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

VIROLOGY 198, 736–740 (1994)

Mouse Hepatitis Virus Strain A59 RNA Polymerase Gene ORF 1a: Heterogeneity among MHV Strains

PEDRO J. BONILLA,* ALEXANDER E. GORBALENYA,† AND SUSAN R. WEISS*,[1]

*Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6076; and †Institute of Poliomyelitis and Viral Encephalitides, Russian Academy of Medical Sciences, 142782 Moscow region, Russia

Received August 10, 1993; accepted November 3, 1993

Gene 1, the putative RNA replicase gene of coronaviruses, is expressed via two large overlapping open reading frames (ORF 1a and ORF 1b). We have determined the nucleotide sequence of ORF 1a, encoded within the first 13.7 kb of gene 1, for the coronavirus mouse hepatitis virus strain A59 (MHV-A59). Putative papain-like protease domains, a picornavirus 3C-like protease domain, two hydrophobic domains, and a domain "X" of unknown function, previously identified in other coronaviruses (1–3), are also present in ORF 1a of MHV-A59. Comparison between the ORF 1a sequence of MHV-A59 and the published sequence of the JHM strain of MHV (*2*) showed a high degree of similarity with the exception of several short regions. We sequenced one region of MHV-JHM that contained an 18 amino acid insertion relative to A59 and four other regions in which the sequences of the two strains differed. The MHV-2 and MHV-3 strains were also sequenced in some of these regions. Our analysis confirmed the presence of only one heterogeneous region in ORF 1a of MHV-A59 and MHV-JHM which is also present in MHV-2. Our findings indicate the need to modify the published sequence of MHV-JHM.

Coronaviruses are a group of RNA viruses of medical importance as enteric, respiratory, and neurologic pathogens (*4*). MHV-A59, a murine coronavirus, contains a single-stranded, positive-sense RNA genome, which is 31.3 kb in length (*2, 5*). Gene 1, the putative RNA-dependent RNA-polymerase gene, located at the 5′ end of the genome, is 21.7 kb while the other six genes encompass about 9.6 kb of the genomic RNA. The sequence of the RNA polymerase gene has been determined for three coronaviruses: the avian infectious bronchitis virus or IBV (*6*), MHV-JHM (*2*), and the human coronavirus 229E (*3*). These genes contain two overlapping ORFs: ORF 1a and ORF 1b. It has been shown that gene 1 contains sequence motifs common to viral polymerases and helicases (*1–3, 6, 7*). ORF 1a contains viral protease domains and ORF 1b contains polymerase, helicase, and zinc-finger motifs. ORF 1b is highly conserved between MHV, IBV, and 229E (*2, 3*), but there is significantly less homology in ORF 1a, particularly at the 5′ end.

Here we report the sequence of ORF 1a for MHV-A59 obtained by sequencing of overlapping cDNA clones (see Fig. 1). The sequence data reported in this paper will appear in the EMBL database under the accession number X73559. This completes the sequence for the MHV-A59 genome which is 31,327 nucleotides long plus a 3′-end poly (A) tail. The preparation of most of the cDNA clones used to determine the sequence of MHV-A59 has been described (*5*). Additional clones were prepared by reverse transcription of viral genome RNA followed by PCR amplification in the presence of sequence-specific primers. Descriptions of the 5′ leader of the viral MHV-A59 genome and the ORF 1a/ORF 1b junction have been reported elsewhere (*5, 7*).

