VIROLOGY 193, 520–523 (1993)

Sequence Analysis of the Nucleocapsid Protein Gene of Rat Coronavirus SDAV-681[1]

SATOSHI KUNITA,[2] MIKI MORI, AND EIJI TERADA

Department of Laboratory Animal Science, School of Hygienic Sciences, Kitasato University, Kitasato, Sagamihara 228, Japan

Received September 14, 1992; accepted October 28, 1992

The nucleotide sequence of the 3'-end of the genomic RNA of sialodacryoadenitis virus strain 681 (SDAV-681) was determined. A large open reading frame encoding a 454-amino-acid protein was identified as the nucleocapsid protein (N) gene, since the predicted protein is similar in size, chemical properties, and amino acid sequence to the N proteins of other coronaviruses. The amino acid variance of the N proteins between SDAV and mouse hepatitis virus (MHV) is not markedly different from that among MHV strains. A high degree of genetic relatedness between SDAV and MHV was revealed in the intergenic and 3'-noncoding sequences as well as in the N gene.

Sialodacryoadenitis virus (SDAV) is a member of the family Coronaviridae and is known to be an important pathogen which infects laboratory rats at high prevalence, causes clinically significant disease associated with rhinitis and sialodacryoadenitis, and interferes with research that uses rats. It shares antigenic relationships with other rodent coronaviruses, such as Parker's rat coronavirus (PRCV) and mouse hepatitis virus (MHV) (*1*). The genomic organization and the unique strategy of replication of MHV have been well characterized (*2*). The virion contains a single-stranded, positive-sense RNA genome of approximately 32 kilobases (kb). The genome is polyadenylated and organized into seven regions, each containing one or more open reading frames (ORFs). In infected cells, a genomic-sized mRNA and six subgenomic mRNA species are synthesized which form a nested-set with a common 3'-terminus but extending to different lengths in the 5'-direction. Only the 5'-proximal gene of each mRNA is utilized for translation. The genes encoding the structural viral proteins have been identified; the spike glycoprotein (S) gene is third from the 5'-end (gene 3), the membrane glycoprotein (M) gene is sixth from the 5'-end (gene 6), and the nucleocapsid protein (N) gene is the 3'-terminal gene (gene 7). Additionally, the hemagglutinin–esterase glycoprotein (HE) gene (gene 2-1) is translated by only some of the MHV strains (*3*). In contrast to MHV, the molecular biology of SDAV is not understood. To elucidate the structure and the molecular biological properties of SDAV, we have initiated sequence analysis of the SDAV genome. We report here the complete nucleotide sequence of the N gene of SDAV-681 and its relatedness to other coronaviruses.

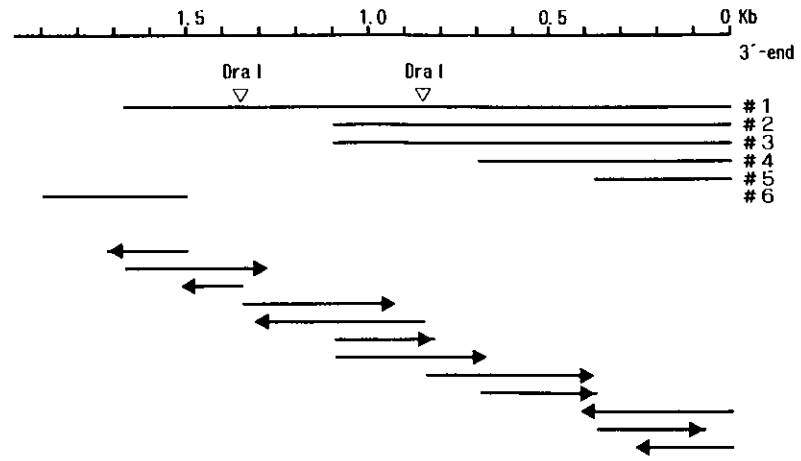
SDAV-681 (*1*) was propagated in a rat mammary tumor cell line (LBC) as described by Hirano *et al.* (*4*). Virus was purified by centrifugation on a discontinuous sucrose gradient consisting of 60 and 30% (w/w) sucrose in TNE buffer (50 m*M* Tris–HCl, pH 7.5, 100 m*M* NaCl, 1 m*M* EDTA) at 110,000 *g* for 2 hr at 4°. The virus recovered from the interface between 60 and 30% sucrose was incubated in TNE buffer containing 0.5 mg/ml proteinase K and 0.5% SDS for 10 min at 56°. The genomic RNA was extracted with phenol–chloroform–isoamyl alcohol and precipitated with ethanol. Double-stranded cDNA was synthesized from SDAV genomic RNA primed with oligo(dT), using cDNA synthesis system plus (Amersham). After addition of *Bam*HI adaptor (Pharmacia) to blunt-ended cDNA, the cDNA was purified on a cDNA spun column (Pharmacia) and ligated into the *Bam*HI-site of plasmid vector pUC19 (Takara). The ligated molecules were used to transform *Escherichia coli* JM109 and cells containing recombinant plasmid were selected on LB agar plates that contained 50 μg ampicillin/ml, 1 m*M* IPTG, and 0.01% X-gal. Recombinant colonies were transferred to nitrocellulose filters and probed with ^{32}P-labeled MHV-N gene cDNA (1362 nucleotides) prepared from MHV-A59 genomic RNA by polymerase chain reaction (PCR) described previously (*5*). Colonies yielding strong signals were analyzed for insert size. Five clones containing 0.4- to 1.7-kb inserts (#1–#5 in Fig. 1A) were sequenced in both directions by the dideoxy method using modified T7 DNA polymerase (Sequenase, U.S. Biochemical). An additional cDNA clone (#6 in Fig. 1A) was obtained using a synthetic oligonu-

