Structure of Rearranged Genome Segment 11 in Two Different Rotavirus Strains Generated by a Similar Mechanism

By SILVIA ADRIANA GONZÁLEZ, 1 NORA MARTA MATTION, 2 RODOLFO BELLINZONI 2 AND OSCAR ROBERTO BURRONE 1,*

1 Instituto de Investigaciones Bioquímicas, Fundación Campomar–CONICET and Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Avenida Patricia Argentinas 435, 1403 Buenos Aires and 2 Centro de Virologia Animal (CEVAN–CONICET), Serrano 661, 1414 Buenos Aires, Argentina

(Accepted 13 February 1989)

SUMMARY

The structures of the rearranged genomic segment 11 of two spontaneous swine rotavirus strains were determined. We found that the rearrangements involved the duplication of normal segment 11 in a head-to-tail orientation, and partial deletions in both monomers. The open reading frame for VP11, the protein encoded by normal segment 11, was maintained. We also showed that the two rearranged genes were transcribed into RNA molecules of the same length as their corresponding genomic segments.

INTRODUCTION

Rotaviruses, recognized as the major cause of acute viral gastroenteritis in human infants and in young animals of many species, possess a genome of 11 segments of dsRNA which can be separated on polyacrylamide gels to give characteristic patterns (Espejo et al., 1980; Estes et al., 1984). Rotaviruses with genome rearrangements have been reported in humans, calves, rabbits and pigs (Pedley et al., 1984; Besselaar et al., 1986; Pocock, 1987; Thouless et al., 1986; Bellinzoni et al., 1987). They have also been generated in tissue culture after serial passage at high multiplicity of infection (Hundley et al., 1985). In most cases, the RNA profiles were modified by the disappearance of an RNA segment and the appearance of a larger one. However, the structure of no rearranged gene has as yet been determined.

Recently, we described four naturally occurring swine rotavirus strains (C60, C134, C117 and Cc86) with rearrangements involving sequences of genomic segment 11 (gs 11) which led to the formation of higher M, segments (Bellinzoni et al., 1987; Mattion et al., 1988). The rearranged segments, named X1 (for strains C60 and C134) and X2 (for strains C117 and Cc86), have a size of approximately 1200 and 1000 bp, respectively, whereas gs 11 is 667 bp long for the UK bovine (Ward et al., 1985) and simian SAI 1 (Mitchell & Both, 1988) strains and 663 bp long for the human Wa strain (Imai et al., 1983).

We have shown that the rearranged genes X1 and X2 are not mosaic structures of the remaining segments and also that VP11, the protein encoded by normal segment 11, is present in cells infected with the rearranged strains (Mattion et al., 1988).

Here we present the structure of the rearranged genes X1 and X2 determined by sequence and primer extension analysis, showing that the rearrangements involved partial duplication of normal gs 11. The transcription products of the normal and rearranged genes were also analysed.

METHODS

Cells and viruses. All the swine rotavirus strains used were grown in cultures of MA104 cells as previously described (Estes et al., 1979).

dsRNA extraction. Virus-infected cells were incubated at 37°C until complete cytopathic effect was observed at which time cells were disrupted by freezing and thawing, nuclei were removed by low speed centrifugation and
dsRNA was purified from the cell lysate by SDS-phenol extraction and concentrated by ethanol precipitation. For cDNA synthesis dsRNA was further purified by CF-11 cellulose chromatography (Bishop & Koch, 1969).

**Enzymes and reagents.** Restriction endonucleases were purchased from Bethesda Research Laboratories and New England Biolabs. *Escherichia coli* DNA polymerase I, T4 DNA ligase, RAV-2 reverse transcriptase and RNase H were from Amersham. Exonuclease III was from Boehringer Mannheim and *E. coli* poly(A) polymerase was from Bethesda Research Laboratories. [α-32P]dATP (3000 Ci/mmol) and [γ-32P]ATP (3000 Ci/mmol) were purchased from New England Nuclear.

