
Sequence of the nucleocapsid gene from murine coronavirus MHV-A59

John Armstrong⁺, Sjeff Smeekens and Peter Rottier

⁺European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG, and Institute of Virology, Veterinary Faculty, State University of Utrecht, 3508 TD Utrecht, The Netherlands

Received 1 December 1982; Accepted 5 January 1983

ABSTRACT

The nucleotide sequence of the RNA encoding the nucleocapsid protein of coronavirus MHV-A59 has been determined. Copy DNA was prepared from mRNA isolated from virally infected cells, fragmented and cloned in the phage vector M13 mp8 for direct sequence determination. A sequence of 1817 nucleotides, adjacent to the viral poly-A tail, was obtained. It contains a single long open reading frame encoding a protein of mol. wt. 49660, which is enriched in basic residues.

INTRODUCTION

The coronaviruses comprise a large group of enveloped RNA viruses isolated from a range of animal hosts (for review see 1). They have been associated with a variety of respiratory and gastro-intestinal infections and neurological disorders, and may also provide a model for the study of persistent viral infection.

The coronavirus strain A59, a mouse hepatitis virus, can be propagated in cell culture, and its molecular biology has been studied in some detail. The virion contains a single positive-stranded RNA, 18kB in length, associated with a nucleocapsid protein (2). The viral lipid envelope contains two glycoproteins: E1, of mol. wt. 24,000, occurring in unglycosylated as well as well as O-glycosylated forms (3,4,5), and E2, (mol. wt. 90,000/180,000), which forms the surface projections or peplomers characteristic of the corona virion (2,4).

Two features of the life cycle of coronaviruses are of particular interest at the molecular level. First, the viral mRNA's produced during infection form a nested set, corresponding to the 3' end of the virion RNA but extending to different lengths towards the 5' end: the largest is identical to the virion RNA (6-12). From each of the RNA's (seven in the case of MHV-A59), only the 5' gene is translated (7,13,14). Thus, the coronaviruses have a replication strategy which differs from any so far reported for RNA viruses. Second-

dly, the virus buds intracellularly, in endoplasmic reticulum and perhaps Golgi membranes (15,16). The factors which specify this site of assembly, rather than the plasma membrane, are at present unknown.

Here we report a nucleotide sequence of cloned copy DNA prepared from MHV-A59 mRNA. A sequence of 1817 nucleotides, ending in the polyadenylation site, has been determined. Translation of the sequence predicts a polypeptide whose size, general features and genetic location are consistent with its being the viral nucleocapsid protein.

MATERIALS AND METHODS

cDNA preparation

Total poly-A⁺ RNA from MHV-A59 infected Sac⁻ cells was prepared as described (17). First-strand cDNA was synthesized in a mixture containing RNA (50 μ g), Tris-Cl pH 8.3 (50mM), KCl (50mM), MgCl₂ (8mM), dithiothreitol (1mM), oligo-dT₁₂₋₁₈ (1 μ g, P.L. Biochemicals), sodium pyrophosphate (2mM), dATP, dGTP, dTTP (each 1mM), dCTP (0.5mM), α -³²P-dCTP (50 μ Ci: Amersham) and AMV reverse transcriptase (350 units; kindly provided by Dr. J. Beard) in a total volume of 200 μ l. The mixture was incubated for 30 min. at 41°C, then for 15 min at 45°C. EDTA was added to 20mM, the material extracted twice with phenol/chloroform and the pooled aqueous phases extracted twice with ether. The products were precipitated with ethanol and redissolved in 50 μ l 5mM Tris, 1mM EDTA pH 7.5. The material was loaded on a 1% low-melting-temperature agarose (BRL) gel, and electrophoresis carried out at 20v/cm for 60 min, to remove low molecular weight material. Regions of the gel containing cDNA were identified by autoradiography, cut into 1mm slices, and each slice melted, phenol-extracted and the cDNA precipitated with ethanol. RNA was hydrolysed by incubating the material from each slice in 0.2M KOH for 10 min at 65°C in a volume of 20 μ l, and the mixture was neutralized with HCl. The cDNA was converted to the double-stranded form in a mixture containing HEPES/KOH, pH 6.9 (100mM), MgCl₂ (4mM), dithiothreitol (0.5mM), dATP, dCTP, dGTP and dTTP (each 1mM), KCl (50mM), Klenow polymerase (20 units; Boehringer) in a volume of 100 μ l. The mixture was incubated for 3 hrs at 17°C, and the DNA precipitated with ethanol. Yields obtained from each gel fraction were measured by TCA precipitation of aliquots on filters, and scintillation counting.

