EVOLUTION AND TROPISM OF TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS

Luis Enjuanes,[1] Carlos Sánchez,[1] Fátima Gebauer,[1] Ana Méndez, [1] Joaquín Dopazo,[2] and María L. Ballesteros[1]

[1]Centro Nacional de Biotecnología, CSIC. Campus Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain
[2]Departamento de Sanidad Animal, INIA, Embajadores, 68 28012 Madrid, Spain

ABSTRACT

Transmissible gastroenteritis coronavirus (TGEV) is an enteropathogenic coronavirus isolated for the first time in 1946. Nonenteropathogenic porcine respiratory coronaviruses (PRCVs) have been derived from TGEV. The genetic relationship among six European PRCVs and five coronaviruses of the TGEV antigenic cluster has been determined based on their RNA sequences. The S proteins of six European PRCVs have an identical deletion of 224 amino acids starting at position 21. The deleted area includes the antigenic sites C and B of TGEV S glycoprotein. Interestingly, two viruses (NEB72 and TOY56) with respiratory tropism have the S protein with a similar size to the enteric viruses. NEB72 and TOY56 viruses have 2 and 15 specific amino acid differences with the enteric viruses, respectively. Four of the residues changed are located within the deletion present in the PRCVs and may influence the enteric tropism of TGEV *in vivo*. A receptor binding site (RBS) used by the virus to infect ST and other cell types might be located between sites A and D of the S glycoprotein, since monoclonal antibodies (MAbs) specific for these sites inhibit the binding of the virus to ST cells. An evolutionary tree relating 13 enteric and respiratory isolates has been proposed. According to this tree, a main virus lineage evolved from a recent progenitor which was circulating around 1941. From this, secondary lineages originated PUR46, NEB72, TOY56, MIL65, BRI70, and the PRCVs, in this order. Least squares estimation of the origin of TGEV-related coronaviruses showed a significant constancy in the mutation fixation rate. This rate was $7\pm2 \times 10^{-4}$ nucleotide substitutions per site and per year and falls in the range reported for other RNA viruses. Point mutations and probably recombination events have occurred during TGEV evolution.

INTRODUCTION

TGEV replicates in both villus epithelia cells of the small intestine and in lung cells. A nonenteropathogenic coronavirus related to TGEV appeared in 1984 in Europe[1]. This virus replicates in the respiratory tract and undergoes only limited replication in unidentified submucosal cell types of the small intestine. Both enteric and respiratory viruses cross-react[2]. The analysis of the genetic relationship among these respiratory isolates and others

Coronaviruses, Edited by H. Laude and J.F. Vautherot
Plenum Press, New York, 1994

with respiratory tropism will allow to determine the molecular bases of their tropism and evolution. In this paper, we describe the genetic homology among eight respiratory and five enteric isolates of the TGEV antigenic cluster, which identified amino acids potentially involved in receptor binding sites, and conserved areas of the S gene. Based on these viral sequences, an evolutionary tree, and mechanism for TGEV evolution, have been proposed.

MATERIALS AND METHODS

Cells, Viruses, and proteins

All viruses were grown on swine testis (ST) cells and their characteristics have been previously described[2]. Viral proteins were obtained from purified virions, incubated in the presence of proteinase inhibitors and deglycosylated with N-glycosidase F as described[3,4].

RNA Sequencing

RNA extracted from purified virions[3] was sequenced by oligodeoxynucleotide primer extension and dideoxynucleotide chain termination procedure[5].

Evolutionary Tree

Sequence information has been analyzed following standard phylogenetic methods[6]. The phylogenetic tree was obtained by the neighbor-joining[7] and the least squares methods. The reliability of the tree, i. e., the confidence levels in the branching order, was determined by the bootstrap method[8,9]. The origin of the phylogenetic tree was estimated by a linear least squares procedure[10].

RESULTS AND DISCUSSION

Structural Proteins and Sequence analysis of Enteric and Respiratory Porcine Coronaviruses

Enteric and respiratory TGEVs have been studied. According to the apparent size of the S protein of these viruses they could be grouped into two clusters, one including the enteric (PUR46) and the respiratory viruses (NEB72 and TOY56) with an S protein of standard size, and a second group, including the PRCVs with an S protein of smaller size (Figure 1). The first group of viruses have antigenic sites B and C, while in the second these antigenic sites were not detected by radioimmunoassay[2]. These results indicate that isolates with an almost exclusive respiratory tropism (NEB72 and TOY56) do not have a reduction in the molecular weight of the spike protein as the one detected in the PRCVs isolates.