Within the first 13.7 kb of MHV-A59, a long ORF is present between nucleotides 210–13616 and potentially encodes a 4468-amino-acid polypeptide. We compared the predicted amino acid sequence of MHV-A59 ORF 1a with the published sequence of MHV-JHM (*2*). An amino acid homology plot (see Fig. 2) showed the high degree of sequence similarity between the two strains. Conserved functional domains, encoded within ORF 1a of the RNA polymerase genes of the coronaviruses IBV, 229E, and MHV-JHM (*1–3*), are present in ORF 1a of MHV-A59. The location of these domains in MHV-A59 ORF 1a is shown in Fig. 1. Within ORF 1a of MHV there are two putative papain-like protease domains and one picornavirus 3C-like protease domain that is flanked by two hydrophobic domains. There is also a domain of unknown function conserved among alphaviruses, rubiviruses, hepatitis E virus, and coronaviruses designated as the "X" domain (*8, 9,* and references therein). The MHV-A59 first and second papain-like protease domains (amino acids 1084–1333 and 1681–1935) and the picornavirus 3C-like protease domain (amino acids 3334–3635) share 95, 95, and 98% identical residues with the corresponding sequences in MHV-JHM. Moreover, the amino acid se-

[1] To whom correspondence and reprint requests should be addressed.

0042-6822/94 $6.00

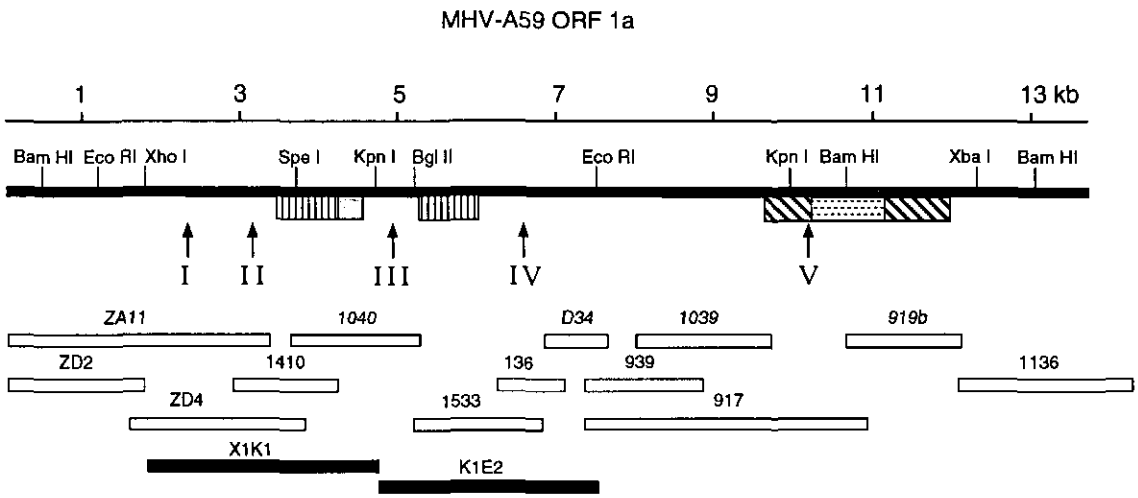

FIG. 1. Map of MHV-A59 ORF 1a. Significant restriction sites are indicated. Papain-like protease domains are represented as (▥), 3C-like protease domain as (▤), hydrophobic domains as (▨), and the domain "X" as (▩). Arrows below the map indicate regions of apparent sequence heterogeneity between MHV-A59 and the published sequence of MHV-JHM (*2*) which were further investigated by additional sequencing. cDNAs are represented as open rectangles and RT–PCR clones by closed rectangles.

quences surrounding the predicted catalytic residues for both the papain-like (Cys^{1121} and His^{1272} for domain one; Cys^{1716} and His^{1873} for domain two) and picornavirus 3C-like domains (His^{3374} and Cys^{3475}) are conserved. The presence of protease motifs in ORF 1a suggests that proteolytic processing may be important in the regulation of synthesis of replicase proteins. Protease activity has only been demonstrated for the 5′ most papain-like protease domain which cleaves p28, a protein encoded in the 5′ end of ORF 1a (*10, 11*, our unpublished results).

