Comparison of group B rotavirus genes 9 and 11

Martin Petric, Kumudini Mayur, Steven Vonderfecht and Joseph J. Eiden*

Division of Infectious Disease, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, U.S.A.

Group B rotaviruses (GBRs) were recognized recently as causative agents of gastroenteritis. Investigations into the relatedness of various heterologous GBR strains have been hindered by the difficulty of growing these viruses in cell culture. Viral RNA extracted from experimentally infected rats was used to prepare cDNA clones. From these, the nucleotide sequences of genes 9 and 11 of the IDIR strain of GBR were determined and compared with the corresponding sequences of the human ADRV strain of GBR. IDIR gene 11 is 643 bp in length with a single open reading frame (ORF) encoding 174 amino acids; IDIR gene 9 is 804 bp in length with a single ORF encoding 246 amino acids. Comparison of the IDIR sequences with those of ADRV showed that nucleotide sequence similarity was 60.6% and 71-9% for genes 9 and 11, respectively. The deduced amino acid sequence similarity was 51-2% for the gene 9 and 66-5% for the gene 11 product. This sequence diversity indicates that GBRs are more distantly related than strains of group A rotavirus.

Rotaviruses are currently classified into antigenically distinct groups, A to F (Bridger, 1987). Group B rotaviruses (GBRs) have been detected in a number of mammalian species including humans, rats, lambs, calves and pigs (Hung et al., 1984; Vonderfecht et al., 1984, 1986; Chasey & Banks, 1984; Bridger, 1987; Theil et al., 1985; Snodgrass et al., 1984). These viruses are antigenically and genetically distinct from the group A rotaviruses (GARs) commonly associated with epidemics of infantile gastroenteritis in winter. Investigations of their molecular biology and epidemiology have been hindered by the inability to grow these viruses in cell culture. To date, cDNA clones have been synthesized from faecal isolates of the IDIR and ADRV strains of GBR (Sato et al., 1989; Chen et al., 1990a). Using these clones, GBR nucleotide sequences have been determined for gene 3 of the IDIR strain, and genes 9 and 11 of the ADRV strain. However, the nucleotide and amino acid sequences of equivalent genomic segments of heterologous GBR strains have not been compared. We have determined the nucleotide sequence of genomic segments 9 and 11 of IDIR and compared them to those of the equivalent ADRV gene segments, as well as GAR sequences.

The IDIR strain of GBR was propagated in infant CD-1 rats (Charles River Laboratories) as previously described (Vonderfecht et al., 1984). Viral genomic RNA was extracted from infected intestinal contents and purified by CF-11 cellulose chromatography (Theil et al., 1981). Synthesis of IDIR cDNA clones and their insertion into the plasmid pUC9 were accomplished by oligo(T) priming (Both et al., 1982; Gething et al., 1979). Northern blot analysis was used to identify cDNA clones of each IDIR genome segment (Eiden et al., 1989), the sequences of which were determined by dideoxynucleotide sequencing (Sanger et al., 1977) using synthetic oligodeoxynucleotide primers and direct primer extension from recombinant pUC clones. Additional sequences were analysed after subcloning into M13 vectors to enable complete sequence analysis from both cDNA strands. The 5'- and 3'-terminal sequences were confirmed by direct RNA sequencing (Bassel-Duby et al., 1986). Analyses of the IDIR nucleotide sequences and deduced primary protein sequences were performed with the GCG sequence analysis software package (Genetics Computer Group, Inc., Madison, Wis., U.S.A.), and comparison with the EMBL and GenBank databases.

The sequence of IDIR gene 11 (available under D00912:IDIR11) was determined by sequence analysis from two nearly full-length cDNA clones. The length of gene 11 was 643 bp with one major open reading frame (ORF) encoding a translation product of 174 amino acids, or 20·1K if unmodified (Fig. 1a). There was a 58 bp untranslated sequence at the 5' end, with the first AUG codon at base 59, in the optimal sequence context for translation initiation [(A/G)NNAUGG; Kozak, 1981]. A second in-frame AUG at base 158 was also within an optimal initiation context. No significant translation products were deduced to be encoded by other reading frames, as has been shown for GAR gene...
11 (Mitchell & Both, 1988). Within the deduced amino acid sequence were three polypeptide sequences encoding potential glycosylation sites conforming to the sequence Asn-X-(Ser/Thr) (Spiegelberg et al., 1970). The deduced amino acid sequence contained a single cysteine and three proline residues. The first termination codon was UAG which was followed by a 63 bp untranslated region at the 3' terminus. No signal sequence for poly(A) polymerase was evident at the 3' end (Proudfoot & Brownlee, 1976).

The sequence of IDIR gene 9 (accession number D00911:IDIR9) was also determined from two nearly full-length cDNA clones of the dsRNA genome segment. The sequence was composed of 804 bp with a single, long ORF encoding a polypeptide of 246 amino acids of 28.5K (Fig. 1b). The AUG codons at bases 9 and 28 were both within sequences optimal for initiation of translation (Kozak, 1981). There were two potential glycosylation sites, six cysteines and eight prolines. The 3' terminus contained 58 untranslated bases and there was no apparent polyadenylation signal.

