Volume 12 Number 4 1984 Nucleic Acids Research

Cloning and nucleotide sequence of the simian rotavirus gene 6 that codes for the major inner capsid protein

Mary K.Estes*, Bruce B.Mason+, Sue Crawford and Jean Cohen§

Departments of Virology and Epidemiology and *Medicine, Baylor College of Medicine, Houston, TX 77030, USA

Received 8 December 1983; Revised and Accepted 19 January 1984

ABSTRACT

The nucleotide sequence of the gene that codes for the major inner capsid protein of the simian rotavirus SA11 has been determined. A DNA copy of mRNA from gene 6 was cloned in the E. coli plasmid pBR322. The full-length gene is 1357 nucleotides long with a 5'-noncoding region of 23 nucleotides and a 3'-noncoding region of 140 nucleotides. The gene contains a single, long, open reading-frame of 1194 nucleotides capable of coding for a protein of 397 amino acids with a molecular weight of 44,816. The predicted protein product is relatively proline-rich with a net charge at neutral pH of -3.5. One stretch of 53 amino acids (encoded by nucleotides 327-485) is basic.

INTRODUCTION

Rotaviruses, members of the Reoviridae family, have a genome that consists of 11 segments of double-stranded (ds) RNA (1). These viruses are major human and veterinary pathogens that cause serious diarrheal disease throughout the world (1,2). It is anticipated that understanding the structure and function of rotavirus genes will help formulate effective strategies of disease prevention or control.

The 11 segments of genome RNA found in particles are apparently monocistronic and they code for either structural proteins found in virus particles (VP1-9) or for nonstructural (NS) proteins found only in infected cells (1,2). Gene-coding assignments and information on the synthesis, post-translational processing, and functions of many of the rotavirus gene protein products are now known (1-5). Nucleotide sequences have been reported (for one or another rotavirus strain) for the RNA segments coding for the major structural outer capsid glycoprotein [VP7, segment 8 or 9, (6,7)], the nonstructural glycoprotein [NS29, segment 10, (8)], another nonstructural protein [NS35, segment 7 or 8 (9,10)] and a postulated precursor to a minor outer capsid protein [segment 11 (11)]. We report here the cloning and nucleotide sequence of gene 6 of SA11 that codes for

the major internal structural protein (VP6) of virus particles. This gene was cloned from purified mRNA.

VP6 is present in large amounts on the inner shell of the double-capsid of the virion (approximately 80% of the protein mass) (12). It may be a trimer in its native form (12), and it is one of the cross-reacting (common) antigens of rotaviruses that can readily be detected with a variety of serologic assay systems including complement fixation, immune adherence hemagglutination assay (IAHA), radioimmunoassay (RIA), and enzyme-linked immunosorbent assay (ELISA). VP6 contains at least two non-overlapping antigenic areas and one of these has been used to characterize rotaviruses serologically into one of two defined subgroups (13,14). VP6 is not the major neutralization antigen, but it has been suggested that antibodies to VP6 may play a role in neutralization assays in vitro or in protection from infection (13-17). Studies with protein expressed from this cloned gene should help clarify these questions.

MATERIALS AND METHODS

Preparation of viral RNA transcripts and synthesis of cDNA

Simian rotavirus SA11 viral mRNA was prepared from purified virus using the endogenous viral RNA polymerase as previously described (18). Plaque-purified SA11, clone 3 (19) was grown in MA104 cells, and purified by fluorocarbon extraction followed by equilibrium centrifugation in CsCl gradients (18). Purified virus was treated with Na_2EDTA to activate the endogenous polymerase, RNA transcripts were synthesized at 40°C, and purified by electrophoresis on 2% agarose tube gels (Seaplaque agarose, 20-30 cm long, run at 600 volts for 18 hours) (3,18). Individual transcript bands were purified by extraction with phenol after melting the agarose at 65°C (3). For cloning, gene 6 transcripts were polyadenylated with E. coli poly A polymerase (Bethesda Research Labs, Gaithersburg, MD or PL Biochemicals, Milwaukee, WI). Each lot of enzyme was calibrated to determine the time required to add approximately 20 AMP residues. The standard reaction mixture of 100 µl contained (final concentrations): 100mM Tris-HCl (pH 7.9); 10mM $MgCl_2$; 2.5mM $MnCl_2$; 200mM NaCl; 100 µg/ml bovine serum albumin; 0.2mM ATP including 200 µCi of ^{3}H-ATP (Amersham, Arlington Heights, IL); and 8 µg (20 picomoles) of SA11 messenger RNA 6. The tailing reaction was monitored by the incorporation of ^{3}H-ATP into trichloroacetic-acid precipitable product and was terminated by the addition of 40mM EDTA and 0.4% sodium dodecyl sulfate (SDS). The RNA was