Against a background of extremely high amino acid sequence conservation, some regions of sequence heterogeneity were observed. The longest five of these regions were chosen for further analysis and included (1) an in-frame 18-amino-acid insertion in MHV-JHM (MHV-JHM amino acids 671–688), (2) a 27-amino-acid frameshift (MHV-A59 amino acids 915–941), (3) a 37-amino-acid frameshift (MHV-A59 amino acids 1530–1566), (4) a region containing both a 17-amino-acid frameshift followed by stretch of different amino acids resulting from a cluster of nucleotide substitutions (between MHV-A59 amino acids 2069–2122), and (5) a 14-amino-acid frameshift (MHV-A59 amino acids 3314–3327). We will refer to these areas as regions I, II, III, IV (A and B), and V, respectively (see Fig. 1). The first four regions of interstrain heterogeneity between MHV-A59 and JHM coincide with the previously identified most variable regions in ORF 1a of IBV, the human coronavirus 229E and MHV-JHM (*2, 3*, unpublished re-

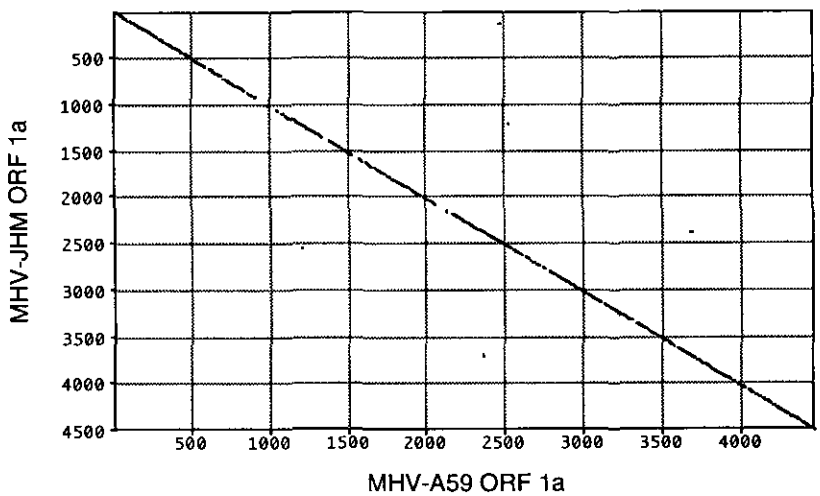

FIG. 2. Homology plot of the predicted amino acid sequence of ORF 1a between MHV strains A59 and JHM. The plot was generated from a Mac Vector 3.5 program with a window size of 10, homology = 90%, and a hash value of 2. The homology percentage scores were obtained with a protein scoring matrix (pam250 matrix) supplied with the program. Scales on both axes indicate amino acid numbers.

```
                REGION  I                          REGION  II
A59      2271 GAGGTGCCT[00]GAGCTTGTC    2943 GCTTTACTTGATAGGTTGGCAGGAGATTATGTCTATCTTTTTGATGAGGGAGGCGATGAAGTGATC
JHMPCR        .........[00].........          ...C.......A.......TG.AC.....................A........A.....T..T
JHMPUB   2276 .........[54].........    3002 ...C.....-.A.......TG.AC.....................A........A.....T..T
MHV-2         ..T..T...[00].....G...
MHV-3         .........[00].........          ..C..............A...A...C........C..C.........................T
F82           .........[00].........

                                                    REGION  III
A59      GCCCCGAGG-ATGTATTGTTCCTTTTCTGCTCCT    4788 GAAGCGCTGCGACATGATATACAATTGGATGATGATGCTCGTGTCTTTGTGCAGGCTAAT
JHMPCR   ...T.CC.T-.............C.........          ............................................................
JHMPUB   ...CTCCC.T......-GT-CCT-C.........    4841 .........-............................................................
MHV-2
MHV-3    .....TC.C-........C.....C..C......

                                                                                          REGION  IV-A
A59      ATGGACTGTCTGCCCACAGACTGGCGTCTCGTTAACAAATTTGATAGTGTTGATGGTGTTA-GAACCATT    6411 GACAAGGGGCCTGTGCCTGCTGCA
JHMPCR   ..........................T..C..T.........................-........          ..T.G.A..........GT.....
JHMPUB   ..........................T..C..T.........................A........    6464 ..T.G...-.........GT.....
MHV-2                                                                                     ..T.............G......

                                                                REGION  IV-B
A59      GTCCTTGTTACCGGCGTCCCT---GGAGCTGATGCGTCAGCTGGT---------GCCGGTATTGCCAAGGAGCAAAAAGCCTGTGCTTCTGCTAGTGTGGAG
JHMPCR   ..............T.CTG.GAGT..T..............ATATCCACCGAGC......C..............G..G..........A...C......
JHMPUB   ..............T.CTG.GAGT..T.....-.........ATATCCATTGAGC......C..............G..G..........A...C......
MHV-2    ........N.....T.CTCTGAGT..T....CCA.TG.G----------------C......C..............G.TG..C.......A.........T.

