VIROLOGY 169, 217–221 (1989)

Phylogeny of Antigenic Variants of Avian Coronavirus IBV

J. G. KUSTERS,*,†,‡,[1] H. G. M. NIESTERS,†,[2] J. A. LENSTRA,* M. C. HORZINEK,† AND B. A. M. VAN DER ZEIJST*

*Departments of *Bacteriology and †Virology, Institute of Infectious Diseases and Immunology, University of Utrecht, P.O. Box 80.165, 3508 TD Utrecht, The Netherlands, and ‡Institute of Molecular Biology and Medical Biotechnology, University of Utrecht, P. O. Box 80.063, 3508 TB Utrecht, The Netherlands*

Received July 14, 1988; accepted November 7, 1988

The sequences of the peplomeric S1 protein of four serologically distinct strains of the infectious bronchitis virus (IBV), an avian coronavirus, have been determined. The S1 protein is thought to contain the serotype-specific neutralization epitopes and to be the main target of antigenic variation. An alignment with sequences of three strains published previously showed that from the 545 amino acid residues only 243 have been conserved. Clustering of substitutions suggests that most serotype determinants are located within the first 300 amino acid residues of S1. A phylogenetic tree of the S1 sequences showed very variable rates of divergence. Differences in topology with a tree based on RNAse-T1 fingerprint data indicate that some of the IBV strains have arisen by genetic recombination.

Avian infectious bronchitis (IB) is a worldwide disease, which is caused by a coronavirus and results in a highly contagious respiratory affliction of young chickens or in a decrease in egg production (*35*). Initially, IB was effectively controlled by vaccination with live attenuated IB virus (IBV). However, this did not prevent outbreaks caused by variant viruses (*10, 13*). RNase T1 fingerprinting analysis showed that field strains isolated from such outbreaks are related to vaccine strains (*24*), suggesting that the new strains have originated from vaccine virus by antigenic variation. As in other RNA viruses, antigenic variation is probably facilitated by the relatively high error rate of transcription (10^{-3}) during the transcription of the RNA template and the absence of a proofreading mechanism (*18, 33*).

It is generally assumed that the serotype of IBV is determined by the glycoprotein E2, which is the structural component of the peplomers, the typical club-shaped structures projecting from the surface of the virus. E2 is processed proteolytically to two noncovalently bound peptide chains, S1 and S2 (*6, 34*). S2 contains the C-terminal half of the sequence, including the transmembrane anchor and two long α-helices that form the stalk of the peplomer (*14*). S1 forms the top part of the peplomer and presumably carries the serotypical determinants. This assumption is based on the findings that all strongly neutralizing monoclonal antibodies recognize S1, and that immunization with purified S1, but not with virus lacking S1, induces neutralizing antibodies (*5, 7, 29*).

To investigate the serotypic variation of IBV at the molecular level, we determined the S1 sequences of four IBV strains (*13, 24*): H120, an attenuated vaccine strain of serotype A; D207, the reference strain of serotype B; D1466, a vaccine strain of serotype C; and V1397, a recent Dutch isolate from serotype A/C. These sequences were compared with the sequence of three strains published previously (*3, 4, 3*): M41, a pathogenic strain from serotype A used in a killed-virus vaccine; M42, a nonpathogenic laboratory strain of serotype A; and 6/82, a recent British field isolate of serotype B (*10*).

Virus strains were obtained from the Poultry Health Institute, Doorn, The Netherlands. Details on the isolation and passage history have been described (*24*). Strains were passaged once in the allantoic cavity of 10-day-old chicken embryos. Virus stocks were stored at −70°. Virus growth, isolation of genomic RNA, cDNA synthesis, cloning, and sequencing were carried out essentially as described previously (*31*). With V1397, cDNA synthesis was primed using random calf thymus pentanucleotides (Pharmacia). By screening of colonies with probes containing S1 sequences of the M41 strain, S1 clones of D207 and H120 were detected. Partial sequencing of a D1466 clone with a large insert yielded S1 sequences that could be used as a probe to obtain D1466 as well as V1397 clones.

Most sequences were based on two or more independent cDNA clones. Only with strain D207, three nu-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. J04329.

[1] To whom reprint requests should be addressed at Department of Bacteriology, University of Utrecht, P.O. Box 80.165, 3508 TD Utrecht, The Netherlands.

[2] Present address: California Institute of Technology, Division of Biology 156-29, Pasadena, CA 91125.