extracted with phenol:chloroform:isoamyl alcohol and precipitated with ethanol. To synthesize cDNA, the poly A-tailed m6 RNA was treated with methyl mercuric hydroxide (13mM, Alfa Products, Danvers, MA) at 25°C for 3 minutes and the denatured RNA was diluted threefold into a reaction volume of 100 µl that contained (final concentrations): 50mM Tris-HCl (pH 8.3); 10mM $MgCl_2$; 5mM dithiothreitol (DTT); 0.5mM each of dATP, dCTP and dGTP; 0.2mM dTTP including the 200 µCi of ^{3}H-TTP (Amersham, 20 Ci/mmol); 400U RNasin™ (Promega Biotec, Madison, WI); 0.1 µg/ml oligo dT 12-18; 10 µg/ml actinomycin D; 4mM sodium pyrophosphate and 127U AMV reverse transcriptase (Life Sciences, St. Petersburg, FL). The cDNA reaction was incubated at 46°C for 30 minutes and terminated by the addition of 40mM EDTA and 1% SDS. The cDNA was extracted with phenol, and the unincorporated triphosphates were removed by passage through a Sephadex G-50 column and precipitation with 2.0M ammonium acetate and 66% ethanol. The cDNA strand of the cDNA-RNA heteroduplex was dC-tailed with terminal transferase essentially as described by Land et al. (20). The mixture of dC-tailed DNA was freed of RNA by alkali treatment (0.2M potassium hydroxide for 40 minutes at room temperature). Following neutralization, the mixture was passed through a Sephadex G-50 column and precipitated with ammonium acetate:ethanol. Double-stranded (ds)-cDNA was synthesized in a 60 minute reaction at 42°C that was primed with 100µg/ml of oligo dG and contained approximately 4 picomoles of gene 6 dC-tailed cDNA; 10mM $MgCl_2$; 60mM KCl; 50mM Tris (pH. 8.3); 5mM DTT; 38U reverse transcriptase and 1mM of each unlabeled nucleoside triphosphate. The reaction was terminated with EDTA and SDS, extracted with phenol and freed of triphosphates as described above. The ds-cDNA was dC-tailed as des- cribed above and inserted into Pst-1 cleaved dG-tailed pBR322 (Bethesda Research Laboratories). The pBR322 was annealed to the dC-tailed ds cDNA (approximately 1:1 molar ratio) after heating to 65°C for 2 minutes, followed by an incubation at 56°C for 1 hour. This preparation was used to transform E. coli RR1.

Identification of bacterial colonies containing gene 6 inserts

Bacteria containing recombinant plasmids were identified by antibiotic (Tet^r, Amp^s) sensitivity. Tetracycline resistant colonies were lysed on nitrocellulose filters and screened for the presence of gene 6 sequences by hybridization with ^{32}P-labeled random-primed cDNA synthesized from isolated gene 6 genome RNA or transcript RNA as described by Street et al.(21). Colonies preliminarily identified as containing inserts to gene 6 by this procedure were further characterized by preparing plasmid DNA labeled with ^{32}P

by nick-translation and using it to probe ds RNA segments separated by polyacrylamide gel electrophoresis and transferred to activated 2-aminophenylthioether (DPT) paper (22). Finally, plasmid DNA covalently bound to DPT paper was used to select mRNA transcripts to program the translation of the cloned gene product in cell-free systems derived from wheat germ or rabbit reticulocyte lysates as previously described (18,22).

DNA sequence determination

A recombinant plasmid (pSA11-6) containing the largest cDNA insert encoding the SA11 gene 6 was used in this study. Plasmid DNA was prepared from cleared lysates of transformed bacteria by cesium chloride equilibrium centrifugation in the presence of ethidium bromide. Restriction enzyme sites were mapped by digestion with one or two enzymes.

DNA sequencing was performed by the chemical method of Maxam and Gilbert (23). The DNA fragments were end-labeled as isolated fragments purified from gels by electroelution. Dephosphorylated 5'-ends were labeled using T4 polynucleotide kinase and γ^{32}P-ATP and 3'-ends were labeled with terminal transferase and α^{32}P-dideoxyATP. DNA fragments labeled at only one end, generated by secondary restriction enzyme digestions, were purified by electroelution following electrophoresis on 5% polyacrylamide gels. The sequencing reaction products were resolved under denaturing conditions on 8 or 20% acrylamide gels containing 8M urea. DNA sequence data were analyzed by computer analysis (24,25).