Fragmentation and cloning of cDNA

Portions (30ng) of the double-stranded cDNA were cleaved with one of the restriction enzymes HaeIII (kindly supplied by V. Pirrotta), Fnu DII or RsaI (both New England Biolabs) under standard conditions. The DNA was purified

by phenol extraction and ethanol precipitation, and randomly ligated to 10ng M13 mp8 (18) replicative-form DNA which had previously been linearized with SmaI (New England Biolabs), in a mixture containing Tris-HCl, pH 7.5 (50mM), MgCl₂ (10mM), dithiothreitol (1mM), ATP (0.2mM) and T4 ligase (6 units; kindly supplied by R. Brown) in a volume of 10 μ l. The ligations were incubated overnight at room temperature.

Alternatively, cDNA was treated with nuclease S1. cDNA (80ng) was incubated in 60 μ l of a solution containing sodium acetate (30mM, pH 5.2), NaCl (0.3M), Zn Cl₂ (2mM) and 10 units S1 nuclease (BRL), for 3 min at 37°C and 10 min at room temperature. EDTA was added to 5mM, the mixture extracted twice with phenol/chloroform, washed with chloroform and ether, and the DNA precipitated with ethanol; approximately 50% of the input radiolabel remained TCA-precipitable. The DNA was then treated with Klenow polymerase in a 10 μ l volume containing Tris-Cl (10mM, pH 7.5), MgCl₂ (10mM), NaCl (50mM), dithiothreitol (0.5mM), all four deoxynucleotides (each 0.4mM) and Klenow polymerase (1 unit) for 15 min. at room temperature. EDTA was added to 20mM and the mixture phenol-extracted and ethanol-precipitated as before. The DNA was then ligated to SmaI-digested M13 mp8 exactly as for restricted cDNA.

The products of the ligation reactions were then used to transfect competent *E. coli* cells, lac⁻ plaques picked and grown, the viral DNA isolated and used directly for single-nucleotide dideoxy-sequencing of the inserted DNA (all as in 19), without any prior screening. Clones of interest were sequenced, using a 15-base synthetic single-strand oligonucleotide (P.L. Biochemicals) as universal primer. Reaction products were analysed on 0.2mm thick thermostatted 6% acrylamide gels (20), 40cm or 60cm in length. Sequences were stored, and overlaps identified, by computer, using the program package of Staden (21).

RESULTS AND DISCUSSION

"Shotgun" cloning of viral cDNA

A total of 48 recombinant clones were sequenced; of these sequences, 33 could be assembled into a contiguous consensus of 1817 nucleotides, followed by a poly-A tract of variable length (Fig.1). The remaining 15 clones, none of which contained sequence overlaps with each other, presumably arise from cellular mRNA present in the starting material, or from regions towards the 5' end of the viral genome.

Several of the clones generated by S1-nuclease digestion of the cDNA diverged at one end from the main consensus (Fig.1). With the possible exception of clone S9 (see below), these were probably due to short 3' extensions remain-

ning on some cDNA fragments after treatment with S1 nuclease and Klenow polymerase; these could promote simultaneous ligation of more than one fragment to a single vector molecule. Therefore sequences obtained from S1-derived clones were used only to confirm regions of the consensus obtained from restricted-cDNA clones.

Structure of the MHV-A59 RNA's

A sequence of 1817 nucleotides, adjoining a poly-A tract, is shown in Fig.2. The viral origin and strand orientation of the sequence was confirmed by dot hybridization of radiolabelled, fragmented virion DNA to the single-strand recombinant DNA's (not shown).

Comparison of the sequence with the analysis of RNase-T1 oligonucleotides from various viral RNA's reported by Lai et al. (22) showed good, although not perfect, agreement (Table 1) between the two sets of data. Of the nine oligonucleotides isolated from RNA7, seven have similar corresponding sequences in Fig.1. The two which were not identified, spots 10 and 19, probably correspond to two spots which are also not found in 3' fragments of the virion RNA (6). It has been proposed that these oligonucleotides reflect the existence of a short 5' sequence common to all seven viral RNA's, and that oligonucleotides 19 and 19a cross the joining site between the leader sequence and the coding region in RNA's 7 and 6 respectively (8,6). Nucleotides 137-152 from Fig.1 show some similarity to both of these oligonucleotides, and also to spot 17, which is found in all the larger RNA's, but not RNA7 (see Table 1). If the