0042-6822/89 $3.00

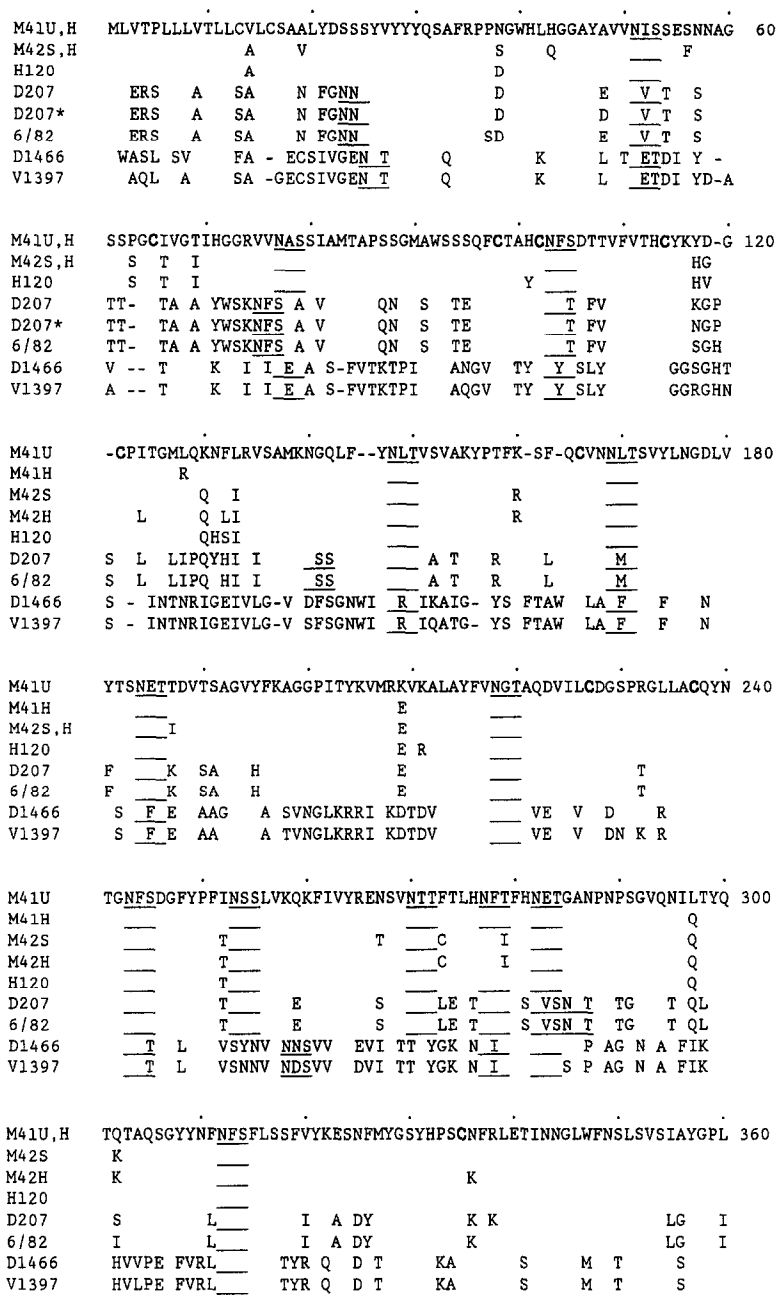

FIG. 1. Amino acid sequences of IBV S1 proteins. The sequence from M41 variant M41U has been listed completely; from other strains only the differences with this sequence are shown. The sequences from M41U, its cDNA variant M41U* (Ala on position 398, unpublished) and from the M42 variant M42S are from Niesters *et al.* (*31*). The sequence of M42 variant M42H is from Binns *et al.* (*3*). The sequences of M41 variant M41H and of strain 6/82 are from Binns *et al.* (*4*). D207* represents a cDNA variant of strain D207 (Glu-48 as main variant was found by sequencing independent clones, Lys-117 by direct RNA sequencing). Dashes were introduced to align the sequences. Potential glycosylation sites (NXS or NXT, X ≠ P) are underlined. Conserved cysteine residues are in boldface.