RESULTS

Isolation of clones representing SA11 rotavirus gene 6

Most procedures to clone double-stranded RNA genes have involved the synthesis of single-stranded cDNA from denatured dsRNA, hybridization of the complementary cDNA molecules, and repair of postulated single-stranded gaps in the dsDNA hybrids with DNA polymerase I or reverse transcriptase prior to tailing and insertion into a cloning vector (6-11, 26). This approach sometimes succeeded in producing full-length clones (10-11); in most cases, the 5'-ends of the genes were missing (6-9).

We produced full-length clones to the SA11 gene 6 by the synthesis of cDNA from purified mRNA synthesized *in vitro* with the endogenous viral RNA polymerase. This method is particularly useful when the amount of purified virus is limited. Characterization of the molecules during the steps of synthesis suggested our method yielded relatively homogenous molecules for cloning (Figure 1). Using this procedure, we obtained over 900 bacterial

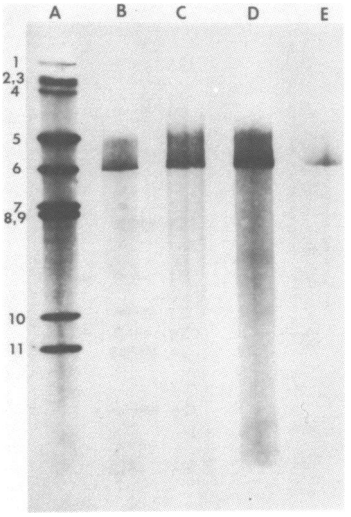

Figure 1. Analysis of SA11-gene 6 cDNA during different stages of its synthesis and cloning from purified gene 6 mRNA. The nucleic acids were analyzed by electrophoresis in a 4% polyacrylamide gel containing 7M urea. This experiment compared the migration of ^{3}H-uridine labeled mRNA from all the viral genes (lane A), the RNA:cDNA product (lane B), the RNA:cDNA product after dC-tailing (lane C), the ss-cDNA (lane D) and ds-cDNA (lane E). The samples in lanes B, C and E were denatured by boiling prior to electrophoresis.

colonies containing gene 6 from 2 μg of purified mRNA. Analysis of 187 of these colonies showed that 34% contained inserts of sufficient size (1000-1500 base pairs) to contain full-length clones. Prior to sequencing, the clones were unequivocally identified as containing gene 6 inserts by hybridization of nick-translated plasmid DNA to separated RNA segments immobilized on DPT-paper (Figure 2 A,B) and by demonstration that mRNA selected to the cloned DNA programmed the synthesis of the expected 41,000 molecular weight protein product (Figure 2 C,D, and E). The clone pSA11-6 analyzed in these experiments was chosen for DNA sequence analysis. Other clones of similar size had identical restriction endonuclease maps.

Nucleotide sequence of SA11 gene 6 cloned DNA

A partial restriction map for pSA11-6 and the strategy used for sequencing is shown in Figure 3. The sequence was determined on both strands of the DNA for 88% of the gene and the rest of the sequence was confirmed by repeated sequencing of overlapping fragments. The sequence of 1357 base pairs (contained between 28G, and 37A-tails made during the process of

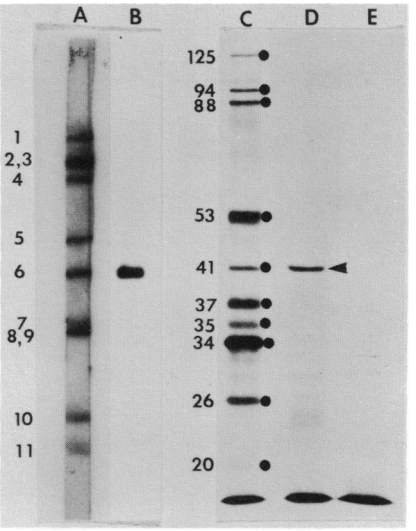

Figure 2. Identification of the plasmids containing DNA inserts of SA11 gene 6. Plasmids containing gene 6 DNA inserts were identified by hybridization to double-stranded RNA segments immobilized on DPT-paper (lanes A and B). The double-stranded RNA segments were hybridized with random-primed cDNA synthesized from all 11 genome segments (lane A) or with ^{32}P-labeled nick-translated plasmid pSA11-6 DNA (lane B). Only RNA segment 6 hybridized with the plasmid DNA. Plasmid DNA used in a hybrid-selection experiment also isolated a mRNA that programmed the synthesis of the gene 6 41,000 molecular weight polypeptide (lane D) in a wheat germ cell-free translation system. The primary translation products synthesized from total transcripts or from no RNA added are shown in lanes C and E, respectively, as previously described (3,18).

