JOURNAL OF VIROLOGY, Sept. 1984, p. 856–859
0022-538X/84/090856-04$02.00/0

Vol. 51, No. 3

NOTES

Nucleotide Sequence of Human Rotavirus Genome Segment 10, an RNA Encoding a Glycosylated Virus Protein

Y. OKADA,[1] M. A. RICHARDSON,[2] N. IKEGAMI,[3] A. NOMOTO,[1] AND Y. FURUICHI[2]*

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

Received 8 March 1984/Accepted 16 May 1984

The complete nucleotide sequence of human rotavirus (Wa strain) genome segment 10 was determined by using a cloned DNA copy. The sequence data indicated that segment 10 is A+T rich (65%) and consists of 750 base pairs. The positive strand of segment 10 contains a single open reading frame that extends 175 codons from the first AUG triplet (residues 42 through 44). The amino acid sequence of the segment 10 product was deduced from the nucleotide sequence. There are two distinct glycosylation sites at the N-terminal hydrophobic region, consistent with previous findings that this protein exists in a glycosylated form. The apparent molecular weight (20,000) of the unglycosylated, precursor polypeptide is in good agreement with the one calculated from the predicted amino acid sequence. Structural analysis of the positive strand (mRNA from segment 10) showed that it could form, like mRNA from segment 11, a stable panhandle structure involving the 5′ and 3′-terminal regions. The nucleotide sequence of segment 10 from simian rotavirus, recently determined by Both et al. (J. Virol. 48:335–339, 1983) was found to be highly homologous to, and to share several important features with, segment 10 of human rotavirus.

Rotaviruses, members of the Reoviridae family, are an important cause of infantile diarrhea in many mammalian species, including humans (8, 9). The rotavirus genome consists of 11 monocistronic double-stranded RNA (dsRNA) segments that are enclosed within a double-layered protein capsid. The segments are transcribed by virus-associated RNA polymerase to form capped mRNAs (11) that also function as templates for a putative replicase in virus-infected cells. The coding assignments for most segments have been established for simian rotavirus (SA11) (1, 6, 14, 16, 23) and bovine rotavirus (18) by translating mRNAs or denatured dsRNAs in in vitro protein-synthesizing systems.

Analysis of rotavirus-specific polypeptides in infected cells in the presence and absence of tunicamycin, a potent inhibitor of N-linked glycosylation, indicated that there were two glycosylated proteins, the structural protein VP7 (38,000 daltons [38K]) and the putative nonstructural protein NCVP5 (28K) (7, 8). VP7, the product of genome segment 9, constitutes the major outer capsid protein and elicits type-specific and neutralizing antibodies, whereas NCVP5, the product of genome segment 10, has not been found in the particles of either purified simian or human rotaviruses. Their non-glycosylated precursors formed in the presence of tunicamycin were found to have molecular masses of 35.5K and 20K, respectively (1, 7).

The function of NCVP5 in infected cells is not yet known, although it was suggested for SA11 virus that this protein might be involved in virus maturation (21). Recently, Both et al. (4) determined the nucleotide sequence of simian genome segment 10 and deduced the amino acid sequence of the precursor polypeptide.

Nucleotide sequence of rotavirus genome segment 10. Human rotavirus genome segments were cloned into pBR322 essentially as described previously (13). Briefly, cDNA to rotavirus genome RNAs were prepared as described previously (13), except that viral dsRNAs were 3′ tailed with polyadenylic acid by using *Escherichia coli* polymerase (3, 5) rather than oligocytidylic acid-tailed by RNA ligase (13). Denaturation of the tailed dsRNA to facilitate cDNA synthesis by reverse transcription was accomplished by incubation with methylmercury hydroxide (20). Cloning of cDNAs into pBR322 and screening for *E. coli* transformants that harbored the recombinant plasmids were carried out as described previously (13).

The recombinant plasmid DNA (pBR322/R1073) that contained the sequence of rotavirus genome segment 10 was purified by CsCl buoyant density gradient centrifugation and Sepharose CL4B column chromatography. Dot-blot hybridization of ^{32}P-labeled, nick-translated R1073 DNA to the individual rotavirus genome segments was performed as described previously (13), and it confirmed that the cloned cDNA was specific for genome segment 10 (data not shown).