To determine the relationship among the enteric and respiratory isolates, the nucleotide sequences of the complete S gene or of the half 5'-end were determined in 8 viral isolates and compared with sequences published for other isolates. The deduced amino acid sequences of 13 isolates were aligned, and the first 720 aminoacids are presented (Figure 2) and the results were diagrammatically summarized (Figure 3A). The amino acid positions reported in this manuscript refer to the location of equivalent residues in the sequence of MIL65 virus. All the PRCVs showed a deletion of 224 amino acids starting in the same position (aa 21), suggesting that all them have a common ancestor. In contrast, two North American PRCVs recently isolated (ISU-1, also designated IND89, and AR310)[11,12] have deletions of different size (227 and 207aa) starting in residues 23 and 17, respectively. These results indicate that the American and the European PRCVs have being independently originated. Although the NEB72 and TOY56 isolates have respiratory tropism, they do not have the large deletion on the S protein.

These viruses have only 2 and 15 specific amino acid differences with the enteric virus[4]. Interestingly, the NEB72 (Table 1) and TOY56 isolates have each one aminoacid dif-

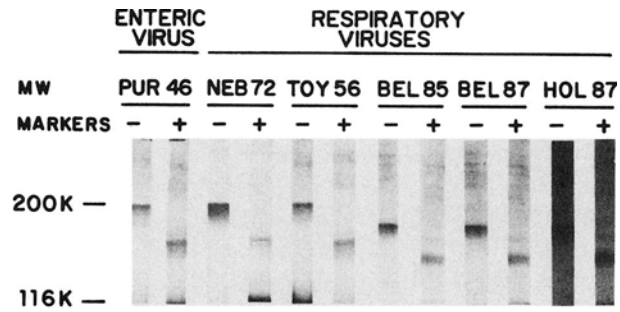


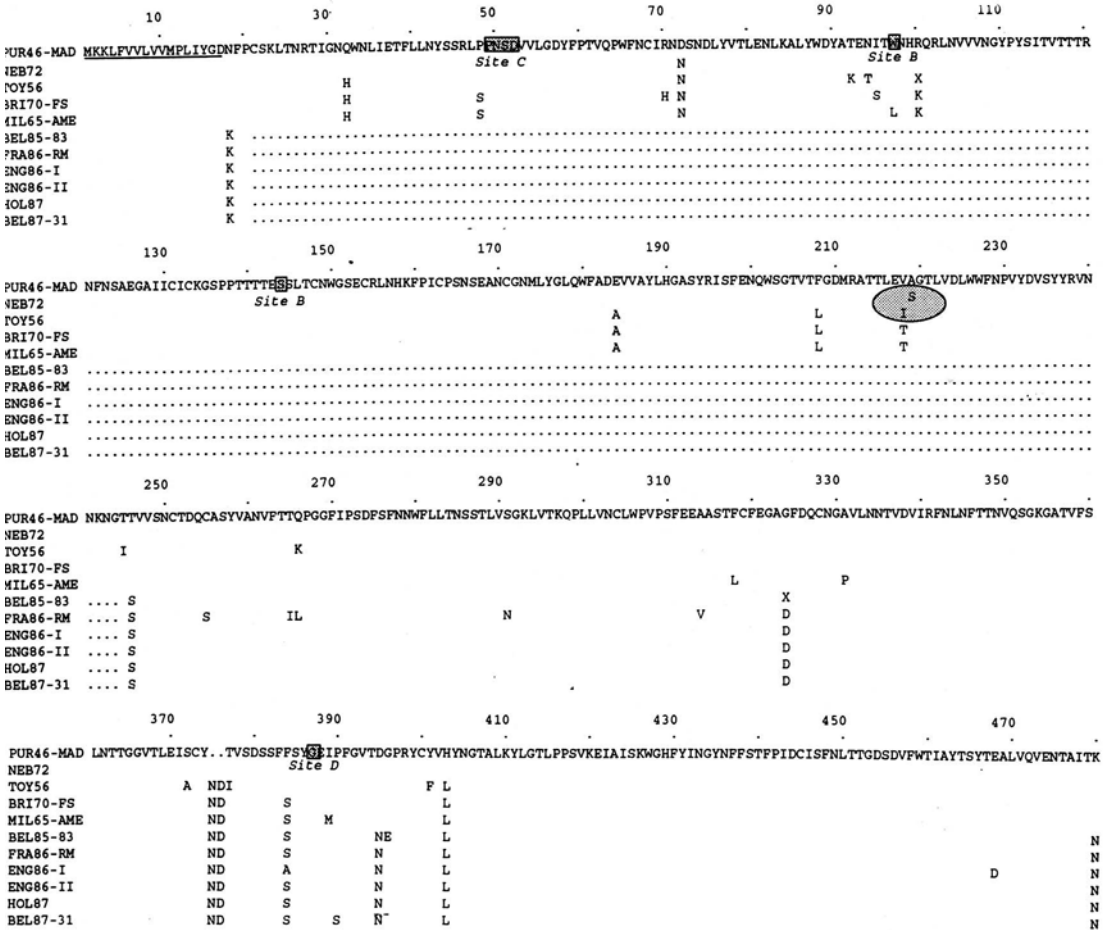
FIGURE 1. PAGE analysis of the spike protein of TGEV related coronaviruses before and after deglycosylation. Purified viruses were dissociated (1 μg/20μl) in 0.1 M sodium acetate, pH 7, with 0.5% SDS and protease inhibitors, and incubated overnight at 37° in the presence (+) or absence (-) of protein N-glycosidase F (0.04 U/μl). The proteins were separated by 7.5% PAGE in the presence of 0.1% SDS and 2-mercaptoethanol and detected using silver staining[13]. Only the gel area corresponding to the S glycoprotein is shown.