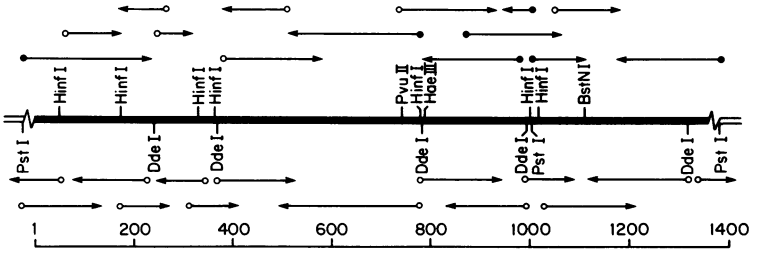

Figure 3. Diagram showing the restriction enzyme fragments used for sequencing. Solid arrows show the direction and extent of sequence obtained from each fragment. The circles indicate whether the sequence was determined from 5'-(○) or 3'-(●) end-labeled DNA fragments.

```
                                MET ASP VAL LEU TYR SER LEU SER LYS THR LEU LYS ASP ALA     14
5'-GGCTTTTAAACGAAGTCTTCAAC ATG GAT GTC CTA TAC TCT TTG TCA AAG ACT CTT AAA GAC GCT      65

ARG ASP LYS ILE VAL GLU GLY THR LEU TYR SER ASN VAL SER ASP LEU ILE GLN GLN PHE     34
AGA GAC AAA ATT GTC GAA GGC ACA TTG TAT TCT AAC GTG AGT GAT CTA ATT CAA CAA TTT    125

ASN GLN MET ILE ILE THR MET ASN GLY ASN GLU PHE GLN THR GLY GLY ILE GLY ASN LEU     54
AAT CAA ATG ATA ATT ACT ATG AAT GGA AAT GAA TTT CAA ACT GGA GGA ATC GGT AAT TTG    185

PRO ILE ARG ASN TRP ASN PHE ASN PHE GLY LEU LEU GLY THR THR LEU LEU ASN LEU ASP     74
CCA ATT AGA AAC TGG AAT TTT AAT TTC GGG TTA CTT GGA ACA ACT TTG CTG AAC TTA GAC    245

ALA ASN TYR VAL GLU THR ALA ARG ASN THR ILE ASP TYR PHE VAL ASP PHE VAL ASP ASN     94
GCT AAT TAT GTT GAA ACG GCA AGA AAT ACA ATT GAT TAT TTC GTG GAT TTT GTA GAC AAT    305

VAL CYS MET ASP GLU MET VAL ARG GLU SER GLN ARG ASN GLY ILE ALA PRO GLN SER ASP    114
GTA TGC ATG GAT GAG ATG GTT AGA GAA TCA CAA AGG AAC GGA ATT GCA CCT CAA TCA GAC    365

SER LEU ARG LYS LEU SER ALA ILE LYS PHE LYS ARG ILE ASN PHE ASP ASN SER SER GLU    134
TCG CTA AGA AAG CTG TCA GCC ATT AAA TTC AAA AGA ATA AAT TTT GAT AAT TCG TCG GAA    425

TYR ILE GLU ASN TRP ASN LEU GLN ASN ARG ARG GLN ARG THR GLY PHE THR PHE HIS LYS    154
TAC ATA GAA AAC TGG AAT TTG CAA AAT AGA AGA CAG AGG ACA GGT TTC ACT TTT CAT AAA    485

PRO ASN ILE PHE PRO TYR SER ALA SER PHE THR LEU ASN ARG SER GLN PRO ALA HIS ASP    174
CCA AAC ATT TTT CCT TAT TCA GCA TCA TTT ACA CTA AAT AGA TCA CAA CCC GCT CAT GAT    545

ASN LEU MET GLY THR MET TRP LEU ASN ALA GLY SER GLU ILE GLN VAL ALA GLY PHE ASP    194
AAT TTG ATG GGC ACA ATG TGG TTA AAC GCA GGA TCG GAA ATT CAA GTC GCT GGA TTT GAC    605

TYR SER CYS ALA ILE ASN ALA PRO ALA ASN ILE GLN GLN PHE GLU HIS ILE VAL PRO LEU    214
TAC TCA TGT GCT ATT AAC GCA CCA GCC AAT ATA CAA CAA TTT GAG CAT ATT GTG CCA CTC    665

ARG ARG VAL LEU THR THR ALA THR ILE THR LEU LEU PRO ASP ALA GLU ARG PHE SER PHE    234
CGA AGA GTG TTA ACT ACA GCT ACG ATA ACT CTT CTA CCA GAC GCG GAA AGG TTT AGT TTT    725

PRO ARG VAL ILE ASN SER ALA ASP GLY ALA THR THR TRP PHE PHE ASN PRO VAL ILE LEU    254
CCA AGA GTG ATC AAT TCA GCT GAC GGG GCA ACT ACA TGG TTT TTC AAC CCA GTG ATT CTC    785

ARG PRO ASN ASN VAL GLU VAL GLU PHE LEU LEU ASN GLY GLN ILE ILE ASN THR TYR GLN    274
AGG CCG AAT AAC GTT GAA GTG GAG TTT CTA TTG AAT GGA CAG ATA ATA AAC ACT TAT CAA    845

ALA ARG PHE GLY THR ILE VAL ALA ARG ASN PHE ASP THR ILE ARG LEU SER PHE GLN LEU    294
GCA AGA TTT GGA ACT ATC GTA GCT AGA AAT TTT GAT ACT ATT AGA CTA TCA TTC CAG TTA    905

MET ARG PRO PRO ASN MET THR PRO ALA VAL ALA VAL LEU PHE PRO ASN ALA GLN PRO PHE    314
ATG AGA CCA CCA AAC ATG ACA CCA GCA GTA GCA GTA CTA TTC CCG AAT GCA CAG CCA TTC    965

GLU HIS HIS ALA THR VAL GLY LEU THR LEU ARG ILE GLU SER ALA VAL CYS GLU SER VAL    334
GAA CAT CAT GCA ACA GTG GGA TTG ACA CTT AGA ATT GAG TCT GCA GTT TGT GAG TCT GTA   1025

LEU ALA ASP ALA SER GLU THR LEU LEU ALA ASN VAL THR SER VAL ARG GLN GLU TYR ALA    354
CTC GCC GAT GCA AGT GAA ACT CTA TTA GCA AAT GTA ACA TCC GTT AGG CAA GAG TAC GCA   1085

ILE PRO VAL GLY PRO VAL PHE PRO PRO GLY MET ASN TRP THR ASP LEU ILE THR ASN TYR    374
ATA CCA GTT GGA CCA GTC TTT CCA CCA GGT ATG AAC TGG ACT GAT TTA ATC ACC AAT TAT   1145

SER PRO SER ARG GLU ASP ASN LEU GLN ARG VAL PHE THR VAL ALA SER ILE ARG SER MET    394
TCA CCG TCT AGG GAG GAC AAT TTG CAA CGC GTA TTT ACA GTG GCT TCC ATT AGA AGC ATG   1205

LEU ILE LYS @@@                                                                    397
CTC ATT AAA TGA GGACCAAGCTAACAACTTGGTATCCAACTTTGGTGAGTATGTAGCTATATCAAGCTGTTTGAA   1280

CTCTGTAAGTAAGGATGCGTATACGCATTCGCTACACTGAGTTAATCACTCTGATGGTATAGTGAGAGGATGTGACC-3'  1357
```

