the genome of mouse hepatitis virus. *J Virol* 73:6128–6135
- Hutchison CA III, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999) Global transposon mutagenesis and a minimal mycoplasma genome. *Science* 286:2165–2169
- Izeta A, Smerdou C, Alonso S, Penzes Z, Méndez A, Plana-Durán J, Enjuanes L (1999) Replication and packaging of transmissible gastroenteritis coronavirus-derived synthetic minigenomes. *J Virol* 73:1535–1545
- Ksiazek TG, Erdman D, Goldsmith C, Zaki S, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell S, Ling A-E, Humphrey C, Shieh W-J, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang J-Y, Cox N, Hughes J, LeDuc JW, Bellini WJ, Anderson LJ (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348:1953–1966
- Kuo L, Godeke G-J, Raamsman MJB, Masters PS, Rottier PJM (2000) Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J Virol* 74:1393–1406
- Lai MMC, Cavanagh D (1997) The molecular biology of coronaviruses. *Adv Virus Res* 48:1–100
- Laude H, Rasschaert D, Delmas B, Godet M, Gelfi J, Bernard C (1990) Molecular biology of transmissible gastroenteritis virus. *Vet Microbiol* 23:147–154
- Lee HJ, Shieh CK, Gorbalenya AE, Koonin EV, Lamonica N, Tuler J, Bagdzhadzhyan A, Lai MMC (1991) The complete sequence (22 kilobases) of murine coronavirus gene-1 encoding the putative proteases and RNA polymerase. *Virology* 180:567–582
- Leparc-Goffart I, Hingley ST, Chua MM, Phillips J, Lavi E, Weiss SR (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
- Masters PS (1999) Reverse genetics of the largest RNA viruses. *Adv Virus Res* 53:245–264
- McGoldrick A, Lowings JP, Paton DJ (1999) Characterisation of a recent virulent transmissible gastroenteritis virus from Britain with a deleted ORF 3a. *Arch Virol* 144:763–770
- Narayanan K, Makino S (2001) Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J Virol* 75:9059–9067
- Ng LFP, Liu DX (2002) Membrane association and dimerization of a cysteine-rich, 16-kilodalton polypeptide released from the C-terminal region of the coronavirus infectious bronchitis virus 1a polyprotein. *J Virol* 76:6257–6267
- Penzes Z, González JM, Calvo E, Izeta A, Smerdou C, Mendez A, Sánchez CM, Sola I, Almazán F, Enjuanes L (2001) Complete genome sequence of transmissible gastroenteritis coronavirus PUR46-MAD clone and evolution of the Purdue virus cluster. *Virus Genes* 23:105–118
- Peters CJ, Sanchez A, Rollin PE, Ksiazek TG, Murphy FA (1996) Filoviridae: Marburg and Ebola Viruses. In: Fields BN, Knipe DM, Howley PM, Chanock RM, Melnick JL, Monath TP, Roizman B and Straus SE (eds) *Field's Virology*. Lippincott Williams and Wilkens, Philadelphia, pp 1161–1176

- Pingoud A, Jeltsch A (2001) Structure and function of type II restriction endonucleases. *Nucl Acids Res* 29:3705–3727
- Repass JF, Makino S (1998) Importance of the positive-strand RNA secondary structure of a murine coronavirus defective interfering RNA internal replication signal in positive-strand RNA synthesis. *J Virol* 72:7926–7933
- Rice CM, Grakoui A, Galler R, Chambers TJ (1989) Transcription of infectious yellow fever RNA from full-length cDNA templates produced by in vitro ligation. *New Biol* 1:285–296
- Risco C, Antón IM, Enjuanes L, Carrascosa JL (1996) The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. *J Virol* 70:4773–4777
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: A laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Sánchez CM, Gebauer F, Suñé C, Méndez A, Dopazo J, Enjuanes L (1992) Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology* 190:92–105
- Sánchez CM, Izeta A, Sánchez-Morgado JM, Alonso S, Sola I, Balasch M, Planadurán J, Enjuanes L (1999) Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. *J Virol* 73:7607–7618
- Schaad M, Baric RS (1994) Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. *J Virol* 68:8169–8179
- Siddell SG (1995) The Coronaviridae: an introduction. In: Siddell SG (ed) *The Coronaviridae*. Plenum Press, New York *The Viruses*, pp 1–10
- Siddell SG, Sawicki D, Meyer Y, Thiel V, Sawicki S (2001) Identification of the mutations responsible for the phenotype of three MHV RNA-negative ts mutants. *Adv Exp Med Biol* 494:453–458
- Smith GA, Enquist LW (2000) A self-recombining bacterial artificial chromosome and its application for analysis of herpesvirus pathogenesis. *Proc Natl Acad Sci USA* 97:4873–4878
- Stalcup RP, Baric RS, Leibowitz JL (1998) Genetic complementation among three panels of mouse hepatitis virus gene 1 mutants. *Virology* 241:112–121
- Thiel V, Herold J, Schelle B, Siddell SG (2001) Viral replicase gene products suffice for coronavirus discontinuous transcription. *J Virol* 75:6676–6681
- Tresnan DB, Levis R, Holmes KV (1996) Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I. *J Virol* 70:8669–8674
- van Zijl M, Quint W, Briaire J, de Rover T, Gielkens A, Berns A (1988) Regeneration of herpesviruses from molecularly cloned subgenomic fragments. *J Virol* 62:2191–2195
- Wesley RD, Woods RD, Cheung AK (1991) Genetic analysis of porcine respiratory coronavirus, an attenuated variant of transmissible gastroenteritis virus. *J Virol* 65:3369–3373
- Williams GD, Chang R-Y, Brian DA (1999) A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. *J Virol* 73:8349–8355

- Yount B, Curtis KM, Baric RS (2000) Strategy for systematic assembly of large RNA and DNA genomes: the transmissible gastroenteritis virus model. *J Virol* 74:10600–10611
- Yount B, Denison MR, Weiss SR, Baric RS (2002) Systematic assembly of a full length infectious cDNA of mouse hepatitis virus stain A59. *J Virol* 76:11065–11078
- Yount B, Curtis KM, Fritz EA, Hensley LE, Jahrling PB, Prentice E, Denison MR, Geisbert TW, Baric RS (2003) Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci USA* 100:12995–13000