Initially, the *Pst*I-excised cDNA insert was ^{32}P-labeled by incubation with [γ-^{32}P]ATP and polynucleotide kinase, and the nucleotide sequences of the 5′-terminal region was determined. The cDNA insert had terminal sequences of:

$$^{5'}(G)_{\sim 30}\text{-}(T)_{26}\text{-GGCTTTTAAA--//--TGTGACC-}(A)_{16}\text{-}(C)_{\sim 30}\,^{3'}$$

$$^{3'}(C)_{\sim 30}\text{-}(A)_{26}\text{-CCGAAAATTT--//--ACACTGG-}(T)_{16}\text{-}(G)_{\sim 30}\,^{5'}$$

in which 5′ GGCUUU and 5′ GGACTCA were the diagnostic 5′-terminal sequences common to all rotavirus genome dsRNA segments (11, 19). These results indicate that R1073 DNA contained a full-length copy of rotavirus gene 10. The complete nucleotide sequence was then determined by using the Maxam and Gilbert method (17).

Genome segment 10 is A+U rich (65%) and consists of 750 base pairs (Fig. 1). It possesses one long open reading frame

* Corresponding author.

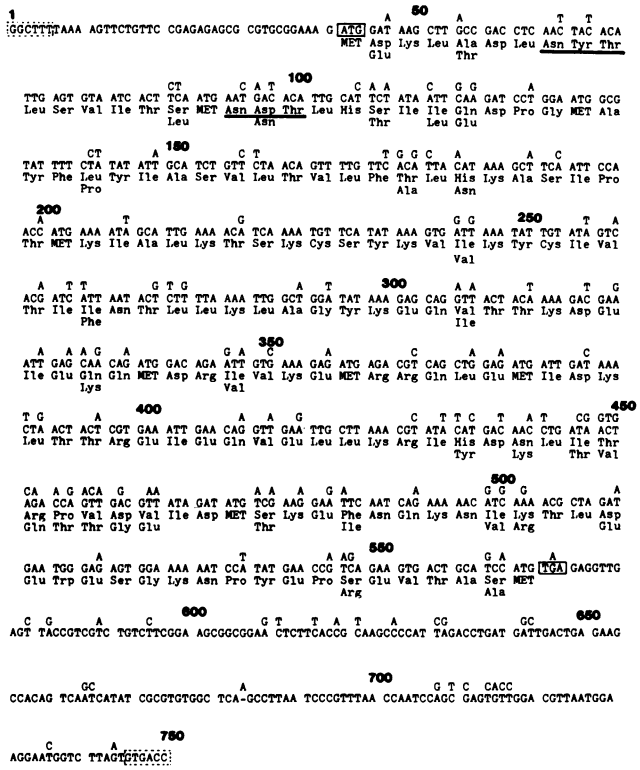

FIG. 1. Nucleotide sequence of the coding stand of human rotavirus (Wa) gene 10 and the predicted amino acid sequence of the NCVP5 protein. For comparison, the nucleotide and amino acid differences that occur in simian rotavirus are also shown. The boxes drawn by broken lines represent the terminal consensus sequences observed for genome segments 7 through 11 (3, 11), and the boxes with solid lines show the putative protein synthesis initiation and termination codons. The underlined regions indicate the glycosylation sites. For convenience, T (in DNA) is used in place of uridine in RNA. The A residue shown between residues 683 and 684 has been reported in simian rotavirus genome segment 10 by Both et al. (4). This base is absent in human rotavirus genome segment 10.

that starts with the first AUG (at residues 42 through 44) and encodes a protein of 175 amino acids. The other two reading frames contain multiple termination codons, with the longest reading frame extending only eight codons (residues 516 through 539). The 5'-proximal AUG appears to be an efficient initiator since it is surrounded by a purine at position −3 and a guanine at position +4. The second AUG, which is in phase with the first (positions 90 through 92), is a relatively weak initiator according to the Kozak rule (15), since it contains C and A in the −3 and +4 positions, respectively. A single terminator occurs at positions 567 through 569. Assuming that the first AUG is the initiator of protein synthesis, the segment 10 product was estimated to have a molecular mass of 20,240, consistent with a reported molecular mass of 20K for the unglycosylated precursor (pNCVP5) produced in tunicamycin-treated rotavirus-infected cells (7). Two potential glycosylation sites, Asn-Tyr-Thr (residues 63 through 71) and Asn-Asp-Thr (residues 93 through 101), were present near the N terminus of the predicted amino acid sequence. There was a long 3'-noncoding region of 185 bases, which may account for the observation that the primary polypeptide coded for by gene 10 mRNA comigrates in sodium dodecyl sulfate (SDS)-polyacrylamide gels with the protein encoded by the shorter gene 11 mRNA.