TABLE 1. Sequence differences between the S genes of enteric (PUR46) and respiratory (NEB72-RT) strain of TGEV

NUCLEOTIDE CHANGED	AMINO ACID CHANGED	CHANGED IN OTHER ENTERIC VIRUSES
214 G → A	70 Asp → Asn	+
655 G → T	219 Arg → Leu	-
2104 C → A	701 Gln → Lys	-
3263 A → R	-	-
3267 A → R	?	-
3270 N → X	?	-

ference in the S protein in positions 219 and 218, respectively. This changes are located within the deletion presented by the PRCVs, and could be responsible of the lack of enteric tropism shown by these viruses.

A RBS in the S glycoprotein of TGEV that interacts with ST cells probably maps between sites A and D, since TGEV binding to ST cells is best inhibited by MAbs specific for these sites[14]. Candidate domains for the localization of this RBS could be the highly conserved area identified between amino acids 405 and 465 (Figure 3B), although other domains around this area can not be ruled out. This RBS may mediate the infection of ST, and other cells growing in culture, by viruses with enteric and respiratory tropism, since it is present in all of them, and could interact with the aminopeptidase N described as the main TGEV receptor present on ST cells[15]. *In vivo*, a second RBS might be required to infect enteric cells. This RBS could be located around either aa 92, 94, and 218 or aa 219, changed in the TOY56 or in the NEB72 isolates, respectively. The two RBSs could mediate cell binding and fusion between the virus and cell membranes. A possibility, is that there is a unique RBS in the S protein of enteric and respiratory TGEV isolates, located between sites A and D. In this case, the deletion present in the PRCVs, or the residue changes within this area, described above, could modify this RBS resulting in viruses unable to infect epithelial intestinal cells *in vivo*. An alternative possibility is that the modifications around residues 218 and 219 could influence other viral or cellular regulatory mechanism affecting essential steps of virus replication, other than the virus-to-cell binding.