                                                                                                  REGION  V
A59      GATCAGGTTGTTACGGAGGTTCGTCAAGAGCCATCTGTTTCAGCTGCTGATGTCAAAGAG      10140 AACCATAATAATGGTAATGATGTTCTCTAT
JHMPCR   ......A......T.....CC.AGA.GAR.T.......CA..A.C.T..C..........          ............................
JHMPUB   ......A......T.....CC.AGA.GA..T.......CA..A.C.T..C..........      10202 ........C.-...................
MHV-2    ...........GT...G.T.CTTATC...C.TG.G..G.G..A.C.TG............

A59      CAGCCTCCAACCGCCTCTGTTACTACATCA
JHMPCR   ...........T...................
JHMPUB   ....-..-.-.T..-.....-.........
```

FIG. 3. Alignment of nucleotide sequences among MHV strains A59, JHM, MHV-2, and MHV-3. The consensus PCR sequence for both variants of MHV-JHM is indicated by the PCR subscript. The deduced nucleotide sequence of both MHV-JHM variants was identical with one exception in region IV-B (see text; R, nucleotides A or G). The PUB subscript refers to the published sequence of MHV-JHM (*2*). Also shown are the sequences that we determined for MHV-2 in regions I and IV and for MHV-3 in regions I and II. F82 is a cDNA derived from MHV-JHM-X (*2*) kindly provided by Dr. M. M. C. Lai. In the alignment, dots (.) indicate sequence identity to the MHV-A59 sequence. Dashed lines (-) indicate a gap introduced for optimal alignment. Gaps in the A59 sequence were introduced to indicate insertions in other MHV strains. The numbers within the brackets in region I indicate the number of nucleotides present between the adjacent sequences. The numbers to the left of each region refer to the position in the corresponding genome of the first nucleotide in each region.

sults). We reasoned that interstrain differences could result from a weak evolutionary pressure in particular regions of the coronavirus genome. Most of the differences between MHV-A59 and JHM in ORF 1a were replacements of amino acid blocks resulting from frameshifts. The frameshifts were caused by pairs of point mutations, an insertion and a deletion, flanking these regions (see Fig. 3). Therefore the apparent interstrain heterogeneity was not a result of a high rate of nucleotide substitutions and/or in-frame insertions/deletions, a well-known phenomenon in other systems (*12*). Instead the frameshifts suggested an unusual pattern of evolution in coronaviruses. We are aware of only one report of the presence of frameshift mutations in viable RNA viral variants (*13*). These observations encouraged us to further analyze these MHV interstrain differences in ORF 1a. Two variants of MHV-JHM were sequenced in these five regions. We sequenced MHV JHM-X, the same virus from which the published MHV-JHM sequence was derived (kindly provided by Dr. M. M. C. Lai). Another variant was obtained from Dr. J. L. Leibowitz. Cytoplasmic RNA from infected cells was isolated (*14*) and reverse transcribed (*15*). Five sets of primers (see Table 1) flanking the regions of interest were used to amplify DNA fragments by PCR. To decrease the probability of sequencing mistakes due to nucleotide misincorporation during amplification, we used Vent DNA polymerase, a thermostable DNA polymerase with a higher fidelity than *Taq* DNA polymerase due to the presence of 3′ to 5′ proofreading exonuclease activity (*16*). A consensus sequence was then obtained by direct PCR sequencing of both strands of each gel-purified DNA fragment (*17*). In addition, we were able to sequence some of these regions in the genomes of MHV-2 and MHV-3 using the same primer sets.