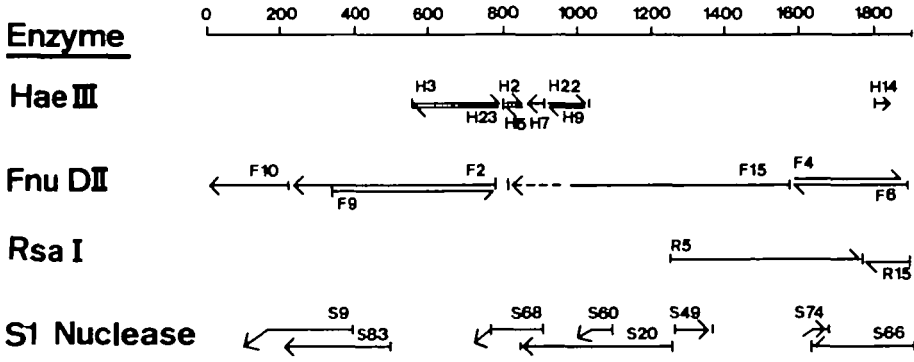


Figure 1. Arrangement of clones used to construct the nucleocapsid sequence. Arrows show the direction in which the sequence was determined (the complement of the M13 single-strand DNA). The 5' end of clone F9 is an anomalous cleavage; all other restricted-cDNA clones are bounded by correct restriction sites. The sequence of clone F15 was readable beyond 600 nucleotides with sufficient accuracy to confirm its overlap with clone H9. Diagonal lines represent regions of divergence from the consensus sequence.

TABLE 1.

Position in sequence	Predicted composition	Reported composition (22)	Oligonucleotide spot No. (22)
137-152	U ₅ C(A ₂ U)(A ₂ G)(A ₃ C)	U ₅ C ₂ (A ₂ C)(A ₂ U)(A ₂ G)(A ₃ C) U ₇ C ₂ (A ₂ U)(A ₂ G)(A ₃ C)(A ₃ U) U ₅ GC ₂ (AU) ₂ (A ₂ U)(A ₃ C)(A ₃ U)	17 19 19a
264-281	C ₂ (AG)(AC)(A ₂ U)(A ₃ U)(A ₄ U)	C ₂ (AC)(AG)(A ₂ U)(A ₃ U)(A ₄ U)	36
311-331	U ₃ GC ₆ (AC)(A ₂ C) ₃	U ₂ GC ₅ (AG)(A ₂ C) ₃	34
341-356	U ₅ GC ₆ (AC)(AU)	U ₄ GC ₅ (AC)(AU)	56*
482-498	U ₇ GC ₂ (AC)(A ₄ C)	U ₅ GC ₂ (AC)(A ₄ N)	57
531-546	U ₈ GC(AC)(AU) ₂	U ₄ GC ₅ (AC)(AU)	56*
622-634	GC ₃ (AC) ₃ (A ₂ U)	GC ₂ (AC) ₃ (A ₂ U)	51
672-685	U ₄ C ₃ (AC)(AU)(AC)	U ₃ C ₂ (AC) ₂ (AU)(AG)	54 ⁺
949-963	U ₃ (A ₂ G)(A ₃ C)(A ₄ U)	U ₂ (A ₂ G)(A ₃ C)(A ₄ N)	55
1084-1096	U ₅ GC ₅ (AU)	U ₄ GC ₅ (AC)(AU)	56*
1577-1590	U ₃ C ₃ (AC) ₂ (AU)(AG)	U ₃ C ₂ (AC) ₂ (AU)(AG)	54 ⁺

Comparison of ribonuclease T₁ oligonucleotides predicted from the cDNA sequence shown in fig.2, and those reported by Lai et al. (22) from MHV-A59 RNA's 5, 6 and 7. Spot 19 was found only in RNA7, spot 19a only in RNA 6 and 5; all the others were reported in all 3 RNA's. For spots 54 (+) and 56 (*), more than one region of the cDNA sequence showed similarity to the reported compositions.

G Q P K Q V T K Q S A K K V R Q K I L N
 GGCCAGCCCCAAGCAAGTAACGAAGCAAAGTGCCAAAAAAGTCAGGCAGAAAAATTTAAAC
 910 920 930 940 950 960

K P R Q K R T P N K Q C P V Q Q C F G K
 AAGCCTCGCCAAAAGAGGACTCCAACAACAGCAGTGCCCCAGTCAGCAGTGTTTTGGAAAG
 970 980 990 1000 1010 1020

R G P N O N F G G S E M L K L G T S D P
 AGAGGCCCAATCAGAATTTTGGAGGCTCTGAAATGTTAAAACTTGGAACTAGTGATCCA
 1030 1040 1050 1060 1070 1080