Figure 4. Nucleotide sequence of cloned copy of SA11 gene 6. The (+) sense strand (corresponding to the mRNA) is shown. The predicted amino acid sequence of the protein product is shown and the termination sites are underlined.

cloning the cDNA) is presented in Figure 4. The sequence revealed that we cloned a full-length copy of mRNA 6 as both the conserved 5' terminal sequence (GGCTTTTAAA) and the conserved 3' terminal sequence (ATGTGACC) of rotavirus gene segments were present (6-11). A 5'-noncoding sequence of 23 base pairs preceded the first AUG codon, which initiated an open reading frame of 1194 bases. The sequence AACAUGG flanking and including the first AUG is reminiscent of the consensus PuXXAUGG sequence found near functional eukaryotic translational initiation sites (27). The polypeptide encoded by this open reading frame contains 397 amino acid residues with a calculated molecular weight of 44,816. This value is close to the apparent molecular weight of 41,000 observed in virus particles (3) or following translation of the mRNA for gene 6 (Figure 2B). The first termination codon in this reading frame was found at nucleotide 1215 and was followed by 140 base pairs of 3'-noncoding sequence. No polyadenylation signal was found at the 3' end. No other open reading frame of greater than 100 base pairs is found in the sequence. The nucleotide composition of the gene 6 sequence is 20.0% T, 28.3% G, 33.1% A and 18.6% C.

