Volume 12 Number 3 1984 Nucleic Acids Research

Nucleotide sequence of the dsRNA genomic segment 7 of Simian 11 rotavirus

G.W.Both, A.R.Bellamy* and L.J.Siegman

CSIRO Division of Molecular Biology, P.O. Box 184, North Ryde, N.S.W. 2113, Australia, and
*Department of Cell Biology, University of Auckland, Private Bag, Auckland, New Zealand

Received 21 November 1983; Accepted 19 December 1983

ABSTRACT

A full length cDNA copy of dsRNA segment seven of Simian 11 rotavirus has been obtained by standard molecular cloning techniques. Segment seven codes for the non-structural viral protein NCVP4 and is 1104 nucleotides in length with putative 5'- and 3'- terminal non-coding regions of 25 and 134 residues respectively. The longest open reading frame of 315 codons extends from nucleotide 26 to 970 inclusive. However, the presence nearby of two other AUG codons makes it unclear which codon is used for initiation. The second AUG conforms to the Kozak consensus sequence and if utilised, would yield a protein 312 amino acids in length with a nett charge at pH7 of -2.5. Determination of the gene 7 sequence indicates that terminal sequence conservation among rotavirus gene segments is limited to three and two nucleotides at the 5' and 3' ends of the plus strand, respectively.

INTRODUCTION

Simian 11 rotavirus (SA11), a member of the Reoviridae family posesses a genome of eleven segments of double-stranded (ds) RNA. Six dsRNA segments encode structural proteins and five code for non-structural polypeptides in the infected cell. In common with other members of the Reoviridae and paramyxoviruses e.g. influenza, the RNA segments of rotaviruses reassort independently during infection, a property demonstrated by the high rate of genetic interchange found amongst the virus progeny derived from cells mixedly infected with two different virus strains (1,2). The mechanism for genetic reassortment and allocation of the eleven genomic segments to each progeny virus particle is not known.

The determination of nucleotide sequences of influenza virus genes revealed the potential multiple coding capacities of some segments (3). Similar studies on rotavirus gene segments may complement future work on the processes of infection and particle morphogenesis and are essential for the understanding of the structure and function of rotavirus genes and any features recognised during reassortment. Recently we devised a generalised

© IRL Press Limited, Oxford, England.

strategy for cloning dsRNA (4), and similar procedures since have been applied to other members of the Reoviridae family (5,6,7,8). For SA11 virus, the availability of cloned copies of various genomic segments enabled the definitive assignment of proteins NCVP4, NCVP3 and VP7 to genomic segments 7, 8 and 9, respectively (9) and the determination of the nucleotide sequences of segments 8 and 9 (4,9). Here we describe the gene structure and nucleotide sequence for segment 7 which codes for the non-structural protein NCVP4. We also discuss the partial conservation of terminal sequences between genomic segments of SA11 and other rotaviruses.

METHODS

Production and identification of gene clones

Methods for cloning cDNA copies of rotavirus genomic dsRNA segments into the PstI site of pBR322 have been described (4). A DNA clone of dsRNA segment 7 (G17) was identified by Northern hybridisation analysis in which radioactively-labelled plasmid G17 was annealed to SA11 genomic segments that had been resolved by electrophoresis and immobilised by blotting on to DBM paper (4).

DNA sequencing

The nucleotide sequence of cloned gene segment 7 was determined by a combination of the three methods described previously (4) (see Legend to Figure 1).

RESULTS

Assignment of NCVP4 to genomic segment 7

The non-structural protein NCVP4 was previously assigned to segment 7 of SA11 using the recombinant plasmid G17 to prepare gene 7-specific mRNA by hybrid selection. *In vitro* translation of this mRNA in a reticulocyte lysate system in the presence and absence of canine pancreatic microsomes clearly identified NCVP4 as the polypeptide product of segment 7 (9).