**Comparison of the nucleotide sequence of IDIR gene segment 11 with any of the published nucleotide sequences of GARs (Estes & Cohen, 1989) revealed little or no similarity. However, the similarity between the nucleotide sequences of the gene 11 segment of IDIR and that of ADRV (Chen et al., 1990a) is 71.9%, and the terminal sequences of these two strains are even more conserved. The untranslated regions at the 5' and 3' termini of IDIR and ADRV are identical in length, and the first 38 bases at the 5' terminus of the positive-strand RNA of IDIR gene 11 are identical to those of ADRV. Similarly, 36 of the 43 bases at the 3' end are conserved. Overall, there is extensive conservation of the bases at either end of the gene segment. These highly conserved regions are consistent with the report that primers complementary to the terminal sequence of ADRV gene 11 can recognize IDIR gene 11 in a polymerase chain reaction (Eiden et al., 1991).

The deduced amino acid sequences of IDIR and ADRV gene 11 also show substantial similarity, 66.5% overall. IDIR gene 11 contains 12 more nucleotides than ADRV. These bases are within the reading frame, yielding an IDIR polypeptide which contains four more amino acids than the comparable ADRV polypeptide. Both ADRV and IDIR sequences are rich in serine, with 21 serine residues in the IDIR and 20 in the ADRV sequence; its position is highly conserved in 15 of 20 sites. This similarity is particularly evident in hydrophobicity plots of the ADRV and IDIR gene 11 polypeptides, reflecting the conserved nature of the amino acid substitutions between the two strains (Fig. 2a). The role of the gene 11 product has not been established, although the GAR gene 11 product (NS26) has been shown to be a non-structural phosphoprotein (Welch et al., 1989).

**Fig. 1.** The deduced amino acid sequences of IDIR genes 11 (a) and 9 (b), and comparison with ADRV genes 11 and 9. Colons indicate identical residues. Potential N-linked glycosylation sites are at amino acids 58, 148 and 156 for the IDIR gene 11 polypeptide (underlined), and at amino acids 43 and 129 for the IDIR gene 9 product (underlined). Asterisks denote identically positioned cysteine residues.
Fig. 2. Comparative hydrophilicity/hydrophobicity plots of the deduced amino acid sequences of GBR genes 11 (a) and 9 (b). The plots were generated by the method of Kyte & Doolittle (1982) using a computer analysis program (Devereux et al., 1984).

conserved, being at identical equivalent positions, which implies that the polypeptide product of gene 9 may be a molecule with a high level of cross-linkage. The ADRV gene 9 product has three potential glycosylation sites, at amino acids 45, 91 and 105, whereas that of IDIR has only two sites, at amino acids 43 and 129.

It has been reported recently that the ADRV gene 9 protein sequence is 28% identical to the VP7 consensus sequence of the major structural glycoprotein of GARs (Chen et al., 1990b). If substitutions of similar amino acids are included, this similarity increases to 78%, despite the disparity in size (ADRV, 249 amino acids; GAR VP7, 326 amino acids). When the IDIR gene 9 protein sequence was aligned with the VP7 sequence in the same manner as described for ADRV, only 18% of the amino acids were identical. However, four of the six cysteines of the IDIR gene 9 product have the same positions in the VP7 sequence. If GBR gene 9 encodes a protein which is functionally and structurally equivalent to GAR VP7, IDIR may be more distantly related than ADRV to GARs. Future efforts should be aimed at identifying the gene encoding the GBR major outer capsid protein.

Although both the nucleotide and amino acid sequences of genes 9 and 11 of ADRV and IDIR are similar, neither pair shares the same degree of sequence conservation that is observed among heterologous strains of GARs. For example, a review of the EMBL and GenBank databases showed that the nucleotide and amino acid sequences of gene 11 of the Wa, OSU and UK strains of GAR are ≥85.9% and ≥82.4% conserved. In contrast, only 66.5% of the amino acids and 71.9% of the nucleotide sequence of ADRV and IDIR gene 11 are identical. Gene 9 of the two GBR strains is 60.6% and 51.2% identical at the nucleotide and amino acid sequence levels. This degree of similarity is substantially less than the 85% reported for heterologous VP7 genes obtained from different serotypes of GAR (Kapikian & Chanock, 1990). However, the differences noted between the ADRV and IDIR genes are largely due to substitutions of amino acids which conserve charge and/or polarity, as illustrated in Fig. 1(b) and 2(b).

The differences between the genes of IDIR and ADRV indicate that strains of GBR are more distantly related than different strains of GAR. This diversity should be confirmed by sequence analysis of additional GBR genomic segments from IDIR and ADRV, as well as other GBR strains. Since each of the rotavirus groups possesses very similar morphology, as determined by electron microscopic examination, the differences within the individual rotavirus groups as well as characteristics shared among the rotavirus groups should aid in our understanding of the mechanisms of pathogenesis and replication, and of structure.

This work was supported in part by funds provided by the National Institutes of Health grant no. 1 R29 AI24922-01A1. M.P. is on sabbatical leave from The Hospital for Sick Children, Toronto, Canada.

References


(Received 28 March 1991; Accepted 30 July 1991)