**cDNA library construction and screening procedure.** A genomic cDNA library from porcine C60 rotavirus was constructed using pUC13 as cloning vector. In brief, total viral genomic dsRNA was 3'polyadenylated with *E. coli* poly(A) polymerase (Both et al., 1982; Cashdollor et al., 1982) and reverse transcribed by priming with oligo(dT) (Maniatis et al., 1982). The second cDNA strand was synthesized according to the manufacturer's recommendation. This primer was used to direct the synthesis of ss cDNA by reverse transcriptase on the viral mRNA template. The procedure was adapted from previously published methods (Agarwal et al., 1981; Virtanen & Pettersson, 1985). The primer (2 pmol), 5' end-labelled with [γ-32P]ATP and T4 polynucleotide kinase (Maniatis et al., 1982), was annealed to 5 μg total cytoplasmic RNA from virus-infected cells by heating at 90°C for 1 min in the presence of 0.25 mM-EDTA and allowing to cool to room temperature. The mixture was then adjusted to 100 mM-Tris-HCl pH 8.2, 10 mM-MgCl₂, 140 mM-KCl, 10 mM-dithiothreitol and 250 μM each of the four unlabelled deoxynucleoside triphosphates. cDNA synthesis was initiated by addition of 30 units of RAV-2 reverse transcriptase and performed at 42°C for 90 min. The elongation products were analysed by electrophoresis on a sequencing 6% polyacrylamide-7 M-urea gel together with size markers.

**ViralmRNA extraction and Northern blotting.** Confluent MA104 cell monolayers were infected with 5 to 10 p.f.u. of virus per cell. At 6 h post-infection total cytoplasmic RNA was extracted by the urea-LiCl method (Auffray & Rougeon, 1980), resolved under denaturing conditions on 1.6% agarose gels and transferred to nitrocellulose filters using the method of Thomas (1980).

An M13mp18 subclone was used to produce two different strand-specific probes for the rearranged genes. To detect X1 and X2 plus strands an M13 primer which anneals 5' of the polylinker cloning sites was used. Synthesis was carried out with [α-32P]ATP to make a complementary copy of the M13 vector DNA and thus leaving the inserted DNA single-stranded. To detect specifically the minus strands of the rearranged segments the same M13 subclone was annealed to the 17-mer M13 sequencing primer and extended in the presence of [γ-32P]ATP. After digestion with EcoRI the labelled ssDNA was eluted from a polyacrylamide gel (Maxam & Gilbert, 1980) and used as probe.

Hybridizations were performed at 65°C in 3x SSC, 5x Denhardt's solution, 0.1% SDS and 50 μg/ml sonicated salmon sperm DNA. Blots were washed in 0.1x SSC-0.1% SDS at 65°C.

**In vitro transcription of virus particles.** C17 virus was purified from infected MA104 cells by density equilibrium banding in CsCl (Smith et al., 1969) and used for in vitro synthesis of mRNA. Virus particles were activated by preincubation at 40°C for 30 min after the addition of 6 mM-EDTA. Subsequently, in vitro transcription was performed as described previously (Mason et al., 1980).

**RESULTS**

**Sequence analysis of gene X1**

Two clones (pX1.1 and pX1.2) containing inserts for the rearranged gene X1 were identified by colony hybridization of a cDNA library obtained from genomic dsRNA of strain C60. Each cDNA insert was subcloned in M13 DNA in both orientations and sequenced by the dideoxynucleotide chain termination technique. The complete nucleotide sequence of gs X1 was determined using pX1.1 (a full-length cDNA clone), pX1.2 (which is 100% homologous to pX1.1) and exonuclease III deletion clones derived from pX1.1 (see Methods).
Segment X1 consisted of 1211 nt (Fig. 1a) with a major open reading frame which extended from the first AUG at bases 22 to 24 to the UGA stop codon at bases 613 to 615, encoding a potential polypeptide of 197 amino acids (Fig. 1c). The gene also had a second (+1) open reading frame as well as the stop codons found in the 3' unit (in the same reading frame as the 5' unit). The arrowhead indicates the end of the 5' unit. Arrows indicate the beginning of the 3' unit (nt 625 to 630) and its homologous region in the 5' unit (nt 78 to 83). (b) Alignment of the overlapping sequences of the 5' and 3' units. The base changes in the 3' unit are shown below the sequence of the 5' unit. (c) Predicted translation product from the major open reading frame of gene X1 (nt 22 to 613). Arrowhead indicates the single potential N-glycosylation site at amino acid 20. Numbers indicate the amino acid position.