Site C
Site B
Site B
Site D

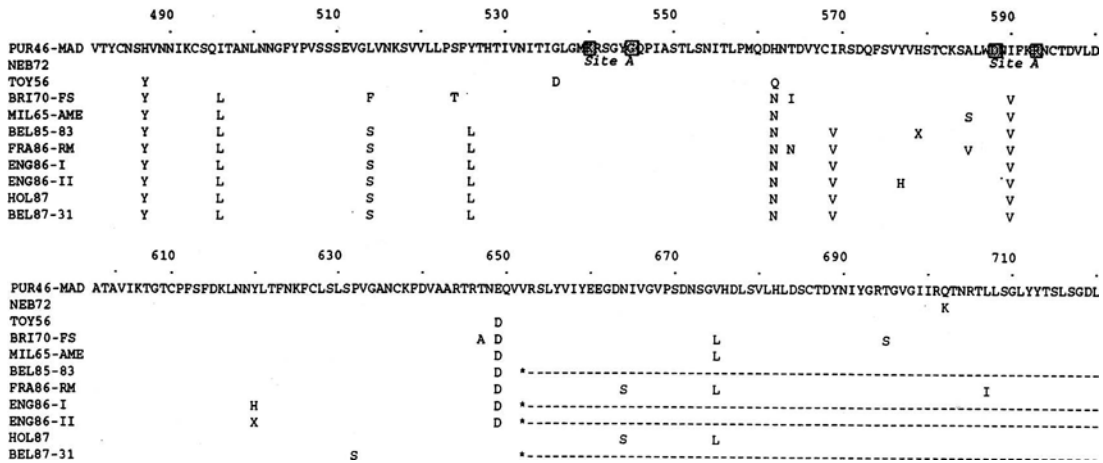


FIGURE 2. Sequence alignment of spike (S) protein of TGEVs and PRCVs. The sequence of the first 720 amino acids of the S protein of PUR46-MAD virus is shown in the first line. In the other lines, the aa changes in the sequence of the S proteins of other viruses have been indicated. In the alignment deleted residues have been filled out with points. Sequence numbers indicate the positions that the aa would have in the MIL65 virus. For simplicity, the sequences of two clones of the PUR46 isolate (PUR46-PAR and PUR46-UTR) have been omitted in this series of sequences, since they show minor changes and their sequences were previously published. The sequences of the strains PUR46-MAD, NEB72, TOY56, HOL87, FRA86-RM, MIL65-AME, BRI70-FS, PUR46-PAR, and PUR46-UTR have been previously reported[4,16,17,18,19,20]. Sequence indeterminations have been coded as X, for any amino acid. Underlined amino acids correspond to the signal peptide. Residues in boxes are involved in the indicated antigenic sites. Asterisks indicate the carboxi-terminus of the segments sequenced. Dashes indicate non-sequenced segments. The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accesion numbers: PUR46-MAD, M94101; NEB72-RT, M94099; TOY56, M94103; HOL87, M94097; BEL85-83, M94096; BEL87-31, M94098; ENG86-I, M94100; ENG86-II, M94102.

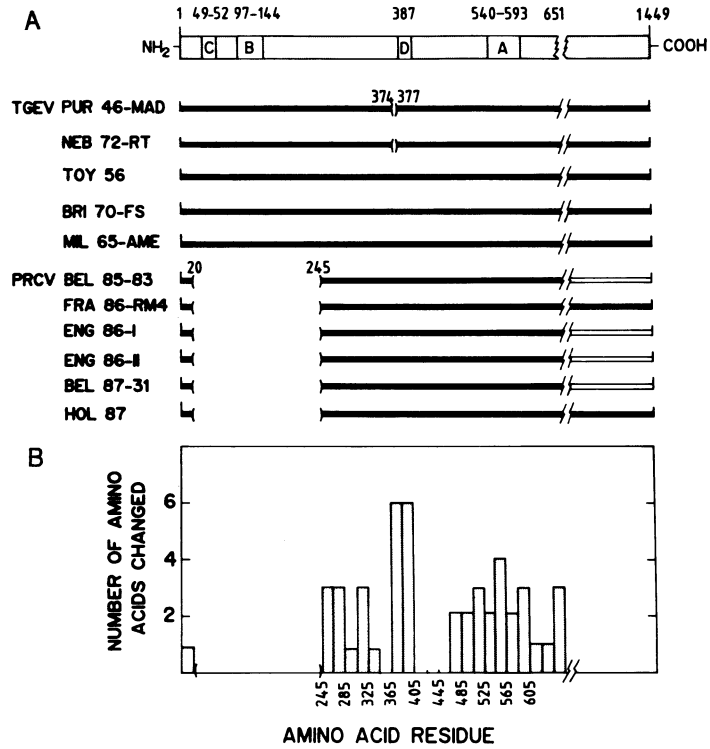


FIGURE 3. Summary of the deletions and amino acid changes present in the S glycoprotein of TGEVs and PRCVs. A. Location of the deletion. Full and empty bars indicate the sequences known and undetermined, respectively. Letters indicate the approximate location of the antigenic sites. The numbers above these letters indicate amino acid residues involved in the formation of these sites. The position of the deletions is indicated by brackets, and the numbers next to the brackets show the amino acids flanking the residues deleted. B. Number of amino acid changes in sequential fragments of 20 aa each, in relation with the PUR46-MAD virus sequence. Only the segments for which the sequences of the 13 virus strains were available have been included in the comparison. Amino acid residues have been numbered according to their position in the MIL65 virus after the alignment. The origin of the sequences of the different strains has been indicated in the legend of Fig. 2.