Recently, Both et al. completed the nucleotide sequence of cloned genome segment 10 of simian rotavirus SA11 (4). When this sequence was compared with that of human rotavirus genome segment 10, many similarities were noted. The two genes were virtually identical in size (750 base pairs for the human rotavirus and 751 base pairs for the simian rotavirus) with an overall sequence homology of 82%. The 5' noncoding sequences were totally conserved, as were the positions of the predicted protein synthesis initiation and termination codons and the two potential glycosylation sites.

There were 137 base changes between the human and simian sequences, with purine-to-purine and pyrimidine-to-pyrimidine changes strongly favored (i.e., 63%). Of the purine to pyrimidine base changes there were an exceptionally high rate of A-U changes. These preferential A-U substitutions may, to some extent, account for the maintenance of the high A+U base composition of segment 10.

As expected from the nucleotide sequence homology, the predicted amino acid sequences of the two viral polypeptides were significantly conserved. There were only 30 changes out of a total of 175 amino acids. Hence, whereas there were 108 base substitutions in the coding region, most of these base changes occurred at the third position of codons. About 50% of the total amino acid changes involved the substitution of amino acids with similar physicochemical properties, e.g., Asp-Glu in positions 45 and 511; Lys-Arg in position 502; Ser-Thr in positions 108 and 475; and Ile-Leu and Ile-Val in positions 14 and 242, respectively. There was, however, a notable region (nucleic acid residues 433 through 465, Fig. 1) near the C terminus in which very little, if any, homology was found. About one-third of the total amino acid changes were concentrated in this highly mutated region.

A Hopp and Wood (10) hydropathicity plot indicated that the human rotavirus segment 10 gene product (pNCVP5), like the simian segment 10 gene product, consisted of two major domains, namely, a hydrophobic N-terminal half and a hydrophilic C-terminal half. The two potential glycosylation sites were located near the N terminus, the first in a marked hydrophobic region and the second at a slightly hydrophilic region. A striking conservation of charged amino acids between the human and simian pNCVP5s was also noted. Apart from four amino acids, one in the N terminus (amino acid residue 19) and three in the highly mutated region (i.e., nucleic acid residues 433 through 465), the remaining 39 charged amino acids were identical with respect to their location and charge.

Structural features of gene 10 mRNA. Previously, we detemined the sequence of human rotavirus genome segment 11 and found that the positive strand (mRNA) has the potential to form a stable "panhandle" structure involving its 5'- and 3'-terminal regions (12). To know whether the positive strand of segment 10 shared such a feature, its secondary structure was investigated (Fig. 2). In Fig. 2, the locations of base changes between the human and simian sequences are also indicated. There were three regions potentially capable of stable intramolecular base pairing ($\Delta G \leq -40.74$ kJ/mol). The region with the greatest stability (residues 82 through 136 and 748 through 707; $\Delta G = -84$ kJ/mol) was formed by base pairing between the 3'-terminal and the 5'-proximal coding nucleotides, resulting in a panhandle structure similar to that predicted for segment 11 mRNA (12).

Simian rotavirus mRNA 10 is also capable of forming such a panhandle structure, but with a shorter panhandle stem region. Panhandles of this type may be a common structural feature of rotavirus mRNAs. We have recently analyzed the