Fig. 1. (a) Nucleotide sequence of the cloned cDNA of C60 rotavirus gs X1. The sequence corresponds to the plus strand (mRNA sense) of the genomic dsRNA. The initiation and stop codons from the major open reading frame as well as the stop codons found in the 3' unit (in the same reading frame as the 5' unit) are underlined. The arrowhead indicates the end of the 5' unit. Arrows indicate the beginning of the 3' unit (nt 625 to 630) and its homologous region in the 5' unit (nt 78 to 83). (b) Alignment of the overlapping sequences of the 5' and 3' units. The base changes in the 3' unit are shown below the sequence of the 5' unit. (c) Predicted translation product from the major open reading frame of gene X1 (nt 22 to 613). Arrowhead indicates the single potential N-glycosylation site at amino acid 20. Numbers indicate the amino acid position.
Normal segment (human Wa strain)

Rearranged segment (porcine C60 strain)

Fig. 2. Schematic diagram of the structure of gsp X1 compared to the normal segment 11 of a standard strain. The duplicated unit of the C60 rearranged segment (bases 625 to 1211) corresponds to the region between bases 78 to 663 of normal segment 11 (indicated with dashed lines).

Table 1. Base changes found in the region shared by the 5' and the 3' units (nt 78 to 620 and nt 625 to 1167)

<table>
<thead>
<tr>
<th>Base position</th>
<th>Base change</th>
<th>Amino acid change*</th>
</tr>
</thead>
<tbody>
<tr>
<td>93 640</td>
<td>T A</td>
<td>S S</td>
</tr>
<tr>
<td>95 642</td>
<td>C T</td>
<td>T I</td>
</tr>
<tr>
<td>103 650</td>
<td>A G</td>
<td>T A</td>
</tr>
<tr>
<td>106 653</td>
<td>C T</td>
<td>L F</td>
</tr>
<tr>
<td>131 678</td>
<td>G A</td>
<td>S N</td>
</tr>
<tr>
<td>157 740</td>
<td>G A</td>
<td>E K</td>
</tr>
<tr>
<td>163 710</td>
<td>T C</td>
<td>F L</td>
</tr>
<tr>
<td>177 724</td>
<td>G A</td>
<td>M I</td>
</tr>
<tr>
<td>217 764</td>
<td>G A</td>
<td>A T</td>
</tr>
<tr>
<td>221 768</td>
<td>C T</td>
<td>S L</td>
</tr>
<tr>
<td>225 772</td>
<td>C T</td>
<td>N N</td>
</tr>
<tr>
<td>232 779</td>
<td>C T</td>
<td>L F</td>
</tr>
<tr>
<td>284 831</td>
<td>G A</td>
<td>G D</td>
</tr>
<tr>
<td>288 835</td>
<td>G A</td>
<td>V V</td>
</tr>
<tr>
<td>290 837</td>
<td>C T</td>
<td>S F</td>
</tr>
<tr>
<td>302 849</td>
<td>C T</td>
<td>S L</td>
</tr>
<tr>
<td>307 854</td>
<td>C T</td>
<td>Q Stop</td>
</tr>
<tr>
<td>324 871</td>
<td>T C</td>
<td>S S</td>
</tr>
<tr>
<td>339 886</td>
<td>T G</td>
<td>D E</td>
</tr>
<tr>
<td>377 924</td>
<td>A G</td>
<td>N S</td>
</tr>
<tr>
<td>408 955</td>
<td>A T</td>
<td>S S</td>
</tr>
<tr>
<td>420 967</td>
<td>G A</td>
<td>K K</td>
</tr>
<tr>
<td>435 982</td>
<td>G A</td>
<td>K K</td>
</tr>
<tr>
<td>438 985</td>
<td>G A</td>
<td>K K</td>
</tr>
<tr>
<td>441 988</td>
<td>T C</td>
<td>D D</td>
</tr>
<tr>
<td>459 1006</td>
<td>C T</td>
<td>Y Y</td>
</tr>
<tr>
<td>477 1024</td>
<td>T C</td>
<td>D Y</td>
</tr>
<tr>
<td>499 1046</td>
<td>T C</td>
<td>L L</td>
</tr>
<tr>
<td>513 1060</td>
<td>T C</td>
<td>D D</td>
</tr>
<tr>
<td>537 1084</td>
<td>T A</td>
<td>N K</td>
</tr>
<tr>
<td>563 1110</td>
<td>C T</td>
<td>A V</td>
</tr>
<tr>
<td>605 1152</td>
<td>A T</td>
<td>E V</td>
</tr>
<tr>
<td>614 1161</td>
<td>G A</td>
<td>Stop</td>
</tr>
</tbody>
</table>