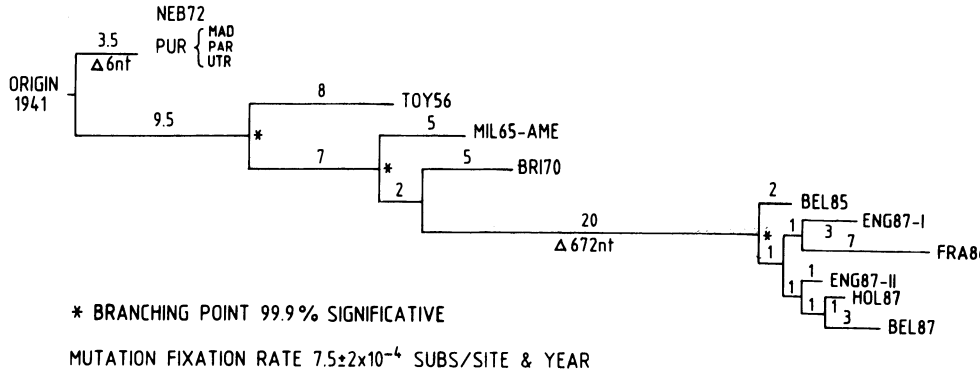


FIGURE 4. Evolutionary tree of TGEV related coronaviruses. Neighbour-joining and least squares methods of tree reconstruction procedures were applied to the first 1956 nucleotides of 13 virus isolates (the 11 isolates indicated in Fig. 2, and the clones PUR46-PAR and PUR46-UTR previously reported)[17,18]. Numbers in the diagram indicate residue substitutions between branching points. Δ, indicates the introduction of a deletion between branching points. *, indicates that all the descendents of this fork have, with a probability of 99.9%, a recent common ancestor.

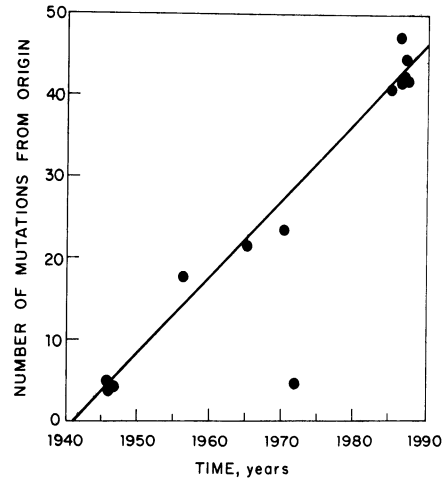


FIGURE 5. Relationship between mutation fixation rate and year of isolation. The line relating the number of mutations from origin with the year of isolation was plotted. Line and origin were estimated at the same time by linear least squares fit. The expression for the line was: d= 0.95 t - 1893, r^2= 0.97, where d is the distance to the origin, t is the time in years, and r^2 the Pearson's correlation coefficient[10]. The data correspond to the viral isolates used in the construction of the evolutionary tree (Fig. 4). The line with a minimum square error was determined and represented. The point showing minimum fitness with the line corresponds to the NEB72-RT isolate.

Evolutionary tree for the S gene of TGEVs and PRCVs

The nucleotide sequence of the S glycoprotein of eight respiratory and five enteric TGEVs (three of which were different clones of the same PUR46 virus strain) were aligned taking into account the two deletions of 6 and 672 nts present in the sequence of the PUR 46 and PRCVs, respectively, for maximum fitness. Phylogenetic analysis of the sequences (first 1956 nt) of the viruses described in Figure 2, by either the neighbor-joining or the least