Figure 3 summarizes the sequencing results. The nucleotide sequences of both JHM variants were identical except for a G to A transition in region IV. The JHM variant obtained from Dr. Leibowitz has an arginine codon (AGG) while JHM-X has a lysine codon (AAG). This conservative mutation corresponds to Lys^{2134} in the

TABLE 1

SETS OF PRIMERS USED TO AMPLIFY AND SEQUENCE REGIONS I–V

Primer set	Position	Sequence
I	JFP 2208–2231	GATGGCTTTGCTACCAGCTGGTAA
	JRP 2647–2624	CTCGAAAGAAGAGGGTGGCTTACA
II	JFP 2932–2955	TGTGCTTGACGCAGTTGAGAGTAC
	JRP 3334–3311	ACCGACTTCGGTTTCCTCAGCAGA
III	JFP 4768–4791	TGTGACGAATGCATGTAGTTCGCT
	JRP 5074–5051	CCTTATTTGGCTAACACTCGCCTC
IV	JFP 6361–6384	CCATGAGGAAGCATCGCTGAAATC
	JRP 6781–6758	CCTACACCCTGTCAGGAACATATC
V	JFP 10141–10164	GGATACTGCTACTTATAGAGAGGC
	JRP 10418–10395	GGTCTGTCATGTCAGCTGAAGAAC

Note. Primers designated as JFP correspond to the viral genome sequence. JRP primers are complementary to the sequence in the viral genome. The numbers in the primer names refer to the first and last nucleotide according to the published sequence of MHV-JHM. Intracellular cytoplasmic RNA from MHV-infected cells was reverse transcribed and portions were amplified by PCR. PCR reactions were performed in 20 m*M* Tris (pH 8.8), 10 m*M* KCl, 10 m*M* $(NH_4)SO_4$, 2 m*M* $MgSO_4$, 0.1% Triton X-100, 400 μM each dNTP, 0.5 μM forward and reverse primer, 0.1 mg/ml BSA, and 1–2 units Vent DNA polymerase. After a hot start, reactions were cycled 25–30 times as follows: 95° for 30 sec, 68° for 30 sec, and 75° for 30 sec. PCR products were gel purified and sequenced from both directions with a fmol DNA sequencing system from Promega (*17*).

MHV-JHM published sequence. We were unable to confirm the presence of the 18-amino-acid insertion in region I in any of the two JHM variants nor from cDNA F82, one of the clones from which the original MHV-JHM sequence was obtained (*2*). Moreover, MHV-2 and MHV-3, like MHV-A59, lack this insertion. For regions II, III, and V we found that the nucleotide insertions and deletions in the published sequence of MHV-JHM, causing the apparent amino acid heterogeneity regions, were not present in the MHV-JHM PCR-derived sequence. As a result our deduced MHV-JHM amino acid sequence was similar to that of MHV-A59. MHV-3 also has a similar amino acid sequence to MHV-A59 in region II. In the case of region IV, we were unable to confirm the reported sequence of MHV-JHM which resulted in a 15-amino-acid frameshift (see Fig. 4). However, we did find that the deduced MHV-A59 sequence is 4 amino acids shorter than the PCR-derived MHV-JHM consensus sequence in region IV-A. Similarly, the MHV-2 sequence also has a 5-amino-acid in-frame deletion compared to the PCR-derived MHV-JHM sequence in this region. The in-frame deletions of MHV strains A59 and 2 are themselves different. Further downstream, in region IV-B, there is a stretch of amino acids in which the sequences of MHV-A59, MHV-JHM, and MHV-2 are heterogeneous. Therefore region IV is a domain of true amino acid heterogeneity in ORF 1a among these strains.