The deduced protein product of this gene contains 75 (19%) charged residues (Table 1). Of the charged residues, 38 are basic (arginine, lysine and histidine) and 37 are acidic (glutamic and aspartic acid) resulting in a slight net negative charge (-3.5 at pH 7.0). The protein is

Table I: Codon usage and amino acid composition of SA11 gene 6/ VP6

TTT	18	TCT	5	TAT	6	TGT	2	ALA	28	(7.0)[a]
TTC	8	TCC	2	TAC	4	TGC	1	ARG	25	(6.3)
TTA	7	TCA	11	TAA	0	TGA	1	ASN	37	(9.3)
TTG	9	TCG	4	TAG	0	TGG	5	ASP	19	(4.8)
								CYS	3	(0.7)
CTT	4	CCT	2	CAT	5	CGT	0	GLN	18	(4.5)
CTC	4	CCC	1	CAC	0	CGC	1	GLU	18	(4.5)
CTA	9	CCA	15	CAA	14	CGA	1	GLY	18	(4.5)
CTG	2	CCG	3	CAG	4	CGG	0	HIS	5	(1.3)
								ILE	28	(7.0)
ATT	16	ACT	13	AAT	24	AGT	3	LEU	35	(8.8)
ATC	4	ACC	1	AAC	13	AGC	1	LYS	8	(2.0)
ATA	8	ACA	13	AAA	6	AGA	17	MET	11	(2.7)
ATG	11	ACG	2	AAG	2	AGG	6	PHE	26	(6.5)
								PRO	21	(5.2)
GTT	6	GCT	9	GAT	10	GGT	3	SER	26	(6.5)
GTC	4	GCC	3	GAC	9	GGC	2	THR	29	(7.3)
GTA	8	GCA	15	GAA	11	GGA	11	TRP	5	(1.3)
GTG	9	GCG	1	GAG	7	GGG	2	TYR	10	(2.5)
								VAL	27	(6.8)

[a] # of residues (percent)

relatively rich in proline (21 residues) and relatively poor in cysteine (3 residues). The codon usage for this protein is biased against those containing the CG dinucleotide and also against NCC codons as noted for other rotavirus genes (6-11).

DISCUSSION

The replication strategy, the gene and protein functions, and the molecular basis for antigenic variation of the rotaviruses should be understood when a number of isolates have been cultivated in tissue culture, characterized serologically and the structure of their genes determined. The cloning and sequencing of rotavirus genes is one step to understanding the viral genome. The cloning of gene 6 from mRNA synthesized *in vitro* represents an alternative method to that used by others for cloning dsRNAs (6,11,26,28,29). Cloning from mRNA may be essential for virus strains that are available only in limited amounts since large amounts of transcripts can be obtained by activating the endogenous polymerase.

The inner capsid of the rotaviruses contains 11 segments of dsRNA and three or four proteins: VP1, (125,000 molecular weight, 125K) encoded by RNA segment 1; VP2, (94K), encoded by RNA segment 2; and VP6, (41K), encoded by RNA segment 6. A fourth protein, the gene product of RNA segment 3, may also be an inner capsid protein (1,3). The principal protein of the inner capsid is VP6 which is probably located on the external surface of the inner capsid (30). VP6 may be the viral polymerase or one of the other enzymes (guanyltransferase, methyltransferases, nucleotide phosphohydrolase) associated with single-shelled particles (2). In addition, VP6 could bind to RNA genome segments or RNA transcripts, or interact with other proteins of the inner and outer capsid during replication and morphogenesis. Immunocytochemical experiments suggest single-shelled particles bud through the endoplasmic reticulum and acquire a transient envelope prior to their maturation into double-shelled particles (31). The deduced amino acid sequence of VP6 should be examined with these possible functions in mind.

The deduced protein is slightly acidic. One cluster of basic residues is found between amino acid residues 102-154 that might allow the interaction of VP6 with RNA. VP6 is relatively rich in proline content, particularly near the carboxy terminus of the protein; amino acid residues 297 to 376 contain 19 prolines. Cysteine residues are also scarce in the deduced protein sequence. In both respects, VP6 shares features with

the nucleocapsid protein of human respiratory syncytial virus (32). VP6 has been shown to be relatively resistant to digestion with proteolytic enzymes including trypsin (4). This resistance is not due to a lack of arginine or lysine residues, and therefore, it must reflect the inaccessibility of these residues to enzymatic digestion. The molecular weight of the protein calculated from the deduced amino acid sequence is slightly higher than the apparent molecular weight seen by electrophoresis of VP6 in SDS-gels, possibly reflecting secondary structure remaining in this protein after treatment with SDS and reducing agents.