cleotide differences between cDNA clones were observed, two of them leading to amino acid substitutions. During the sequence determination of strain M41, one difference between two cDNA clones was observed. From the nucleotide sequences (submitted to the EMBL, GenBank, and DDBJ nucleotide sequence databases), amino acid sequences have been deduced. In Fig. 1, the amino acid sequences are listed together with the S1 sequences published previously. As observed earlier (*31*), sequence differences on a number of positions show that different laboratory variants of a strain can exist; our H120 sequence has

```
M41U,H   QGGCKQSVFSGRATCCYAYSY--GGPSLCKGVYSGELDLNFECGLLVYVTKSGGSRIQTA 420
M41U*                                        A
M42S                                         H
M42H               K                         H
H120                                 L       H
D207               AN              N      R  TKS      F  TD      R
6/82               AN              N      R  TKS      F  TD      R
D1466    K S       NRK        K PTN VQE    N  RNTK     FID TD    I
V1397    K S       NHK        K PTN VQE    N  RNTK     FID TD    I

M41U,H   TEPPVITRHNYNNITLNTCVDYNIYGRTGQGFITNVTDSAVSYNYLADAGLAILDTSGSI 480
M42S             Q    ___                  M  ___
M42H             QN   ___                     ___
H120             Q    ___                     ___
D207     N  FTL Q    ___ DR  E      V        ___NY IN      G M       A
6/82     N  FTL Q    ___ DR  E      V        ___NY IN      G M       A
D1466    EK   Y TNFT   VVGK  N      Y   V S_I_TE FGF-- QGD  V     A
V1397    EK   Y TNFT   VVGK  N      Y   V S___TE FGF-- EGD  V     A

M41U,H   DIFVVQGEYGLTYYKVNPCEDVNQQFVVSGGKLVGILTSRNETGSQLLENQFYIKITNGT 540
M42S,H               N                                              ___
H120           S     N                                  ___          ___
D207                 N                                  ___   P      I___
6/82                 N                                  ___   P      I___
D1466        S KDGPLTH   I   N     Y     NI  L   S___   IQ  D     L  S
V1397          RDGPFTH   I   N     Y     NI  L   S___   IQ  D     L  S

M41U,H   RRFRR 545
M42S,H
H120
D207       S
6/82       S
D1466      H
V1397      H
```

FIG. 1—*Continued.*

five differences with M41 not reported by Cavanagh *et al.* (*9*).

The length of the S1 protein varies between 535 and 538 amino acids, including the signal peptide and the arginine-rich cleavage site between S1 and S2 (*6*). The sequences could be aligned by assuming deletions/insertions at 14 positions. Two S1 proteins can have different amino acids in up to 49% of the positions of the sequence (Table 1). However, from the 17 to 19 cysteine residues, 16 have been completely conserved, as are most of the glycosylation sites. Presumably, the sequence variability is the combined result of the accumulation of neutral substitutions and the positive selection of antigenic variants; the relative high frequency of nonsilent mutations in several parts of the S1 sequence (not shown) suggests a positive selection.

TABLE 1

DIFFERENCE MATRIX OF S1 SEQUENCES OF IBV STRAINS[a]

	M41U	M42S	H120	D207	6/82	D1466	V1397
M41U (A)	—	4.7	3.9	22.6	22.4	44.4	44.4
M42S (A)	4.7	—	3.4	21.6	21.5	45.0	45.0
H120 (A)	3.9	3.4	—	21.5	21.5	44.0	44.0
D207 (B)	22.6	21.6	21.5	—	1.1	48.8	48.0
6/82 (B)	22.4	21.5	21.5	1.1	—	48.4	47.8
D1466 (C)	44.4	45.0	44.0	48.8	48.4	—	5.8
V1397 (A/C)	44.4	45.0	44.0	48.0	47.8	5.8	—

[a] The figures represent the percentages of nonidentical amino acids. The designations of the strains are as in Fig. 1. From each strain, only one variant has been listed.

To localize the most variable regions, the number of different amino acid residues found on all 545 positions is plotted in Fig. 2. On 243 positions, the amino acid residues are conserved in all strains. Although there are no clearly defined hypervariable regions—as for example, in the HIV envelope protein (*30*), the VP1 protein of foot-and mouth-disease virus (*2, 11*), and the rotavirus VP7 protein (*16, 17*)—there are relatively many replacements in the regions 50–170 and 250–310. Insertions/deletions were mainly found in the region 120–170. These observations suggest a localization of most of the serotypic and antigenic determinants in the N-terminal half of the S1 subunit.