[1] Sequence data from this article have been deposited with the DDBJ, EMBL, and GenBank Libraries under Accession No. D10760.

[2] To whom reprint requests should be addressed.

0042-6822/93 $5.00

A



B

```
5'-TTGTTGAGAATCTAATCTAAACTTTAAGGATGTCTTTTGTTCCCGGACAA      50
GAAAACGCCGGTAGCAGAAGCTCCTCTGGAAACCGCGCTGGTAATGGAATCCTCAAGAAG  110
ACCACTTGGGCTGACCAAACCGAGCGCGGACAAAATAATGGAAATAGAGGCAGAAGGAAT  170
CAGCCCAAGCAGACTGCAACTACTCAGCCCAATACCGGGAGTGTGGTTCCCCATTACTCT  230
TGGTTTTCGGGCATTACCCAATTCCAGAAGGGAAAAGAGTTCCAGTTTGCAGGTGGACAA  290
GGAGTGCCTATTGCCAATGGAATCCCACCTTCTGAGCAAAAGGGATATTGGTATAGACAC  350
AACCGTCGTTCTTTTAAAACACCTGATGGGCAGCAGAAGCAACTACTCCCCAGATGGTAT  410
TTTTACTATCTTGGGACGGGCCCCCATGCTGGAGCCAGTTTCGGAGACAGCATTGAGGGA  470
GTCTTCTGGGTTGCAAATAGTCAGGCGGATACCAACACCTCTGCTGACATTGTTGAAAGG  530
GACCCAAGTAGCCATGAGGCTATTCCTACTAGGTTTGCGCCCGGTACGGTATTGCCTCAG  590
GGTTTCTATGTTGAAGGCTCGGGAAGGTCTGCACCTGCTAGTCGATCTGGTTCGCGGTCA  650
CAATCCCGTGGGCCAAATAATCGCGCTAGAAGCAGTTCCAACCAGCGCCAGCCTGCCTCT  710
ACTGTAAAACCTGATATGGCCGAAGAAATTGCTGCTCTTGTTTTGGCTAAGCTAGGCAAA  770
GATGCCGGACAGCCTAAGCAAGTAACTAAGCAAAGTGCCAAAGAAGTCAGGCAGAAAATT  830
TTAAATAAGCCTCGCCAAAAGAGGACTCCAAACAAGCAGTGCCCAGTGCAGCAGTGTTTT  890
GGAAAGAGAGGCCCCAATCAGAATTTTGGAGGCCCTGAAATGTTAAAACTTGGAACTAGT  950
GATCCACAGTTCCCCATTCTTGCAGAGTTGGCCCCAACACCTGGTGCCTTCTTCTTTGGA 1010
TCTAAATTAGAATTGGTCAAAAAGAATTCTGGTGGCGTTGATGAACCCACCAAAGATGTG 1070
TATGAGCTGCAATATTCAGGTGCAGTCAGATTTGATAGTACTCTACCTGGTTTTGAGACT 1130
ATCATGAAAGTGTTGAATGAGAATTTGAACGCCTACCAGAATCAAGCTGGTGGTGCAGAT 1190
GTAGTGAGCCCAAAGCCCCAAAGAAAGAGAGGGACGAAACAAACGGCTCAGAAAGAAGAA 1250
TTAGATAGTATAAGCGTTGCAAAGCCCAAAAGTGCCGTGCAGCGAAATGTAAGCAGAGAA 1310
TTAACCCCAGAGGATAGAAGCCTGTTGGCGCAGATCCTAGATGATGGCGTTGTGCCTGAT 1370
GGGTTAGATGACTCTAATGTGTAAAGAGAATGAATCCTATGTCGGCGCTCGGTGGTAACC 1430
CCTCGCGAGAAAGTCGGGATAGGACACTCTCTATCAGAATGGATGTCTTGCTGTCATAAC 1490
AGATAGAGAAGGTTGTGGCAGACCCTGTATCAATTAGTTGAAAGAGATTGCAAAATAGAG 1550
AATGTGTGAGAGAGTGTAAGGTCCTACGTCTAACCATAAGAACGGCGATAGGCGCCCCCT 1610
GGGAAGAGCTCACATCAGGGTACTATTCCTGCAATGCCCTAGTAAATGAATGAAGTTGAT 1670
CATGGCCAATTGGAAGAATCAC-POLY(A)-3'                           1692
```