Q F P I L A E L A P T V G A F F F G S K
 CAGTTCCCCATTCTTGCAGAGTTGGCTCCAACAGTTGGTGCCTTCTTCTTTGGATCTAAA
 1090 1100 1110 1120 1130 1140

L E L V K K N S G G A D E P T K D V Y E
 TTAGAATTGGTCAAAAAGAATTCTGGTGGTGCTGATGAACCCACCAAAGATGTGTATGAG
 1150 1160 1170 1180 1190 1200

L Q Y S G A V R F D S T L P G F E T I M
 CTGCAATATT CAGGTGCAGTTAGATTGATAGTACTCTACCTGGTTTTGAGACTATCATG
 1210 1220 1230 1240 1250 1260

K V L N E N L N A Y Q K D G G A D V V S
 AAAGTGTGAATGAGAATTTGAATGCCTACCAGAAGGATGGTGGTGCAGATGTGGTGAGC
 1270 1280 1290 1300 1310 1320

P K P Q R K G R R Q A Q E K K D E V D N
 CCAAAGCCCCAAAGAAAAGGGCGTAGACAGGCTCAGGAAAAGAAAGATGAAGTAGATAAT
 1330 1340 1350 1360 1370 1380

V S V A K P K S S V Q R N V S R E L T P
 GTAAGCGTTGCAAAGCCCCAAAAGCTCTGTGCAGCGAAATGTAAGTAGAGAATTAACCCCA
 1390 1400 1410 1420 1430 1440

E D R S L L A Q I L D D G V V P D G L E
 GAGGATAGAAGTCTGTTGGCTCAGATCCTAGATGATGGCGTAGTGCCAGATGGGTTAGAA
 1450 1460 1470 1480 1490 1500

D D S N V *
 GATGACTCTAATGTGTAAGAGAATGAATCCTATGTCGGCGCTCGGTGGTAACCCCTCGCG
 1510 1520 1530 1540 1550 1560
 AGAAAGTCGGGATAGGACACTCTCTATCAGAATGGATGCTTGTCTGCATAACAGATAGA
 1570 1580 1590 1600 1610 1620
 GAAGGTTGTGGCTGCCCTGTATCAATTAGTTGAAAGATTTGCAAAATAGAGAATGTGTG
 1630 1640 1650 1660 1670 1680
 AGAGAAGTTAGCAAGGTCCTACGTCTAACCATAAGAACGGCGATAGGCCGCCCTGGGAA
 1690 1700 1710 1720 1730 1740
 GAGCTCACATCAGGCTACTATTCTTGC AATGCCCTAGTAAATGAATGAAGTTGATCATGG
 1750 1760 1770 1780 1790 1800
 CCAATTGGAAGAATCAAAAAAAAAAAAAAAAAAAAAA
 1810 1820 1830 1840

Fig.2 Sequence of the MHV-A59 nucleocapsid gene and protein.

components of all three oligonucleotides are ordered to maximise homology with the cDNA sequence (Fig.3), the predicted sequences are consistent with the earlier assignments of spots 19 and 19a by Lai et al. (22) and in addition suggest that spot 17 represents an internal region from the larger RNA's which is immediately upstream from the nucleocapsid gene. It is of interest that clone S9 diverges from the consensus sequence (Fig.1) at almost exactly the inferred site of divergence between oligonucleotides 19 and 17, suggesting that the divergence is due not to a cloning artefact, but to its originating from a different viral RNA.

Sequence of the nucleocapsid gene

The extreme 3' gene of MHV-A59, corresponding to RNA7, is known to encode the viral nucleocapsid protein (7,13). Translation of the sequence (Fig.2) gives only one long open reading frame, from nucleotides 154 to 1515, predicting a protein of mol. wt. 49,660 which is enriched in basic residues. These features are entirely consistent with the observed electrophoretic mobility of the nucleocapsid protein (2) and its function of binding to viral RNA. There is also a short open reading frame between nucleotides 218 and 488, predicting a polypeptide of 90 amino acids, but there is at present no evidence for the existence of such a protein.

A search for homology with the nucleocapsid gene sequences of snowshoe hare Bunya virus (23), Sindbis virus (24), influenza virus (25) and vesicular stomatitis virus (26) using the computer program SEQFIT (21) revealed no significant similarities confirming that the coronaviruses form a distinct viral group.