* The amino acid changes were determined assuming the 3' unit has the same reading frame as the 5' unit.
Fig. 3. (a) Autoradiograph of primer extension products (a, b, c) using 5 μg total cytoplasmic RNA from MA104 cells, (lanes 1 to 3) infected with strains C60, C117 and OSU or (lane 4) mock-infected, analysed on a sequencing gel. Numbers are sizes in nucleotides. (b) Diagram of the extension products (a, b, c) obtained with the different viral mRNAs. Details of band assignments are given in the text.

At the nucleotide level, the coding 5' unit of X1 (positions 1 to 615) shared a high degree of homology with gs 11 from both UK (89%) and Wa (91%) strains. There is a significant difference between the coding regions of the UK and Wa genes: a single base deletion at nucleotide 389 in Wa puts its sequence out of frame with respect to UK for 25 bases; the reading frame is then recovered by a single insertion at nucleotide 418. The X1 5' unit maintained the UK gene reading frame leading to a total amino acid sequence homology of 89% with the Wa and 90% with the UK protein. However, the degree of amino acid homology with the Wa protein increased to 95% if the region between residues 123 and 132 (the divergent reading frame region) was not taken into account. Moreover, gene X1 (in both units) and Wa gs 11 shared the characteristic histidine (CAU codon) deletion at nt 442 to 444 with respect to the UK gene, and this resulted in both cases in a protein product one amino acid shorter than the UK VP11. In addition, the single putative N-glycosylation site at amino acid 20 found in the translated product of the known sequences of gs 11 was also present in the X1 protein product.

Analysis of gene X2 by primer extension

To determine whether the rearranged gene X2 of strain C117 arose as a consequence of a duplication mechanism similar to that for X1, primer extension experiments were performed with an oligonucleotide complementary to the X1 plus strand sequence between nt 445 and 461. This sequence is conserved in the two X1 units (but for one mismatch in the 3' unit) and in both bovine (UK) and human (Wa) genes. Total cytoplasmic RNA from virus-infected cells was used in the reverse transcriptase reactions. The cDNA extension products were analysed on a denaturing polyacrylamide gel as shown in Fig. 3(a). If gene X2 were also a dimer of gs 11 in a head-to-tail orientation two extension products would be obtained, each one initiating on the repeated sequence present in both units (Fig. 3b). As expected from the nucleotide sequence, X1 produced two extension products of 461 nt (band a) and 1008 nt (band c) in length. Similarly, X2 also yielded two cDNAs of approximate sizes 461 nt (band a) and 840 nt (band b). The shorter
product (band a), common to both X1 and X2 mRNAs as well as the normal gs 11 mRNA of the OSU porcine strain, originated by priming on the 5' unit (see Fig. 3b). The larger products, which initiated further downstream, were characteristic of the rearranged genes (band c in X1 and band b in X2). We therefore concluded that X2 was also composed of two segment 11 units in a head-to-tail orientation. Since the difference in length between extension products b and c (approximately 168 nt) was the same as that between gs X1 and X2 and since the VP11 coding region must be maintained in X2 as it is in X1 (Mattion et al., 1988), it could be inferred that the 3' unit of X2 was shorter than that of X1.