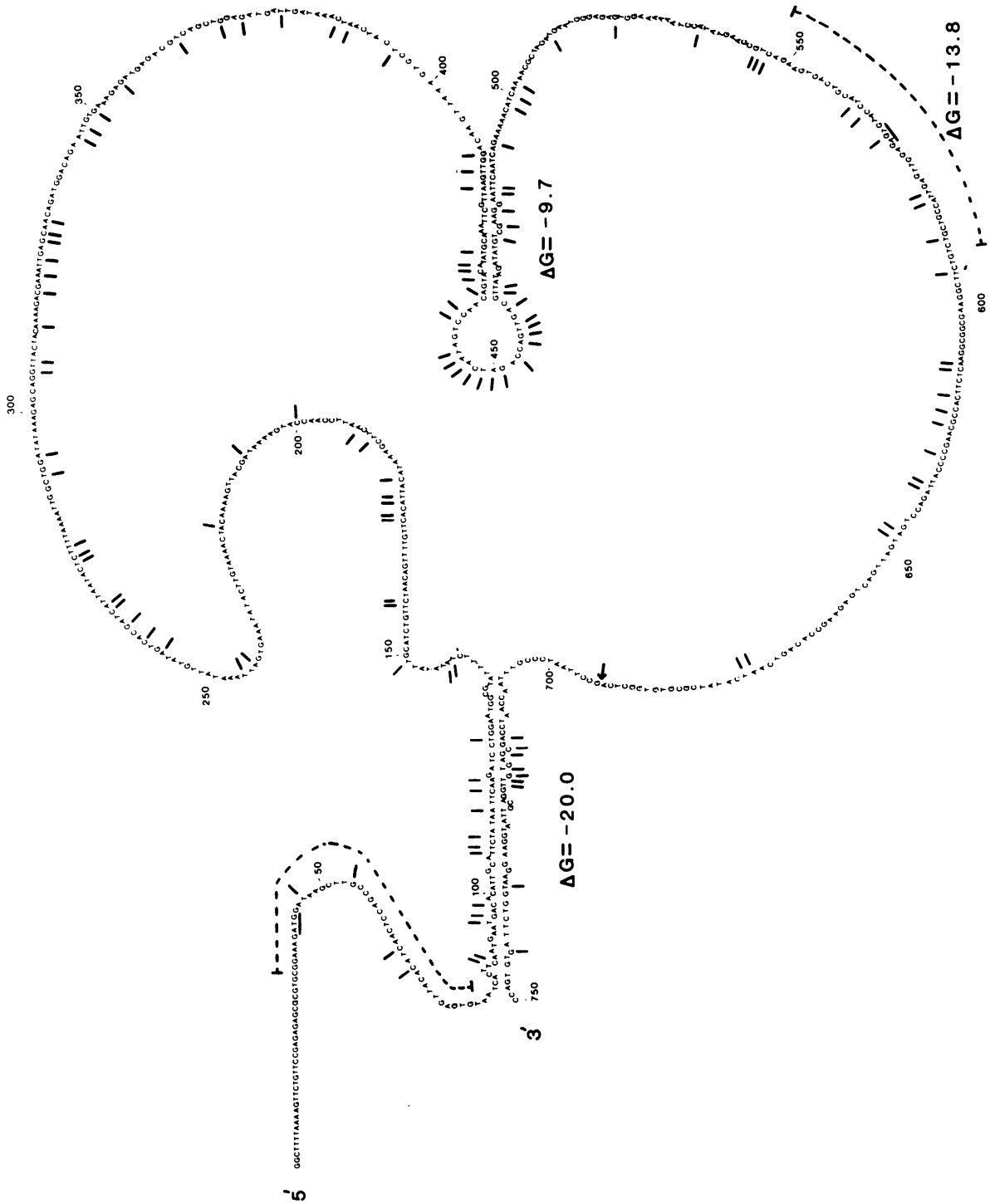

FIG. 2. Potential panhandle structure formed by human rotavirus (Wa) genome segment 10. The lines outside of the sequence indicate those bases that are different in simian rotavirus genome segment 10. The broken lines indicate the possible base paired region which, for simplicity, is not shown on this figure. The underlined region indicates the potential protein synthesis initiation site. Free energy estimations were determined by a computer program (Intelligenetics, Palo Alto, Calif.) based on the parameters of Borer et al. (2).

sequences of rotavirus mRNAs 6 through 11 and found that all have the potential to form such structures. In each case, the 5'- and 3'-terminal regions of the mRNA were capable of base pairing to form a stable stem structure (as measured by their free energies).

Cashdollar et al. and we have determined the nucleotide sequences of genome segments S2 (5) and S3 (22) of human reovirus, another dsRNA virus. Analyses of these reovirus RNAs indicated that reovirus S2 and S3 mRNAs apparently could not form such panhandle structures (22). Therefore, the potential for panhandle configurations is not a general phenomenon of mRNAs from dsRNA-containing viruses. Rather, such structures, like the high A+U base content, may be a characteristic feature of rotavirus mRNAs.

