JOURNAL OF VIROLOGY, Sept. 1984, p. 860–862
0022-538X/84/090860-03$02.00/0

Vol. 51, No. 3

Nucleotide Sequence of the Gene Encoding the Serotype-Specific Antigen of Human (Wa) Rotavirus: Comparison with the Homologous Genes from Simian SA11 and UK Bovine Rotaviruses

MICHAEL A. RICHARDSON,[1] AIKICHI IWAMOTO,[2] NOBUKO IKEGAMI,[3] AKIO NOMOTO,[2] AND YASUHIRO FURUICHI[1]*

Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110[1]; University of Tokyo, Medical School, Tokyo, Japan[2]; and Osaka National Hospital, Clinical Research Institute, Osaka, Japan[3]

Received 29 March 1984/Accepted 22 May 1984

The nucleotide sequence of human (Wa) rotavirus genome segment 9, which encodes the serotype-specific antigen VP7, has been determined. Comparison of the deduced amino acid sequence of Wa VP7 protein to the sequences of simian SA11 and UK bovine VP7 proteins shows that the majority of the amino acid differences are clustered between amino acid residues 37 through 49, 65 through 75, 87 through 105, 122 through 126, 146 through 149, 178 through 181, and 208 through 242. A hydrophilicity profile of the three proteins reveals correlations between hydrophilic peaks, potentially antigenic determinants, and certain clusters of amino acid changes.

Rotaviruses are an important cause of acute gastroenteritis in the young of humans and other mammals (7, 8) and possess a genome of 11 distinct double-stranded RNA segments enclosed within a double-layered protein capsid (reviewed in references 7 and 8). In infected cells the double-stranded RNA segments are transcribed into mRNAs by a virion-associated RNA-dependent RNA polymerase (3). These mRNAs also act as templates for a putative viral replicase. The major component of the outer shell is the glycoprotein VP7, encoded by either genome segment 8 (UK bovine rotavirus; 15) or 9 (human Wa and simian SA11 rotaviruses; 2, 12). VP7 elicits neutralizing antibodies that have been used to distinguish at least four human, two bovine, and three avian rotavirus serotypes (7). Wa rotavirus is representative of human serotype 1 (18).

To investigate the basis for antigenic variation between rotavirus serotypes and to help identify important antigenic determinants, we have sequenced Wa rotavirus genome segment 9 and compared it and its deduced amino acid sequence with the cognate sequences of simian SA11 rotavirus (2; equivalent to human serotype 3 [17]) and UK bovine rotavirus (5).

Cloned cDNA copies of the Wa double-stranded RNA genome were obtained as previously described (11), and the nucleotide sequence of genome segment 9 was determined by the chemical sequencing method of Maxam and Gilbert (14).

Segment 9 is 1062 nucleotides long with two in-phase AUG triplets at residues 49 through 51 and 136 through 138 and a single terminator at residues 1,027 through 1,029 (Fig. 1). Segment 9 thus encodes a protein of either 297 or 326 amino acids, depending on which of these AUGs initiates protein synthesis. The Wa 9 sequence, like all rotavirus segments sequenced to date, is A+T rich (67%).

Nucleotide sequence homology among the three serotypes is approximately 75%, with the two potential initiators of protein synthesis and the single terminator located at identical positions in all three sequences (Fig. 1). Like SA11 and UK bovine mRNAs, the first potential initiation codon (residues 49 through 51) in Wa segment 9 mRNA has the sequence TXXAUGT, regarded as a weak initiator of protein synthesis in eucaryotic systems (13). The second AUG-coding triplet (residues 136 through 138) in all three viruses has the consensus sequence AXXAUGG, characteristic of a strong initiator. It is not yet known whether protein synthesis initiation occurs at one or both AUG codons. Although the flanking nucleotides at the second AUG and features of the second putative signal peptide (see below) suggest this AUG as the initiator of protein synthesis, conflicting results have been obtained by sodium dodecyl sulfate-polyacrylamide gel analysis of viral polypeptides synthesized either in vitro or in the presence of tunicamycin (6, 15). Thus, comparative studies on the migration of the non-glycosylated precursor to VP7 and the 317 amino acid long products (as deduced from nucleotide sequence data; 1, 4) of genome segments 7 (UK) and 8 (SA11), implicated both the first (for SA11) and the second (for UK) AUGs as potential initiators. The possible effect of hydrophobicity on the migration of VP7 could be a further complicating factor (10). Protein sequence data on the in vivo-synthesized precursor to VP7 will be required to resolve this issue.