This is the first primary structure, of either nucleic acid or protein, to be determined from a coronavirus. Experiments are now in progress, using restriction fragments prepared from the clones described above, to determine the precise sequence relationships between the virion and messenger RNA's, and

Sequence No	140	150	Met
Oligonucleotide	G T A G G T A A T C T A A A C T T T A A G G A T G		
19	[T,T,A A A T,C]	T A A T C T A A A C T T T A A G	
19a	[T,T,A A A T,C]	T A A T C T A A A C T A T A T G	
17	[A A C,C]	T A A T C T A A A C T T T A A G	

Figure 3
Alignment of oligonucleotides described by Lai et al. (22) with the sequence from Fig.2. Oligonucleotide 19 was found uniquely in RNA7, 19a uniquely in RNA6 and 17 in all the viral RNA's except RNA7 (6,22).

to done, specifically, the gene adjacent to the nucleocapsid gene, which encodes the viral glycoprotein E1 (13,14).

ACKNOWLEDGEMENTS

We are grateful to Willy Spaan for the RNA preparation, and to Ben van der Zeijst for extensive discussion. P.R. was supported at the EMBL by a short-term fellowship from the European Molecular Biology Organisation, and J.Å. I. by a European Fellowship from the Royal Society. We thank Annie Steiner for typing the manuscript.

REFERENCES

1. Ter Meulen, V., Siddell, S. and Wege, H. Eds. (1981) *Biochemistry and biology of coronavirus*. Advances in Experimental Medicine and Biology, vol. 142, Plenum Press, New York.
2. Sturman, L.S., Holmes, K.V. and Behnke, J. (1980) *J. Virol.* 42, 449-462
3. Niemann, H. and Klenk, H.-D. (1981) *J. Mol. Biol.* 153, 993-1010.
4. Holmes, K.V., Doller, E.W. and Sturman, L.S. (1981) *Virology* 115, 334-344
5. Rottier, P.J.M., Horzinek, M.C. and van der Zeijst, B.A.M. (1981) *J. Virol.* 40, 350-357.
6. Spaan, W.J.M., Rottier, P.J.M., Horzinek, M.C. and van der Zeijst, B.A.M. (1982) *J. Virol.* 42, 432-439
7. Cheley, S., Anderson, R., Cupples, M.J., Lee Chan, E.C.M. and Morris, V.L. (1981) *Virology* 112, 596-604
8. Lai, M.M.C., Brayton, P.R., Armen, R.C., Patton, D.D., Pugh, C. and Stohman, S.A. (1981) *J. Virol.* 39, 823-834
9. Leibowitz, J.L., Wilhelmsen, K.C. and Bond, C.W. (1981) *Virology* 114, 39-51.
10. Stern, D.F. and Kennedy, S.I.T. (1980) *J. Virol.* 34, 665-674
11. Stern, D.F. and Kennedy, S.I.T. (1980) *J. Virol.* 36, 440-449
12. Wege, H., Siddell, S., Sturm, S. and Ter Meulen, V. (1981) *J. Gen. Virol.* 54, 213-217
13. Rottier, P.J.M., Spaan, W.J.M., Horzinek, M.C. and van der Zeijst, B.A.M. (1981) *J. Virol.* 38, 20-26
14. Siddell, S.G., Wege, H., Barthel, A. and Ter Meulen, V. (1980) *J. Virol.* 33, 10-17
15. McIntosh, K. (1974) *Curr. Top. Microbiol. Immunol.* 63, 85-129.
16. Holmes, K.V. and Behnke, J.N. (1981) in Ref.1
17. Spaan, W.J.M., Rottier, P.J.M., Horzinek, M.C. and van der Zeijst, B.A.M. (1981) *Virology* 108, 424-434
18. Messing, J. (1982) In *Genetic Engineering: principles and methods*, Setlow J.K. and Hollaender, A., eds. Vol. IV, pp.19-36, Plenum Press, New York.
19. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.* 143, 161-172
20. Garoff, H. and Ansoerge, W. (1981) *Anal. Biochem.* 115, 450-457
21. Staden, R. (1980) *Nucl. Acids Res.* 8, 3673-3694
22. Lai, M.C., Patton, C.D. and Stohman, S.A. (1982) *J. Virol.* 41, 557-565
23. Bishop, D.H.L., Gould, K.G., Akashi, H. and Clerx-van Haaster, C.M. (1982) *Nucl. Acids Res.* 10, 3703-3712
24. Rice, C.M. and Strauss, J.H. (1981) *Proc. Natl. Acad. Sci. USA* 78, 2062-2066.
25. Winter, G. and Fields, S. (1981) *Virology* 114, 423-428
26. McGeoch, D.J. and Turnbull, N.T. (1978) *Nucl. Acids Res.* 5, 4007-4024