The nucleotide sequence of segment seven

Preliminary sizing of the insert in the recombinant plasmid G17 by PstI excision and agarose gel electrophoresis demonstrated that the insert was close to the predicted size (1080-1120 nucleotides) for a full length copy of segment seven (4). The sequence of the gene was therefore determined by the strategy outlined in Figure 1. Most of the sequence was determined either from both DNA strands of the clone or by a combination of two independent methods, except for a small region from base 1010 onward

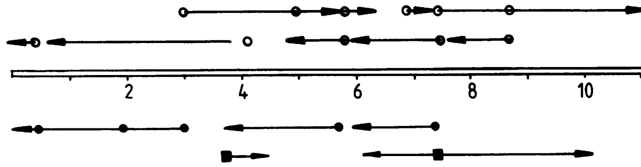

Figure 1. Strategy for determining the nucleotide sequence of a cloned copy of SA11 gene segment 7. The numbers refer to the distance in nucleotides x 10^{-2}. The sequence was determined by copying cDNA from SA11 mRNA (●) (9), by the Maxam and Gilbert method (○) (13), or by the dideoxy method (14) after subcloning into phage M13mp8 (15) (■).

(beyond the termination codon) where the data from one method were unambiguous. At each end of the cloned DNA, we observed the homopolymer tracts of nucleotides expected from the cloning strategy used (4). Moreover, since the 5'-terminal sequence obtained from the clone is in agreement (save for one base, see below) with that obtained from the messenger RNA, clearly the cloned copy of segment 7 includes all of the 5'-terminal sequence of the segment. Similarly the presence of the homopolymer runs and a partly conserved nucleotide sequence at the 3' end of the plus strand (see below) indicate that the clone is full length. The complete nucleotide sequence of the plus strand of segment 7 and the protein sequence it predicts are presented in Figure 2. The 3'- and 5'-terminal non-coding sequences of segment 7 are 134 and 25 nucleotides in length, respectively. However, the latter assumes that the first AUG codon (bases 26-28) of the open reading frame is used for the initiation of translation. There are two other AUG codons at nucleotides 35-37 and 53-55, all in the same reading frame. If the first initiation codon is utilised the gene would code for a polypeptide 315 amino acids in length, corresponding to a molecular weight 34.6 k daltons.