A more accurate definition of the antigenic determinants may be derived from a comparison of similar sequences of strains with different serological properties. It has been suggested (*9, 31*) that in serotype-A strain the clustered substitutions in two regions, HVR 1 (56–

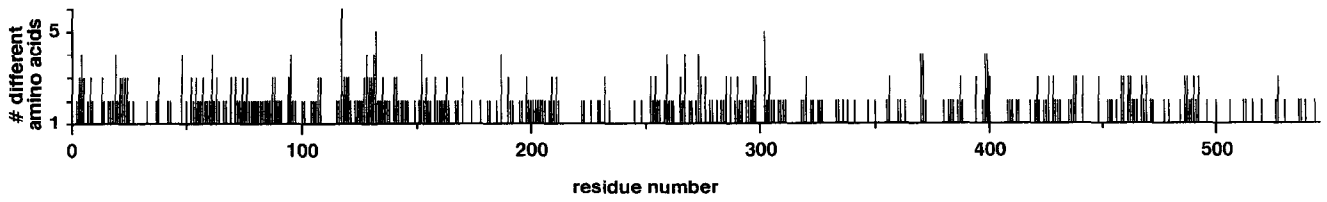

FIG. 2. Number of different amino acid residues per position in the sequences listed in Fig. 1.

69 in the numbering of Fig. 1) and HVR 2 (117–133) coincide with neutralization epitopes. Indeed, a mutation in HVR 1 prevented neutralization by two different monoclonal antibodies (*9*). From the six differences between the serotype-B strains D207 and H6/82, three are in a region corresponding to HVR 2.

The epitopes of IBV recognized by neutralizing monoclonal antibodies against S1 are conformation dependent (unpublished data). As described elsewhere (*25;* Kusters *et al.,* unpublished results), the 30 N-terminal residues of the S2 subunit contain several overlapping conformation-independent epitopes that evoke a weak neutralizing response. Since these epitopes cross-react with antisera against different serotypes, they are not relevant for the serotype of the virus.

An alignment of the S1 nucleotide sequences was used to calculate a distance matrix, from which the most likely phylogenetic tree was inferred by a program distributed as part of the PHYLIP package 2.6 (*15*). The topology and branch lengths of this tree, shown in Fig. 3A, were not affected by shuffling the order of sequences. Intriguingly, there are differences in topology between this tree and a tree based on RNase-T1 fingerprints (Fig. 3B). First, M41, M42, and H120 are