Similar findings of amino acid sequence heterogeneity have been observed elsewhere in the genome of MHV. Sequence comparison of the highly conserved N gene of several MHV revealed two small regions of sequence divergence (*18*). These two have been proposed to serve as spacer regions that connect three highly conserved structural domains of the N gene. Extensive amino acid heterogeneity has also been found in a localized region of the spike glycoprotein among strains of MHV and even among variants of the same strain (*19, 20*). Epitopes associated with neurovirulence have been mapped to the same heterogeneity region in the spike glycoprotein, suggesting a link between heterogeneity and function (*20*). It is possible that the heterogeneity region identified here serves as a spacer between functional domains of the replicase. Alternatively this may represent a functional motif that determines a particular characteristic among different MHV strains.

In conclusion, here we report the sequence for ORF 1a of MHV-A59 that together with the published sequence of MHV-A59 ORF 1b (*7*) completes the sequence of the replicase gene for MHV-A59. Comparison of the deduced amino acid sequences between MHV-A59 and JHM revealed the presence of several regions of apparent heterogeneity. However, sequencing analysis failed to confirm the apparent heterogeneity in all but one of these areas, indicating the need to correct the published sequence of MHV-JHM (*2*). One region of true sequence heterogeneity is present in the middle section of ORF 1a in three strains of MHV, but its functional significance remains to be determined. Recently we have identified ORF 1a-encoded polypep-

```
                REGION  IV-A                REGION  IV-B

A59             DKGPVPAAVLVTGVP GADASAGA   GIAKEQKACASASVEDQVVTEVRQEPSVSAADVKE
JHM_JL          .RR...S......AAS......ISTEP.T.........D.....I.M.AQKRS..TTVA...
JHM_X           .RR...S......AAS......ISTEP.T.........D.....I.M.AQKKS..TTVA...
JHM_PUB         .R.LCRLQS.LPVLRVVL-...ISIEP.T.........D.....I.M.AQKKS..TTVA...
MHV-2           ..............ALS..ATAP--    .T.....V...D..V....SGFLSDLCGATV....
```

FIG. 4. Alignment of amino acid sequences for MHV strains A59, JHM, and 2 in region IV. The sequences correspond to the nucleotide sequences shown in Fig. 3. The MHV-A59 sequence corresponds to amino acids 2068–2125. The amino acid sequences corresponding to MHV-JHM (obtained from Dr. J. L. Leibowitz), MHV-JHM-X, and the published MHV-JHM (*2*) [amino acids 2084–2144] are indicated by JL, X, and PUB subscripts, respectively. To maximize the homology in the alignment gaps were introduced into the MHV-A59 and MHV-2 sequences where MHV-JHM had amino acid insertions. Horizontal dashed lines indicate in-frame amino acid deletions.

tides (*21, 22*). Characterization of these polypeptides may provide insights into the role of this heterogeneity region.

ACKNOWLEDGMENTS

The authors thank Xiurong Wang, Andy Pekosz, and Dr. Susan Kraft for their technical assistance and Drs. Michael M. C. Lai and Julian L. Leibowitz for providing MHV-JHM variants. This work was supported by NIH Grant AI-17418. P.J.B. was supported by NIH Training Grant NS-07180.