DISCUSSION

The biological function of NCVP4 during rotavirus infection is not understood. The nucleotide sequence of segment 7 predicts that NCVP4 could be a protein of 315 amino acids in length, but some uncertainty accompanies this prediction. In the absence of data on the aminoterminal residues of NCVP4, we are unable to determine which of the three methionine codons in the sequence is utilised for initiation. Of the three potential initiation sites, only the second conforms fully to the eukaryotic consensus flanking sequences proposed by Kozak (10) in which a purine (usually A) occupies

```
                          M  L  K  M  E  S  T  Q  Q  M  A  V  S  I  I  N  S  S  F  E  A  A
GGCAUUUAAUGCUUUUCAGUGGUUGAUGCUCAAGAUGGAGUCUACGCAACAGAUGGCCGUCUCAAUUAUUAACUCUUCUUUUGAAGCUGC
         10        20        30        40        50        60        70        80        90

 V  V  A  A  T  S  A  L  E  N  M  G  I  E  Y  D  Y  Q  D  I  Y  S  R  V  K  N  K  F  D  F
AGUUGUAGCUGCAACCUCAGCUCUUGAGAAUAUGGGAAUAGAAUAUGAUUAUCAGGAUAUAUUCUAGAGUAAAGAAUAAAUUUGAUUU
        100       110       120       130       140       150       160       170       180

 V  M  D  D  S  G  V  K  N  N  P  I  G  K  A  I  T  I  D  Q  A  L  N  N  K  F  G  S  A  I
UGUGAUGGACGAUUCUGGUGUUAAAAAUAAUCCGAUUGGUAAAGCAAUAACUAUUGAUCAAGCUUUGAAUAAUAAAUUUGGAUCUGCUAU
        190       200       210       220       230       240       250       260       270

 R  N  R  N  W  L  A  D  T  S  R  P  A  K  L  D  E  D  V  N  K  L  R  M  M  L  S  S  K  G
.AAGAAAUAGAAACUGGCUUGCUGAUACUUCUAGACCAGCUAAAUUGGAUGAGGAUGUAAACAAACUAAGAAUGAUGUUAUCAUCAAAAGG
        280       290       300       310       320       330       340       350       360

 I  D  Q  K  M  R  V  L  N  A  C  F  S  V  K  R  I  P  G  K  S  S  S  I  I  K  C  T  K  L
.AAUUGAUCAAAAAAUGAGAGUUUUAAACGCAUGCUUCAGUGUAAAAAGAAUACCUGGAAAAUCAUCAUCUAUUAUUAAAUGCACAAAAUU
        370       380       390       400       410       420       430       440       450

 M  R  D  K  L  E  R  G  E  V  E  V  D  D  S  F  V  D  E  K  M  E  V  D  T  I  D  W  K  S
.GAUGCGUGAUAAAUUGGAACGUGGUGAAGUUGAAGUGGAUGAUUCAUUUGUGGAUGAAAAAAUGGAAGUGGAUACCAUUGACUGGAAAUC
        460       470       480       490       500       510       520       530       540

 R  Y  E  Q  L  E  Q  R  F  E  S  L  K  S  R  V  N  E  K  Y  N  N  W  V  L  K  A  R  K  M
GCGCUAUGAGCAAUUGGAGCAAAGGUUUGAAUCAUUGAAAUCCAGGGUAAAUGAAAAAUAUAAUAAUUGGGUGUUGAAAGCAAGAAAAAU
        550       560       570       580       590       600       610       620       630

 N  E  N  M  H  S  L  Q  N  v  I  P  Q  Q  Q  A  H  I  A  E  L  Q  V  Y  N  N  K  L  E  R
.GAAUGAAAAUAUGCAUUCUCUUCAAAAUGUCAUCCCUCAACAGCAAGCACAUAUAGCUGAGCUUCAAGUGUACAAUAAUAAACUAGAACG
        640       650       660       670       680       690       700       710       720

 D  L  Q  N  K  I  G  S  L  T  S  S  I  E  W  Y  L  R  S  M  E  L  D  P  E  I  K  A  D  I
.UGAUUUGCAAAAUAAAAUUGGAUCCCUUACUUCUUCGAUUGAAUGGUAUUUAAGAUCAAUGGAAUUAGACCCUGAAAUAAAGGCAGACAU
        730       740       750       760       770       780       790       800       810

 E  Q  Q  I  N  S  I  D  A  T  N  P  L  H  A  F  D  D  L  E  S  V  I  R  N  L  I  S  D  Y
.UGAACAGCAAAUUAACUCAAUUGAUGCGAUAAAUCCAUUGCACGCUUUUGAUGACUUAGAAUCAGUAAUACGUAAUUUGAUAUCUGAUUA
        820       830       840       850       860       870       880       890       900

 D  K  L  F  L  M  F  K  G  L  I  Q  R  C  N  Y  Q  Y  S  F  G  C  E  *
.UGACAAAUUAUUCCUUAUGUUCAAAGGAUUAAUACAGAGAUGUAAUUAUCAAUAUUCAUUUGGUUGCGAAUAACCAUUUUGAUACAUGUU
        910       920       930       940       950       960       970       980       990

.GAACAAUCAAAUACAGUGUUGUAUGUUGUCAUCUAUGCAUAACCCUCUAUGAGCACAAUAGUUAAAAGCUAACACUGUCAAAAACCUAAA
       1000      1010      1020      1030      1040      1050      1060      1070      1080

.UGGCUAUAGGGGCGUUAUGUGGCC
       1090      1100
```

Figure 2. Nucleotide sequence of the (+) (coding) strand of SA11 RNA gene 7 and the predicted amino acid sequence of the NCVP4 protein.