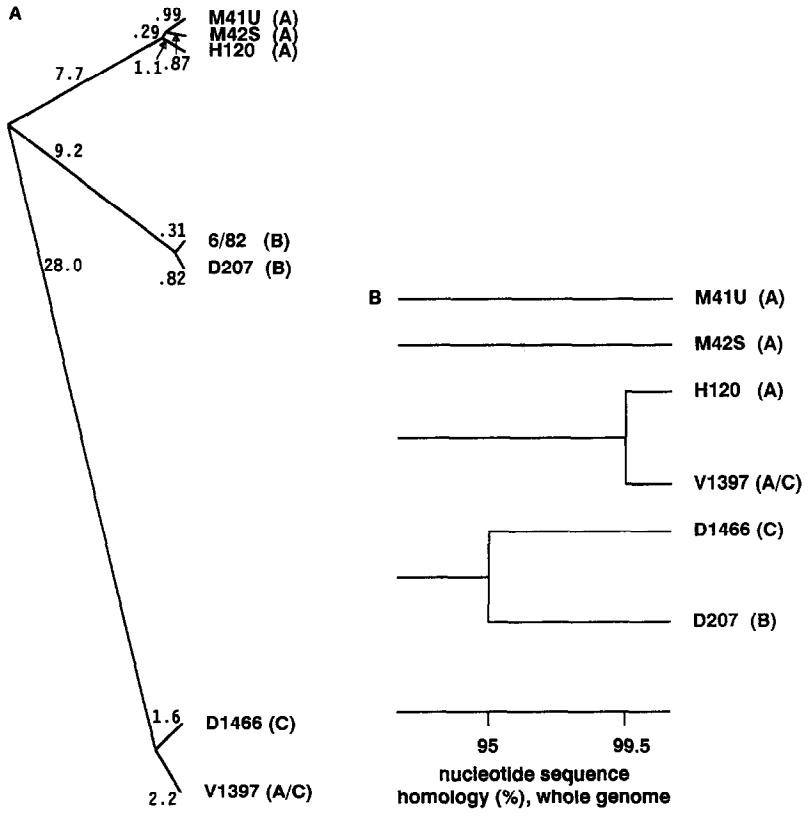

FIG. 3. Phylogenetic relationship between IBV strains. (A) Tree of S1 sequences. The percentage of nucleotide relatedness was calculated for the strains listed in Table 1. The figures indicate percentages of replaced nucleotides. The tree is unrooted, i.e., the position of the hypothetical ancestral IBV sequence is unknown. (B) Tree representing genomic relatedness. Percentages are estimated from RNase-T1 fingerprint analysis (*1*). The serotype is indicated between parentheses. The data are from Kusters *et al.* (*24*), except for M42 (*12;* J.G.K., unpublished).

placed apart in the T1 tree (<95% identity) but have closely related S1 sequences (≥97.5% nucleotide identity). Second, the considerable divergence of the S1 sequences of strains D1466 and D207 is not reflected in the T1 tree, nor in a tree (not shown) based on sequences of the E1 genes (*8*). Third, V1397 is related to H120 in the T1 tree (≥99.5% overall sequence similarity), but to D1466 in the S1 tree.

Theoretically, the first two of these discrepancies might be resolved by assuming extreme variations of the evolutionary rate within the viral genome. However, this would not explain the similar S1 sequences of V1397 and D1466 vs the common RNase-T1 spots of V1397 and H120. Instead, we propose that genetic recombination has played a role in the generation of antigenic variants. For instance, V1397 may have acquired a D1466-like peplomer gene. In the murine coronavirus MHV, recombination occurs at a rather high frequency (*19, 20, 26*) and may alter the serotype of the virus (*27*). For IBV, conditions that favor recombination are created in the field by vaccination of chickens with more than one attenuated IBV strain. Thus, infections of cells with two different strains, leading to the formation of recombinants, may very well have occurred.

RNA recombination has been well documented for picornaviruses (*21–23, 28, 32*). It would be interesting to test our hypothesis that recombination also plays a role in the generation of new IBV variants. Such a test might be based on the localization of the recombination site or on *in vitro* recombination experiments.

We thank Dr. G. Koch for communicating unpublished data. The help of Dr. J. A. M. Leunissen (University of Nijmegen) with the computer programs and the use of the facilities of the Dutch CAOS/CAMM Center, under Grants SON-11-20-700 and STW-NCH44.0703, are gratefully acknowledged. This work was supported in part by a research grant from Duphar B.V., Weesp, The Netherlands (H.G.M.N.).