FIG. 1. Sequencing of the 3'-end of the SDAV-681 genome. (A) cDNA clones and sequencing strategy. Clone #1 was sequenced completely from subcloned fragments generated by restriction endonuclease *Dra*I. Clones #2–#6 were sequenced partly their 5'- or 3'-termini. (B) The primary nucleotide sequence of the 3'-end (1692 nucleotides) of the SDAV-681 genome. The initiation and termination codons for the N open reading frame are underlined and those for the internal open reading frame are shown by the wavy lines. The intergenic consensus sequence (TCTAAAC) and the 3'-conserved sequence (GGAAGAGC) are shown by the double lines. The region indicated by the dotted line was used for preparing a complementary oligonucleotide primer to obtain clone #6.

```
SDAV-681    MSFVPGQENA GSRSSSGNRA GNGILKKTTW ADQTERGQNN GNRGRRNQPK QTATTQPNTG SVVPHYSWFS     70
MHV-A59                 G    V                        P   Q                  S                  70
MHV-JHM                                               L   Q    K             S                  70
MHV-2                                                 --- -      H           A                  86
MHV-S                          S S                   AG                      S                  90
MHV-Nu67                                              PT  Q                  S                  90

SDAV-681    GITQFQKGKE FQFAGGQGVP IANGIPPSEQ KGYWYRHNRR SFKTPDGQQK QLLPRWYFYY LGTGPHAGA[S    140
MHV-A59                    E            A                                           |       140
MHV-JHM                    Q            A Q                                    Y    |E      140
MHV-2                      Q          S A                     H                     |E      136
MHV-S                     VQ            A                                           |E      140
MHV-Nu67                   D            A                                           [       140

SDAV-681    FGDSIEGVFW VANSQADTNT SA]DIVERDPS SHEAIPTRFA PGTVLPQGFY VEGSGRSAPA SRSGSRSQSR    210
MHV-A59     Y                     RS]                                                       210
MHV-JHM     Y  D    V    SQ  E R    |                                           P           210
MHV-2       Y  D    V    SQ    K  T | V                                                     206
MHV-S       Y  D    V    SQ    K  T |                                                       210
MHV-Nu67    Y                     RS]                                                       210

SDAV-681    GPNNRARSSS NQRQPASTVK PDMAEEIAAL VLAKLGKDAG QPKQVTKQSA KEVRQKILNK PRQKRTPNKQ    280
MHV-A59                                                                                     280
MHV-JHM                                                                                     280
MHV-2                          A                                     T                      276
MHV-S                                                                                       280
MHV-Nu67                                                                                    280

SDAV-681    CPVQQCFGKR GPNQNFGGPE MLKLGTSDPQ FPILAELAPT PGAFFFGSKL ELVKKNSGGV DEPTKDVYEL     350
MHV-A59                         S                        V                   A              350
MHV-JHM                                                  A                   A G            350
MHV-2                           S                         S                  A              346
MHV-S                           S                        A                   A        H     350
MHV-Nu67                        S                                                           350

SDAV-681    QYSGAVRFDS TLPGFETIMK VLNENLNAYQ[NQAGGADVVS PKPQRKRGTK QTAQKE]ELDS ISVAKPKSAV    420
MHV-A59                                     [K-D                  GRRQ AQEK D] V N V       S    419
MHV-JHM                                     |  D                     K     D|  V N V       S    420
MHV-2            I                          |  D SV L          P RGRRQ AQEK D|  V N V       L    416
MHV-S                                       [K-D                  GRRQ AQEK D] V N V       S    419
MHV-Nu67         I                          [  T  V               GRRQ AQEK D] V N V       S    420

SDAV-681    QRNVSRELTP EDRSLLAQIL DDGVVPDGL- DDSNV                                          454
MHV-A59                                    E                                                454
MHV-JHM                                    E                                                455
MHV-2                                      E                                                451
MHV-S                                      E                                                454
MHV-Nu67          D                  L     E                                                455
```