Calcium is important in maintaining the integrity of the outer capsid of rotavirus particles (33). One hypothesis for the acquisition of the outer capsid could be that VP6 is a calcium binding protein. However, the deduced amino acid sequence does not reveal evidence of the consensus sequence seen in several known calcium binding proteins (34). The deduced amino acid sequence of VP6 was also analyzed to determine predicted hydrophilic and hydrophobic domains and secondary structure (24,25,35; Figure 5). Hydrophilic domains were predicted at both the amino and carboxy termini (residues 7-17; 373-380) and between residues 93-133; 140-145 and 225-231. Comparison of the gene and deduced protein sequences in these regions with those of other rotavirus strains that possess different antigenic subgroups may allow mapping the location of the common epitope(s) and the distinct subgroup epitopes on VP6 (13,14,36,37). Conservation of specific hydrophobic domains between virus strains may reveal regions where VP6 might interact with the membrane of the endoplasmic reticulum to trigger budding of particles into the endoplasmic reticulum.

Gene 6 contains the conserved terminal sequences reported for other rotavirus genes, but it differs from the other published gene sequences in

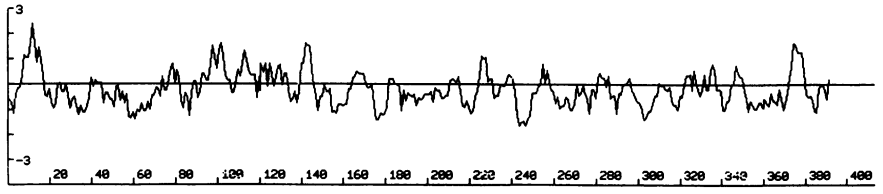

Figure 5. Hydrophilicity plot for deduced SA11 VP6 protein sequence. The average hydrophilicity for the protein sequence of Vp6 was calculated according to Hopp and Woods (35). The y-axis shows the range of hydrophilicity values averaged over 6 amino acids (from 3.0 to -3.0). The sequence numbering on the x-axis corresponds with Figure 4.

that it contains a shorter region of noncoding sequences at the 5' end and a longer region of noncoding sequences at the 3' end. The gene 6 sequence also contains two potential splice sites (38). Some of the other rotavirus genes also contain potential splice sites, although this feature has not been commented on previously. Since the rotaviruses replicate in the cytoplasm and no requirement of the nucleus for replication is known, the significance of this observation is not clear. The occurrence of a potential splice site in several of the rotavirus genes may represent a chance observation.

The production of the major protein and of the common antigen of rotavirus particles from these clones could be useful to develop seroepidemiologic assays or to produce VP6 as a polypeptide vaccine. It has been suggested that vaccination with heterologous rotaviruses may induce protective antibodies (39-41), but the results of other studies do not agree (42,43). The possibility that antibodies directed at VP6 have a protective effect could be tested directly following inoculation of VP6 produced from cloned genes. In addition, this cloned gene may be useful to develop new, rapid diagnostic tests for rotavirus using spot hybridization assays. Finally, the expression of VP6 in either prokaryotes or eukaryotes from cloned genes should help elucidate the function of this protein in virus structure, replication and assembly.

ACKNOWLEDGEMENTS

We thank Drs. Larry Rothblum and Subrahmanyan Chirala for teaching us DNA sequencing and for many helpful discussions. We also thank Brad Ericson for performing the hybrid-selected translation experiment and Gerry Both and his colleagues for sharing his unpublished sequence data on VP6 with us. This work was supported in part by grants from the Thrasher Research Fund, the Diarrheal Disease Program of the World Health Organization, and Public Health Service Grant CA09197 from the National Cancer Institute. J.C. was supported by INRA and NATO fellowships.

+Present address: Microbiology Division, Wyeth Laboratories, Philadelphia, PA 19101, USA

§Present address: Station de Virologie et D'Immunologie INRA, Route de Thiverval, 78850 Thiverval-Grignon, France