REFERENCES

1. AARONSON, R. P., YOUNG, J. F., and PALESE, P., *Nucleic Acids Res.* **10,** 237–246 (1982).
2. BECK, E., FEIL, G., and STROHMAYER, K., *EMBO J.* **2,** 555–559 (1983).
3. BINNS, M. M., BOURSNELL, M. E. G., CAVANAGH, D., PAPPIN, D. J. C., and BROWN, T. D. K., *J. Gen. Virol.* **66,** 719–726 (1985).
4. BINNS, M. M., BOURSNELL, M. E. G., TOMLEY, F. M., and BROWN, T. D. K., *J. Gen. Virol.* **67,** 2825–2831 (1986).
5. CAVANAGH, D., DARBYSHIRE, J. H., DAVIS, P., and PETERS, R. W., *Avian Pathol.* **13,** 573–583 (1984).
6. CAVANAGH, D., DAVIS, P. J., PAPPIN, D. J. C., BINNS, M. M., BOURSNELL, M. E. G., and BROWN, T. D. K., *Virus Res.* **4,** 133–143 (1986a).
7. CAVANAGH, D., DAVIS, P. J., DARBYSHIRE, J. H., and PETERS, R. W., *J. Gen. Virol.* **67,** 1435–1442 (1986b).
8. CAVANAGH, D., and DAVIS, P. J., *J. Gen. Virol.* **69,** 621–629 (1988).
9. CAVANAGH, D., DAVIS, P. J., and MOCKETT, A. P. A., *Virus Res.* **11,** 141–150 (1988).
10. COOK, J. K. A., *Avian Pathol.* **13,** 733–741 (1984).
11. CHEUNG, A., DELMARTES, J., WEISS, S., and KÜPPER, H., *J. Virol.* **48,** 451–459 (1983).
12. CLEWLEY, J. P., MORSER, J., AVERY, R. J., and LOMNICZI, B., *Infect. Immunol.* **32,** 1227–1233 (1981).
13. DAVELAAR, F. G., KOUWENHOVEN, B., and BURGER, A. G., *Vet. Q.* **6,** 114–120 (1984).
14. DE GROOT, R. J., LUYTJES, W., HORZINEK, M. C., VAN DER ZEIJST, B. A. M., SPAAN, W. J. M., and LENSTRA, J. A., *J. Mol. Biol.* **196,** 963–966 (1987).
15. FELSENSTEIN, J., *Evolution* **39,** 783–791 (1985).
16. GREEN, K. Y., MIDTHUN, K., GORZIGLIA, M., HOSHINO, Y., KAPIKIAN, A. Z., CHANOCK, R. M., and FLORES, J., *Virology* **161,** 153–159 (1987).
17. GUNN, P. R., SATO, F., POWELL, K. F. H., BELLAMY, A. R., NAPIER, J. R., HARDING, D. R. K., HANCOCK, W. S., SIEGMAN, L. J., and BOTH, G. W., *J. Virol.* **54,** 791–797 (1985).
18. HOLLAND, J., SPINDLER, K., HORODYSKI, F., GRABAU, E., NICHOL, S., and VANDEPOL, S., *Science* **215,** 1577–1585 (1982).
19. KECK, J. G., STOHLMAN, S. A., SOE, L. H., MAKINO, S., and LAI, M. M. C., *Virology* **156,** 331–341 (1987).
20. KECK, J. G., MATSUSHIMA, G. K., MAKINO, S., FLEMING, J. O., VANNIER, D. M., STOHLMAN, S. A., and LAI, M. C., *J. Virol.* **62,** 1810–1813 (1988).
21. KEW, O. M., and NOTTAY, B. K., *In* "Modern Approaches to Vaccines" (R. Chanock and R. Lerner, Eds.), pp. 357–362. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1984.
22. KING, A. M. Q., MCCAHON, D., SLADE, W. R., and NEWHAM, J. W. I., *Cell* **29,** 921–928 (1982).
23. KIRKEGAARD, K., and BALTIMORE, D., *Cell* **47,** 433–443 (1986).
24. KUSTERS, J. G., NIESTERS, H. G. M., BLEUMINK-PLUYM, N. M. C., DAVELAAR, F. G., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M., *J. Gen. Virol.* **68,** 343–352 (1987).
25. LENSTRA, J. A., KUSTERS, J. G., KOCH, G., and VAN DER ZEIJST, B. A. M., *Mol. Immunol.* **26,** 7–15 (1989).
26. MAKINO, S., KECK, J. G., STOHLMAN, S. A., and LAI, M. M. C., *J. Virol.* **57,** 729–737 (1986).
27. MAKINO, S., FLEMING, J. O., KECK, J. G., STOHLMAN, S. A., and LAI, M. M. C., *Proc. Natl. Acad. Sci. USA* **84,** 6567–6571 (1987).
28. MCCAHON, D., SLADE, W. R., PRISTON, K. A. J., and LAKE, J. R., *J. Gen. Virol.* **35,** 555–565 (1977).
29. MOCKETT, A. P. A., CAVANAGH, D., and BROWN, T. D. K., *J. Gen. Virol.* **65,** 2281–2286 (1984).
30. MODROW, S., HAHN, B. H., SHAW, G. M., GALLO, R. C., WONG-STAAL, F., and WOLF, H., *J. Virol.* **61,** 570–578 (1986).
31. NIESTERS, H. G. M., LENSTRA, J. A., SPAAN, W. J. M., ZIJDERVELD, A. J., BLEUMINK-PLUYM, N. M. C., HONG, F., VAN SCHARRENBURG, G. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M., *Virus Res.* **5,** 253–263 (1986).
32. ROMANOVA, L. I., TOLSKAYA, E. A., KOLESNIKOVA, M. S., and AGOL, V. I., *FEBS Lett.* **118,** 109–112 (1980).
33. STEINHAUER, D. A., and HOLLAND, J. J., *J. Virol.* **57,** 219–228 (1986).
34. STERN, D. F., and SEFTON, B. M., *J. Virol.* **44,** 794–803 (1982).
35. WEGE, H., SIDDELL, S., TER MEULEN, V., *Curr. Top. Microbiol. Immunol.* **99,** 165–200 (1982).