FIG. 2. Amino acid sequence comparison of the N proteins of SDAV-681 and five MHV strains. The deduced sequence of the SDAV-681 N protein is from this work. The sequences for MHV-2 and -Nu67 were obtained from recent work (*5*). The MHV-A59 and -S sequences are taken from Parker and Masters (*6*) and the MHV-JHM sequence is taken from Skinner and Siddell (*7*). Spaces indicate amino acids identical to the SDAV-681 sequence. Hyphens indicate gaps introduced to maximize alignment. The spacer regions A and B (amino acid positions 140–162 and 381–406 in the SDAV-681 sequence, respectively) proposed by Parker and Masters (*6*) are bracketed.

cleotide primer complementary to nucleotides 183–202 (Fig. 1B) for cDNA synthesis and then sequenced in the same manner.

The nucleotide sequence of the 3′-end of SDAV-681 genome covering the complete ORF of 1362 nucleotides is presented in Fig. 1B. This ORF extends from nucleotides 30 to 1391 and potentially encodes for a protein of 454 amino acids. The predicted protein is highly homologous to the N proteins of MHV strains (MHV-A59, -JHM, -2, -S, and -Nu67) aligned in Fig. 2. Nucleotide and amino acid sequences of SDAV-681 and MHV strains are 89.0 to 92.7% and 90.1 to 93.8% identical, respectively. Incidentally, the amino acid sequence identity among the N proteins of MHV strains is 91.8 to 97.1%. Coronaviruses can be divided into several antigenic groups. SDAV belongs to a group including MHV, bovine coronavirus (BCV), human coronavirus strain OC43 (HCV-OC43), and hemagglutinating encephalomyelitis virus of swine (HEV). The N protein of MHV shows an amino acid sequence homology of approximately 70% with BCV (*8*) and HCV-OC43 (*9*). The high degree of sequence identity between SDAV and MHV shown in this study confirmed the close relationship of these two viruses as revealed by serological analysis (*1, 10*). The predicted N protein of SDAV possesses some features which are characteristic of the N proteins of coronaviruses. That is to say, the protein is highly basic, suggesting RNA-binding property of this protein (*11*), and is rich in serine residues, which are presumed to be sites of phosphorylation (*12*). Sixty of the amino acid residues (13%) are basic and forty-two residues (9%) are acidic, giving the protein a net charge of +18 at neutral pH. There are thirty-nine serine residues (9%) located at the almost same positions as MHV.

SDAV-681 has an internal ORF in the +1 reading frame within the N gene (Fig. 1B). This ORF potentially encodes a protein of 207 amino acids, which is basic (a net +7 charge at neutral pH) and rich in hydrophobic amino acids (50%) especially leucine residues (18%). An internal ORF of identical size has been found within the N genes of MHV and BCV (*6, 8*). Recently, it was demonstrated for BCV that a 23-kDa protein encoded by the internal ORF lying within the N gene was ex-

pressed in virus-infected cells (*13*). Further studies are required to reveal the functional significance of this protein and whether similar proteins can be detected for other coronaviruses.

Noncoding regions flanking the N gene on SDAV genome were also compared with those of MHV. The 5′-noncoding sequence of 17 nucleotides (at nucleotide positions 13–29 in Fig. 1B) immediately before the initiation codon, including a stretch of consensus sequence (TCTAAAC), is identical to the intergenic sequence between the M and N genes of MHV (*14, 15*). This stretch of consensus sequence is conserved among most coronaviruses, and among the 5′-end of genome, the intergenic regions, and the mRNA leader sequences of the same coronavirus. Thus it is the proposed leader RNA binding site for the initiation of mRNA transcription (*2, 14, 15*). The SDAV N gene is flanked on its 3′-side by a noncoding region of 301 nucleotides (at nucleotide positions 1392–1692 in Fig. 1B) and a poly(A) tail. The 3′-noncoding sequence of SDAV is highly homologous (more than 97% sequence identity) to MHV (*6, 7*) and contains the conserved sequence, GGAAGAGC, 81 nucleotides from the 3′-end, which has been proposed as a recognition signal for negative-stranded RNA synthesis (*2*).