REFERENCES

1. Estes, M.K., Palmer, E.L. and Obijeski, J.F. (1983) Curr. Topics in Microbiol. and Immunol. 105, 124-184.

2. Holmes, I.H. (1983) in The Reoviruses, Joklik, W.K., Ed., pp. 359-423, Plenum Publ. Co., New York.
3. Mason, B.B., Graham, D.Y. and Estes, M.K. (1983) J. Virol. 46, 413-423.
4. Ericson, B.L., Graham, D.Y., Mason, B.B. and Estes, M.K. (1982) J. Virol. 42, 825-839.
5. Arias, C.F., Lopez, S. and Espejo, R.T. (1982) J. Virol. 41, 42-50.
6. Both, G.W., Mattick, J.S. and Bellamy, A.R. (1983) Proc. Natl. Acad. Sci. USA 80, 3091-3095.
7. Elleman, T.C., Hoyne, P.A., Dyall-Smith, M.L., Holmes, I.H. and Azad, A.A. (1983) Nucleic Acids Res. 11, 4689-4701.
8. Both, G.W., Siegman, L.J., Bellamy, A.R. and Atkinson, P.H. (1983) J. Virol. 48, 335-339.
9. Both, G.W., Bellamy, A.R., Street, J.E. and Siegman, L.J. (1982) Nucleic Acids Res. 10, 7075-7088.
10. Dyall-Smith, M.L., Elleman, T.C., Hoyne, P.A., Holmes, I.A. and Azad, A.A. (1983) Nucleic Acid Res. 11, 3351-3362.
11. Imai, M., Richardson, M.A., Ikegami, N., Shatkin, A.J. and Furuichi, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 373-377.
12. Novo, E. and Esparza, J. (1981) J. Gen. Virol. 56, 325-335.
13. Greenberg, H., McAuliffe, V., Valdesuso, J., Wyatt, R., Flores, J., Kalica, A., Hoshino, Y. and Singh, N. (1983) Inf. Imm. 39, 91-99.
14. Thouless, M.E., Beards, G.M. and Flewett, T.H. (1982) Arch. Virol. 73, 219-230.
15. Bastardo, J.W., McKimm-Breschkin, J.L., Sonza, S., Mercer, L.D. and Holmes, I.H. (1981) Inf. Imm. 34, 641-647.
16. Killen, H.M. and Dimmock, N.J. (1982) J. Gen. Virol. 62, 297-311.
17. Offit, P.A., Clark, H.F. and Plotkin, S.A. (1983) Inf. Imm. 42, 293-300.
18. Mason, B.B., Graham, D.Y. and Estes, M.K. (1980) J. Virol. 33, 1111-1121.
19. Estes, M.K., Graham, D.Y., Ramig, R.F. and Ericson, B.L. (1982) Virology 122, 8-14.
20. Land, M.H., Grez, M., Hauser, H., Lindenmaier, N. and Schutz, G. (1982) Nucleic Acids Res. 9, 2251-2266.
21. Street, J.E., Croxson, M.C., Chadderton, W.F. and Bellamy, A.R. (1982) J. Virol. 43, 369-378.
22. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory.
23. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
24. Larson, R. and Messing, J. (1983) DNA 2: 31-35.
25. Chou, P.Y. and Fasman, G.D.(1978) in Advances in Enzymology, Meister, A., Ed., pp. 45-148, Wiley, New York.
26. Cohen, J., Mason, B.B. and Estes, M.K. (1983) in Double-Stranded RNA Viruses, Compans, R.W. and Bishop, D.H.L. Eds., pp. 65-71, Elsevier Science Publishing Co.
27. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
28. Cashdollar, L.W., Esparza, J., Hudson, G.R., Chmelo, R., Lee, P.W.K. and Joklik, W.K. (1982) Proc. Natl. Acad. Sci. 79, 7644-7648.
29. McCrae, M.A. and McCorquodale, J.G. (1982) J. Virol. 44, 1076-1079.
30. Bican, P., Cohen, J., Charpilienne, A. and Scherrer, R. (1982) J. Virol. 43, 1113-1117.
31. Altenburg, B.C., Graham, D.Y. and Estes, M.K. (1980) J. Gen. Virol. 46, 75-85.
32. Elango, N. and Venkatesan, S. (1983) Nucleic Acids Res. 11, 5941-5951.

33. Cohen, J., Laporte, J., Charpilienne, A., Scherrer, R. (1979) Arch. Virol. 60, 177-186.
34. Gariépy, J. and Hodges, R.S. (1983) FEBS Letters 160, 1-5.
35. Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824-3828.
36. Lambert,J-P., Marissens, D., Marbehant, P. and Zissis, G. (1983) J. Med. Virol. 11, 31-38.
37. Tufvesson, B. (1983) J. Med. Virol. 12, 111-117.
38. Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857.
39. Wyatt, R.G., Mebus, C.A. and Yolken, R.H. et al. (1979) Science 203, 548-550.
40. Vesikari, T., Isolauri, E., Delem, A., D'Hondt, E., Andre, F.E. and Zissis, G. (1983) Lancet ii, 807-811.
41. Wyatt, R.G., Kapikian, A.Z. and Mebus, C.A. (1983) J. Clin. Micro. 18, 505-508.
42. Gaul, S.K., Simpson, T.F., Woode, G.M. and Fulton, F.W. (1982) J. Clin. Micro. 16, 495-503.
43. Woode, G.N., Kelso, N.E., Simpson, T.F., Gaul, S.K., Evans, L.E. and Babiuk, L. (1983) J. Clin. Micro. 18, 358-364.