Amino acid sequence homology among the type-specific antigens of the three serotypes is 82% (Fig. 2). Most of the changes are conservative, involving the exchange of amino acids with similar physiochemical properties. All three viruses have a potential glycosylation site (Asn-X-Thr) at amino acid residues 69 through 71. In addition, there is a second site at residues 238 through 240 for both Wa and UK bovine rotaviruses, and a third (Asn-X-Ser) is present only in UK bovine rotavirus at residues 318 through 320. There are two distinct hydrophobic regions near the N termini of all three proteins (Fig. 3), one following each of the two possible AUG initiators. Both are theoretically capable of acting as signal peptides, with the second (between residues 30 and 50) perhaps more exactly fulfilling the consensus requirements of a signal peptide (16), i.e., amino acid content, length, appropriately positioned arginine residues, and greater β turn potential at the possible cleavage site.

* Corresponding author.

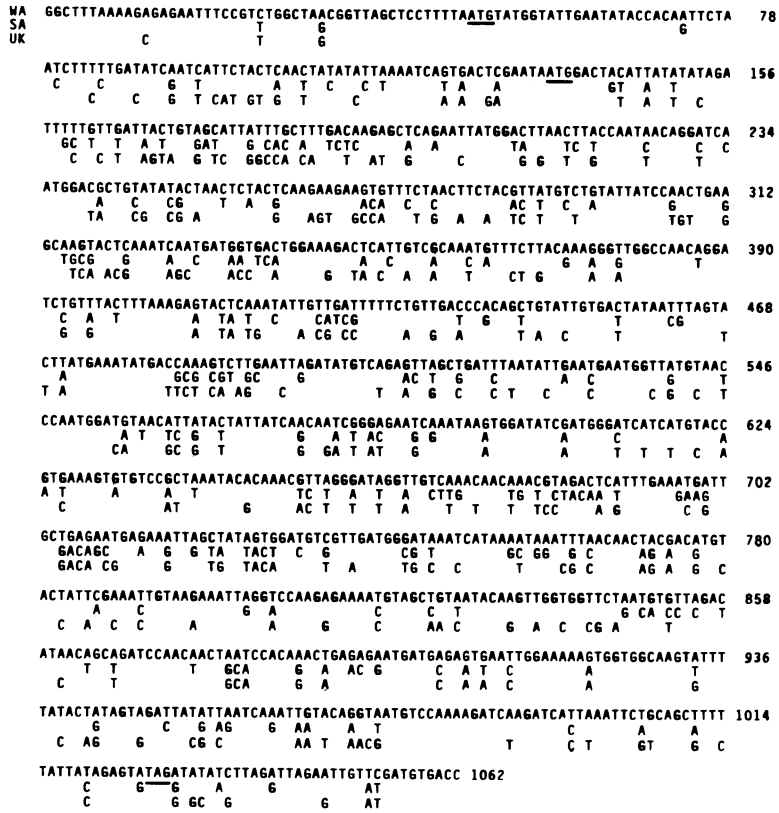

FIG. 1. Nucleotide sequence of segment 9 from Wa rotavirus. A cloned cDNA copy of gene segment 9 was used for sequence determination. Both strands were sequenced. The DNA Sequence of the sense strand is shown 5′ to 3′. Underlined bases indicate the positions of the two possible initiators and the single terminator. For comparison, the base changes found in the homologous segment from SA11 and UK bovine rotaviruses are shown.