REFERENCES

1. GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P., and BLINOV, V. M., *Nucleic Acids Res.* **17,** 4847–4861 (1989).
2. LEE, H. J., SHIEH, C. K., GORBALENYA, A. E., KOONIN, E. V., LA-MONICA, N., TULER, J., BAGDZHADZHYAN, A., and LAI, M. M. C., *Virology* **180,** 567–582 (1991).
3. HEROLD, J., RAABE, T., SCHELLE-PRINZ, B., and SIDDELL, S. G., *Virology* **195,** 680–691 (1993).
4. SPAAN, W., CAVANAGH, D., and HORZINEK, M. C., *In* "Immunochemistry of Viruses. II. The Basis for Serodiagnosis and Vaccines" (M. H. van Regenmortel and A. R. Neurath, Eds.), pp. 359–379. Elseviers Science Publishers, New York, 1990.
5. PACHUK, C. J., BREDENBEEK, P. J., ZOLTICK, P. W., SPAAN, W. J. M., and WEISS, S. R., *Virology* **171,** 141–148 (1989).
6. BOURSNELL, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. H., TOMLEY, F. M., and BINNS, M. M., *J. Gen. Virol.* **68,** 57–77 (1987).
7. BREDENBEEK, P. J., PACHUK, C. J., NOTEN, A. F. H., CHARITE, J., LUYTJES, W., WEISS, S. R., and SPAAN, W. J. M., *Nucleic Acids Res.* **18,** 1825–1832 (1990).
8. KOONIN, E. V., GORBALENYA, A. E., PURDY, M. A., ROZANOV, M. N., REYES, G. R., and BRADLEY, D. W., *Proc. Natl. Acad. Sci. USA* **89,** 8259–8263 (1989).
9. GORBALENYA, A. E., KOONIN, E. V., and LAI, M. M. C., *FEBS Lett.* **288,** 201–205 (1991).
10. BAKER, S. C., SHIEH, C. K., CHANG, M. F., VANNIER, D. M., and LAI, M. M. C., *J. Virol.* **63,** 3693–3699 (1989).
11. BAKER, S. C., YOKOMORI, K., DONG, S., CARLISLE, R., GORBALENYA, A. E., KOONIN, E. V., and LAI, M. M. C., *J. Virol.* **67,** 6056–6063 (1993).
12. DOMINGO, E., *Curr. Opin. Genet. Dev.* **2,** 61–63 (1992).
13. GARCÍA-BARRENO, B., PORTELA, A., DELGADO, T., LÓPEZ, J. A., and MELERO, J. A., *EMBO J.* **9,** 4181–4187 (1990).
14. SAMBROOK, J., FRITSCH, E. F., and MANIATIS, T., *In* "Molecular Cloning: A Laboratory Manual" (J. Sambrook *et al.*, Eds.), pp. 7.12–7.15. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
15. KAWASAKI, E. S., (1990). *In* "PCR Protocols: A Guide to Methods and Applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, Eds.), pp. 21–27. Academic Press, San Diego.
16. KONG, H., KUCERA, R. B., and JACK, W. E., *J. Biol. Chem.* **268,** 1965–1975 (1993).
17. GOMBOLD, J. L., HINGLEY, S. T., and WEISS, S. R., *J. Virol.* **67,** 4504–4512 (1993).
18. PARKER, M. M., and MASTERS, P. S., *Virology* **179,** 463–468 (1990).
19. LUYTJES, W., STURMAN, L., BREDENBEEK, P. J., CHARITE, J., VAN DER ZEIJST, B. A. M., HORZINEK, M. C., and SPAAN, W. J. M., *Virology* **161,** 479–487 (1987).
20. PARKER, S. E., GALLAGHER, T. M., and BUCHMEIER, M. J., *Virology* **173,** 664–673 (1989).
21. DENISON, M. R., ZOLTICK, P. W., HUGHES, S. A., GIANGRECO, B., OLSON, A. L., PERLMAN, S., LEIBOWITZ, J. L., and WEISS, S. R., *Virology* **189,** 274–284 (1992).
22. HUGHES, S. A., DENSION, M. L., BONILLA, P. J., LEIBOWITZ, J. L., and WEISS, S. R., *In* "Proceeding of the Fifth International Coronavirus Symposium," in press.