It has been reported that high frequency RNA recombination occurs among MHV strains during coinfection both *in vitro* (*16*) and *in vivo* (*17*). This recombination event is thought to be caused by reattachment of pausing RNA intermediates during RNA synthesis at sites of high homology (*2*). The conservation of nucleotide sequences between SDAV and MHV in the intergenic and 3′-noncoding regions as well as in the N gene suggests that these two viruses are at least structurally able to undergo RNA recombination with each other. Under experimental conditions, mice are susceptible to intranasal and contact infection with SDAV (*18, 19*) and rats are susceptible to MHV by intranasal inoculation (*20*), but the cell tropisms of two viruses are considerably different. Therefore, RNA recombination between SDAV and MHV is unlikely to occur frequently, it may be possible, in natural infections. Parker and Masters (*6*) suggested that the N protein of MHV consists of three domains separated by two spacer regions A and B, which have less constraint on their amino acid sequences, and these spacers tend to vary among a limited set of two or three alternatives. Interestingly, SDAV-681 has an A59-like spacer A and a JHM-like spacer B (Fig. 2), implying that the N gene of SDAV-681 may have been derived from two prototype N genes of MHVs. It is possible that SDAV may be a rat-adapted virus arisen by recombination between two ancestral MHVs possessing different N genes or that recombination among an ancestral rat coronavirus and two ancestral MHVs may have given rise to SDAV-681.

The results of the present study suggest a high degree of genetic relatedness between SDAV and MHV which is about the same as that among MHV strains. Comparison of the remaining genes with SDAV and MHV and detailed analysis of SDAV structural proteins will provide a clue to understand their evolution and the molecular basis of host cell tropism.

ACKNOWLEDGMENT

This work was supported partly by Grant-in-Aid for Scientific Research from Ministry of Education, Science, and Culture, Japan.

REFERENCES

1. Bhatt, P. N., Percy, D. H., and Jonas, A. M., *J. Infect. Dis.* **126,** 123–130 (1972).
2. Lai, M. M. C., *Annu. Rev. Microbiol.* **44,** 303–333 (1990).
3. Yokomori, K., Banner, L. R., and Lai, M. M. C., *Virology* **183,** 647–657 (1991).
4. Hirano, N., Ono, K., Sada, Y., Inoue, A., and Murakami, T., *Arch. Virol.* **85,** 301–304 (1985).
5. Kunita, S., Terada, E., Goto, K., and Kagiyama, N., *Lab. Anim. Sci.*, **42,** 593–598 (1992).
6. Parker, M. M., and Masters, P. S., *Virology* **179,** 463–468 (1990).
7. Skinner, M. A., and Siddell, S. G., *Nucleic Acids Res.* **11,** 5045–5054 (1983).
8. Lapps, W., Hogue, B. G., and Brian, D. A., *Virology* **157,** 47–57 (1987).
9. Kamahora, T., Soe, L. H., and Lai, M. M. C., *Virus Res.* **12,** 1–9 (1989).
10. Smith, A. L., *Lab. Anim. Sci.* **33,** 157–160 (1983).
11. Stohlman, S. A., Baric, R. S., Nelson, G. N., Soe, L. H., Welter, L. M., and Deans, R. J., *J. Virol.* **62,** 4288–4295 (1988).
12. Stohlman, S. A., and Lai, M. M. C., *J. Virol.* **32,** 672–675 (1979).
13. Senanayake, S. D., Hofmann, M. A., Maki, J. L., and Brian, D. A., *J. Virol.* **66,** 5277–5283 (1992).
14. Budzilowicz, C. J., Wilczynski, S. P., and Weiss, S. R., *J. Virol.* **53,** 834–840 (1985).
15. Shieh, C.-K., Soe, L. H., Makino, S., Chang, M.-F., Stohlman, S. A., and Lai, M. M. C., *Virology* **156,** 321–330 (1987).
16. Makino, S., Keck, J. G., Stohlman, S. A., and Lai, M. M. C., *J. Virol.* **57,** 729–737 (1986).
17. Keck, J. G., Matsushima, G. K., Makino, S., Fleming, J. O., Vannier, D. M., Stohlman, S. A., and Lai, M. M. C., *J. Virol.* **62,** 1810–1813 (1988).
18. Bhatt, P. N., Jacoby, R. O., and Jonas, A. M., *Infect. Immun.* **18,** 823–827 (1977).
19. Barthold, S. W., de Souza, M. S., and Smith, A. L., *Lab. Anim. Sci.* **40,** 481–485 (1990).
20. Taguchi, F., Yamada, A., and Fujiwara, K., *Arch. Virol.* **59,** 275–279 (1979).