**Viral RNA transcripts**

In order to analyse the transcripts from the rearranged segments, total viral mRNA was purified from C60- and C117-infected MA104 cells and used in Northern blots. mRNA from OSU-infected cells was also included as a control.

Hybridizations were performed with an ssDNA probe complementary to the X1 plus strand. Fig. 4(a) shows that the sizes of the transcripts, 1.2 and 1.0 kb, were coincident with the lengths of the corresponding X1 and X2 genes. Moreover, *in vitro* transcription of C117 virus particles by the endogenous viral polymerase also yielded one mRNA product which corresponded in size to the X2 viral mRNA produced in infected cells (Fig. 4b, lane 4). No hybridization signal was obtained when an ssDNA probe complementary to the gene minus strand was used, ruling out the possibility of contamination of the RNA samples with genomic dsRNA (data not shown).

**DISCUSSION**

We studied the structure of the rearranged segments of two naturally occurring porcine rotavirus strains. We compared gs X1 with the normal gs 11 of the human strain Wa because they have a high degree of homology at both the nucleotide and amino acid levels. This let us show that the rearrangements involved the duplication of deleted versions of normal gs 11 but the open reading frame for its product VP11 was maintained. Furthermore the 5' and 3' termini characteristic of every rotavirus segment were conserved in the rearranged gene. It would have
been interesting to compare the rearranged segments with the original gs 11 from which they were generated, but this was not possible since the C60 and C117 strains were obtained from field isolates. Sequence analysis of cDNA clones corresponding to gene X1 revealed that it retained the gene 11 coding region. We have previously shown that the rearranged strains have a polypeptide with the same electrophoretic mobility as VP11 (Mattion et al., 1988). Rearrangements of segment 5 of bovine rotavirus (BRV strain) resulted in the protein product either being absent or appearing as a larger novel protein (Hundley et al., 1985).

Regardless of whether the rearrangements involve an intra- or an intermolecular mechanism it is most likely that, given the conservative mode of replication of the rotavirus genome, gs 11 rearranged during synthesis of the plus strand which in turn is followed by the synthesis of the minus strand.

In the case of ssRNA viruses, a copy choice mechanism involving the jumping of the viral polymerase from one template strand to another has been well documented (Bujarski & KAESBERG, 1986; Makino et al., 1986; Kirkegaard & Baltimore, 1986). According to our results, if a similar mechanism were operating for rotaviruses the viral polymerase would have to retain template specificity, thus avoiding the formation of mosaic gene structures.

We have also demonstrated that the viral mRNAs of the rearranged genes whether produced in virus-infected cells or in vitro with virus particles had the same length as their corresponding genome RNA segments. This is also consistent with the conservative mode of replication of rotaviruses.

The fact that the two rearranged strains originated spontaneously in two different places (Bellinzoni et al., 1987; Mattion et al., 1988) and the demonstration of a similar head-to-tail duplication of segment 11 suggest the existence of a common mechanism which could be relevant to the variability observed in the length of other rotavirus genomic segments (Estes et al., 1984; Hundley et al., 1987; Tanaka et al., 1988).

We would like to thank Dr Eduardo A. Scodeller and Dr José La Torre for helpful discussions and Alejandra Raimondi for synthesis and purification of the oligonucleotides. S.A.G. is specially grateful to Dr José L. Affranchino for his encouragement and suggestions. This work was partly supported by grants from the Programa Nacional de Biotecnología (PNB), Secretaría de Ciencia y Técnica (SECYT) and SAREC (Swedish Agency for Research Cooperation with Developing Countries). S.A.G. and N.M.M. are Fellows and O.R.B. is a Career Investigator of CONICET, and R.B. is a Fellow of the Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC).

REFERENCES


(Received 12 October 1988)