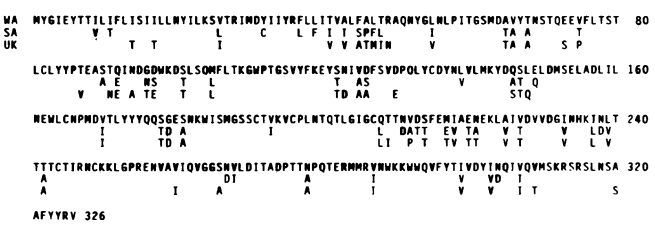

FIG. 2. Deduced amino acid sequence of Wa serotype-specific protein VP7. The amino acid substitutions found in the corresponding proteins from SA11 and UK bovine rotaviruses are indicated for comparative purposes.

A distinct clustering of amino acid substitutions at specific regions of the VP7 proteins is evident, in particular between residues 37 and 49, 65 and 75, 87 and 105, 122 and 126, 146 and 149, 178 and 181, and 208 and 242 (Fig. 2). Comparison of these variable regions with a Hopp and Woods (9) computer-generated hydrophilicity plot on all three proteins (Fig. 3) revealed that regions 65 through 75, 87 through 105, 146 through 149, 178 through 181, and 211 through 227 (indicated by bars in Fig. 3) were located at or adjacent to hydrophilic peaks. It is of interest that the largest peak (310 through 317), the most likely to be associated with an antigenic determinant (9), has an identical amino acid se-

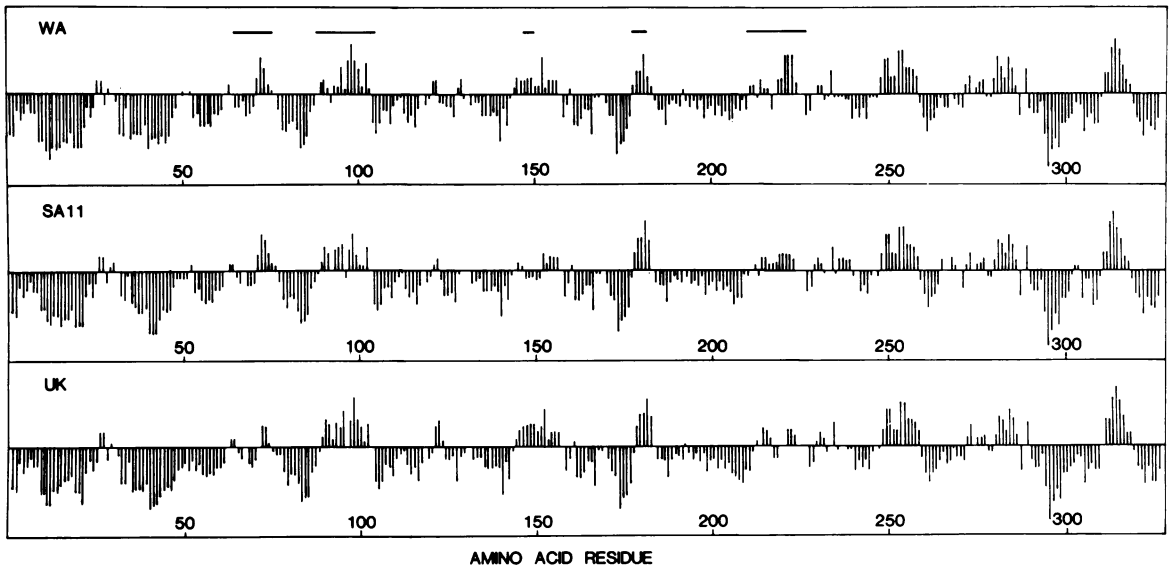

FIG. 3. Hydrophilicity plots on the serotype-specific proteins from Wa, SA11, and UK bovine rotaviruses. The plots were generated by computer by using the parameters of Hopp and Woods (9). The bars denote regions of interest mentioned in the text.

quence for all three proteins, which suggests that this region by itself could not be the basis for anitgenic differences between rotavirus serotypes.