position -3 and a G occupies position +4, i.e. AXXAUGG. NCVP4 is clearly discernable among the ^{35}S-methionine-labelled polypeptide products synthesized in a reticulocyte lysate in response to total SA11 mRNA (see Fig. 1, ref. 9). Proteins NCVP3 and pVP7 are not, although they are of similar size. The trivial explanation of relative methionine content is ruled out since NCVP4, NCVP3 and pVP7 have 14, 10 and 10 methionine

residues, respectively (4,9). Thus gene 7 mRNA appears to be translated very efficiently and the second AUG codon in the sequence is the one most likely to be utilised for initation. This would yield a protein of 34.3 k daltons, 312 amino acids in length. Examination of all three reading frames of the gene 7 nucleotide sequence reveals that the next longest open reading frame is only 41 codons in length. Thus, it is extremely unlikely that gene 7 codes for any protein other than NCVP4. This situation is similar to that observed for SA11 genes 8, 9 and 10 (4,9,12) and indicates that the rotavirus genes sequenced so far are truly monocistronic. Assuming that glutamic and aspartic acid carry a charge of -1 and arginine, lysine and histidine are +1, +1 and +0.5, respectively at pH7, NCVP4 would possess a nett charge of -2.5 making it a slightly acidicprotein in contrast to NCVP3, which by the same calculation is strongly basic (4).

Finally, the terminal sequences for gene 7 each differ by one nucleotide from all other rotavirus genes examined so far, i.e. 5' GGCAUU....UGUGGCC3' as compared with 5' GGCUUU....UGUGACC3', respectively (4,6,8,9,11,12 and Both and Bellamy, unpublished results). The 5' sequence was twice determined indirectly from the mRNA and reflects the sequence for the majority of molecules in a population. However, the 3' sequence was determined only in one cloned gene copy and could represent a point mutation. Nevertheless, this suggests that more extensive conservation of terminal sequences such as that observed for influenza virus RNA segments (3) is not required for replication, packaging or transcription of rotavirus dsRNA molecules.

ACKNOWLEDGEMENT

Grants from the New Zealand Medical Research Council, New Zealand Child Health Research Foundation and World Health Organisation supported this work. We thank Elizabeth Hamilton for excellent assistance.

REFERENCES

1. Estes, M.K., Palmer, B.L. and Obijeski, J.F. (1982) Curr. Top. Microbiol. Immunol. in press.
2. Holmes, I.H. (1983) In The Reoviridae, W.K. Joklik Ed., pp. 359-423, Plenum Pub. Co. New York.
3. Webster, R.G., Laver, W.G., Air, G.M. and Schild, G.C. (1982) Nature 296, 115-121.
4. Both, G.W., Bellamy, A.R., Street, J.E. and Siegman, L.J. (1982) Nucleic Acids Res. 10, 7075-7088.
5. Cashdollar, L.W., Esparza, J., Hudson, G.R., Chmelo, R., Lee, P.W.K. and Joklik, W.K. (1982) Proc. Natl. Acad. Sci. USA. 79, 7644-7648.
6. Imai, M., Richardson, M.A., Ikegami, N., Shatkin, A.J. and Furuichi,

Y. (1983) Proc. Natl. Acad. Sci. USA 80, 373-377.
7. Richardson, M.A. and Furuichi, Y. (1983) Nucleic Acids Res. 11, 6399-6408.
8. Elleman, T.C., Hoyne, P.A., Dyall-Smith, M.L., Holmes, I.H. and Azad, A.A. (1983) Nucleic Acids Res. 11, 4689-4701.
9. Both, G.W., Mattick, J.S. and Bellamy, A.R. (1983) Proc. Natl. Acad. Sci. USA 80, 3091-3095.
10. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5251.
11. Both, G.W., Mattick, J.S., Siegman, L.J., Atkinson, P.H., Weiss, S., Bellamy, A.R., Street, J.E. and Metcalf, P. (1983) in The Double-Stranded RNA Viruses, D.H.L. Bishop and R.W. Compans Eds., pp. 73-82, Academic, New York.
12. Both, G.W., Siegman, L.J., Bellamy, A.R. and Atkinson, P.H. (1983) J. Virology, in press.
13. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
15. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.