squares methods of tree-reconstruction procedures, gave two identical trees, with the same branching order, confidence levels, and branch lengths (Figure 4). The least squares relationship between the number of mutations from origin and the year of isolation was determined (Figure 5). The extrapolation of this line to zero mutations allowed to predict that these TGEVs were originated from a recent common ancestor circulating around 1941. Since then, from a main lineage, the PUR46, TOY56, MIL65, BRI70, and the PRCVs were derived in the indicated order (Figure 4). The accumulation of mutations with time (Figure 5) fits a straight line with high Pearsons correlation coefficient (r^2=0.97). From the slope of this line, the mutation fixation rate can be estimated at 7±2 x 10^{-4} substitutions per nucleotide and per year. This rate falls in the range reported for other RNA viruses[21]. The direction defined for the evolutionary process, from the predicted origin, supports the occurrence of two deletions: one of 6 nt in the lineage from the root to PUR46 strains and another of 672 nt in the lineage leading from TGEV to PRCVs. It may be concluded that the European PRCVs have been derived by a 672-nt deletion from an enteric TGEV, since we have examined isolates preceding the PRCVs. In contrast, it cannot be guaranteed that the PUR46 emerged by a 6-nt deletion from an unknown ancestor. An alternative explanation could be that the other enteric isolates shown in Figure 3 could have been derived from PUR46 by the addition of 6 nt. If recombination has been the cause of the deletion present in the PUR46 and PRCVs, then two mechanisms of evolution would be involved in the antigenic variation of TGEV, point mutations and recombination.

Acknowledgements

This investigation has been founded by grants from the Comisión Interministerial de Ciencia y Tecnología, Comunidad Autónoma de Madrid, European Communities (BAP and Science Projects), NATO, and Fundación Ramón Areces.

REFERENCES

1. P. Callebaut, I. Correa, M. Pensaert, G. Jiménez, and L. Enjuanes, *J. Gen. Virol.* 69:1725 (1988).
2. C.M. Sánchez, G. Jiménez, M.D. Laviada, I. Correa, C. Suñé, M.J. Bullido, F. Gebauer, C. Smerdou, P. Callebaut, J. A. M. Escribano, and L. Enjuanes, *Virology* 174:410(1990).
3. F. Gebauer, W.A.P. Posthumus, I. Correa, C. Suñé, C.M. Sánchez, C. Smerdou, J. A. Lenstra, R. Meloen, and L. Enjuanes, *Virology* 183:225 (1991).
4. C.M. Sánchez, F. Gebauer, C. Suñé, A. Mendez, J. Dopazo, and L. Enjuanes, *Virology* 190:92 (1992).
5. O. Fichot, and M. Girard, *Nucleic Acids Res.* 18:6162 (1990).
6. N.M. Saitou, and M. Nei, *Mol. Biol. Evol.* 4:406 (1987).
7. W.M. Fitch, and E. Margoliash, *Science* 155:279 (1967).
8. B. Efron. "The Jackknife, the bootstrap and other resampling plans," Society for industrial and applied mathematics, Phyladelphia (1982).
9. J. Felsenstein. "PHYLIP manual version version 3.3", University of California (1990).
10. R.R. Sokal, and F. J. Rohlf. "Biometry," Freeman, New York (1981).
11. R.D. Wesley, R.D. Woods, and A.K. Cheung, *J. Virol.* 65:3369 (1991).
12. P.S. Paul, E.M. Vaughn, and P.G. Halbur, Int. Pig Vet.Soc. 15th Meeting (1992).
13. W. Ansorge, *J. Biochem. Bioph. Meth.* 11:13 (1985).
14. C. Suñé, G. Jiménez, I. Correa, M.J. Bullido, F. Gebauer, C. Smerdou, and L. Enjuanes, *Virology* 177:559 (1990).
15. B. Delmas, J. Gelfi, R.L'Haridon, L.K. Vogel, H. Sjöström, O. Norén, and H. Laude, *Nature* 357:417 (1992).
16. P. Britton, K. and W. Page, *Vir. Res.* 18:71 (1990).
17. L. Jacobs, R. de Groot, B.A.M. Van der Zeijst, M.C. Horzinek, and W. Spaan, *Vir. Res.* 8:363 (1987).
18. D. Rasschaert, and H. Laude, *J. Gen. Virol.* 68:1883 (1987).
19. D. Rasschaert, M. Duarte, and H. Laude, *J. Gen. Virol.* 71:2599 (1990).
20. R.D. Wesley, *Adv. Exp. Med. Biol.* 276:301 (1990).
21. E. Domingo, and J.J. Holland, in "RNA genetics," CRC Press, Florida (1988).