The data obtained in this comparative study may help identify significant antigenic determinants on the serotype-specific antigen of rotaviruses. The antigenicity of synthetic peptides corresponding to those variable regions associated with hydrophilic peaks is now under investigation.

LITERATURE CITED

1. **Both, G. W., A. R. Bellamy, J. E. Street, and L. J. Siegman.** 1982. A general strategy for cloning double-stranded RNA: nucleotide sequence of the simian-11 rotavirus gene 8. Nucleic Acids Res. **10**:7075–7088.
2. **Both, G. W., J. S. Mattick, and A. R. Bellamy.** 1983. Serotype-specific glycoprotein of simian 11 rotavirus: coding assignment and gene sequence. Proc. Natl. Acad. Sci. U.S.A. **80**:3091–3095.
3. **Cohen, J.** 1977. Ribonucleic acid polymerase activity associated with purified calf rotavirus. J. Gen. Virol. **36**:395–402.
4. **Dyall-Smith, M. L., T. C. Elleman, P. A. Hoyne, I. H. Holmes, and A. A. Azad.** 1983. Cloning and sequence of UK bovine rotavirus gene segment 7: marked sequence homology with simian rotavirus gene segment 8. Nucleic Acids Res. **11**:3351–3374.
5. **Elleman, T. C., P. A. Hoyne, M. L. Dyall-Smith, I. H. Holmes, and A. A. Azad.** 1983. Nucleotide sequence of the gene encoding the serotype-specific glycoprotein of UK bovine rotavirus. Nucleic Acids Res. **11**:4689–4701.
6. **Ericson, B. L., D. Y. Graham, B. B. Mason, and M. K. Estes.** 1982. Identification, synthesis, and modifications of simian rotavirus SA11 polypeptides in infected cells. J. Virol. **42**:825–839.
7. **Estes, M. K., E. L. Palmer, and J. Obijeski.** 1983. Rotaviruses: a review. Curr. Top. Microbiol. Immunol. **105**:124–184.
8. **Holmes, I. H.** 1983. Rotaviruses, p. 359–423. *In* W. K. Joklik (ed.), The reoviruses. Plenum Publishing Corp., New York.
9. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
10. **Hudspeth, M. E. S., W. M. Ainley, D. S. Shumard, R. A. Butow, and L. T. Grossman.** 1982. Location and structure of the var 1 gene on yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. Cell **30**:617–626.
11. **Imai, M., M. A. Richardson, N. Ikegami, A. J. Shatkin, and Y. Furuichi.** 1983. Molecular cloning of double-stranded RNA virus genomes. Proc. Natl. Acad. Sci. U.S.A. **80**:373–377.
12. **Kalica, A. R., H. B. Greenberg, R. G. Wyatt, J. Flores, M. M. Sereno, A. Z. Kapikian, and R. M. Chanock.** 1981. Genes of human (strain Wa) and bovine (strain UK) rotaviruses that code for neutralization and subgroup antigens. Virology **112**:385–390.
13. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5252.
14. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
15. **McCrae, M. A., and J. G. McCorquodale.** 1982. The molecular biology of rotaviruses. II. Identification of the protein-coding assignments of calf rotavirus genome RNA species. Virology **117**:435–443.
16. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
17. **Wyatt, R. G., H. B. Greenberg, W. D. James, A. L. Pittman, A. R. Kalica, J. Flores, R. M. Chanock, and A. Z. Kapikian.** 1982. Definition of human rotavirus serotypes by plaque reduction assay. Infect. Immun. **37**:110–115.
18. **Wyatt, R. G., H. D. James, A. L. Pittman, Y. Hoshino, H. B. Greenberg, A. R. Kalica, J. Flores, and A. Z. Kapikian.** 1983. Direct isolation in cell culture of human rotaviruses and their characterization into four serotypes. J. Clin. Microbiol. **18**:310–317.