A second potential region of base pairing (residues 35 through 78 and 590 through 547, $\Delta G = -57.96$ kJ/mol) incorporated both the protein synthesis initiation and termination sites. The third possible intragenic base-paired region (residues 410 through 437 and 489 through 462, $\Delta G = -40.74$ kJ/mol) created a single-stranded loop that showed, surprisingly, significant sequence variation between the human and simian rotavirus mRNAs. This region corresponded to that part of the human gene 10 protein showing significant differences in amino acids from the simian gene 10 product (Fig. 1). Of the bases involved in this region, 60% were altered. This represents a mutation rate sixfold higher than for the rest of the molecule. However, apart from this one specific region, the protein encoded by genome segment 10 of human and simian rotaviruses appeared to have been significantly conserved during evolution, suggesting an important role in viral replication. Whether this highly mutated region is the result of evolutionary selection for functional proteins after random mutations of the viral genome or whether the corresponding gene locus is particularly susceptible to base changes during replication remains to be determined.

We thank Kaoru Akatani and Mitsunobu Imai for helpful discussions and expert technical advice throughout this work.

LITERATURE CITED

1. **Arias, C. L., S. Lopez, and R. T. Espejo.** 1982. Gene protein products of the SA11 simian rotavirus genome. J. Virol. **41**:42–50.
2. **Borer, P. N., B. Dengler, and I. Tinoco, Jr.** 1974. Stability of ribonucleic double-stranded helices. J. Mol. Biol. **86**:843–853.
3. **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. **11**:7075–7088.
4. **Both, G. W., L. J. Siegman, A. R. Bellamy, and P. H. Atkinson.** 1983. Coding assignment and nucleotide sequence of simian rotavirus SA11 gene segment 10: location of glycosylation sites suggests that the signal peptide is not cleaved. J. Virol. **48**:335–339.
5. **Cashdollar, L. M., J. Espoza, G. R. Hudson, R. Chmelo, P. N. K. Lee, and W. K. Joklik.** 1982. Cloning the double-stranded RNA genes of reovirus: sequence of the cloned S2 gene. Proc. Natl. Acad. Sci. U.S.A. **79**:7644–7648.
6. **Dyall-Smith, M. L., and I. H. Holmes.** 1981. Gene-coding assignments of rotavirus double-stranded RNA segments 10 and 11. J. Virol. **38**:1099–1103.
7. **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**:825–829.
8. **Estes, M. K., E. L. Palmer, and J. Obijeski.** 1983. Rotaviruses: a review. Curr. Top. Microbiol. Immunol. **105**:123–184.
9. **Holmes, I. H.** 1983. Rotaviruses, p. 359–423. *In* W. K. Joklik (ed.), The reoviridae. Plenum Publishing Corp., New York.
10. **Hopp, T. P., and K. R. Wood.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
11. **Imai, M., N. Ikegami, K. Akatani, and Y. Furuichi.** 1983. Capped and conserved terminal structures in human rotaviruse genome double-stranded RNA segments. J. Virol. **47**:125–136.
12. **Imai, M., M. A. Richardson, N. Ikegami, A. J. Shatkin, and Y. Furuichi.** 1982. Cloning and nucleotide sequencing of *reoviridae* genomes, p. 15–26. *In* R. W. Compans and D. H. L. Bishop (ed.), Double-stranded RNA viruses. Elsevier Biomedical, New York.
13. **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.
14. **Kantharidis, P., M. L. Dyall-Smith, and I. H. Holmes.** 1983. Completion of the gene coding assignments of SA11 rotavirus: gene products of segments 7, 8, and 9. J. Virol. **48**:330–334.
15. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5252.
16. **Mason, B. B., D. Y. Graham, and M. K. Estes.** 1980. In vitro transcription and translation of simian rotavirus SA11 gene products. J. Virol. **33**:1111–1121.
17. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
18. **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.
19. **McCrae, M. A., and J. G. McCorquodale.** 1983. The molecular biology of rotaviruses. V. Terminal structure of viral RNA species. Virology **126**:204–212.
20. **Payvar, F., and R. T. Shimke.** 1979. Methylmercury hydroxide enhancement of translation and trancription of ovalbumin and conalbumin mRNAs. J. Biol. Chem. **254**:7636–7642.
21. **Petrie, B. L.** 1982. Biological activity of rotavirus particles lacking glycosylated proteins, p. 145–156. *In* R. N. Compans and D. H. L. Bishop (ed.), Double-stranded RNA viruses. Elsevier Biomedical, New York.
22. **Richardson, M. A., and Y. Furuichi.** 1983. Nucleotide sequence of reovirus genome segment S3, encoding non-structural protein sigma NS. Nucleic Acids Res. **11**:6399–6408.
23. **Smith, M. L., I. Lazdins, and I. H. Holmes.** 1980. Coding assignments of double-stranded RNA segments of SA11 rotavirus established in vitro translation. J. Virol. **33**:976–982.