

# Three major alleles of rotavirus NSP4 proteins identified by sequence analysis

Yasuo Horie,<sup>1,2</sup> Osamu Masamune<sup>1</sup> and Osamu Nakagomi<sup>2</sup>

The First Department of Internal Medicine<sup>1</sup>, and Department of Microbiology<sup>2</sup>, Akita University School of Medicine, 1-1-1 Hondo, Akita 010, Japan

Comparison of nonstructural glycoprotein NSP4 gene sequences from 22 rotavirus strains originating from six host species and of 14 different combinations of G and P types revealed the presence of three distinct NSP4 alleles, represented by strains Wa, KUN and AU-1. Genetic distances between any of these alleles (18.0%) were significantly greater than those within each allele (5.5%) and phylogenetic analysis suggested that divergence into three distinct alleles had occurred at about the same time during evolution. While amino acid variation among strains was minimal in the amino-

terminal two-thirds of the protein (aa 1–130), variability increased toward the carboxy terminus of the enterotoxigenic peptide region (aa 114–135) and was greatest between residues 135 and 141. Comparison of the amino acid sequences corresponding to the enterotoxigenic peptide region between strains isolated from asymptomatic neonates and those from children with diarrhoea failed to identify any conserved changes that correlated with the capacity of the virus to cause disease. Amino acids were relatively conserved in the domains important for viral morphogenesis.

## Introduction

Group A rotaviruses, members of the genus *Rotavirus* in the family *Reoviridae*, have been established as the single most important aetiological agents of severe diarrhoea of infants and young children worldwide (Kapikian & Chanock, 1996). The rotavirus genome, which is contained in a triple-layered capsid, consists of 11 segments of double-stranded RNA, with six genome segments coding for the structural proteins (VP1, VP2, VP3, VP4, VP6 and VP7) and five segments coding for the nonstructural proteins (NSP1–5) (Estes, 1996). Although most genome segments are monocistronic, evidence indicates that two primary protein products, NSP5 and NSP5a, are expressed in infected cells from alternative long and short open reading frames of genome segment 11 (Mattion *et al.*, 1991). Of the six nonstructural proteins NSP4, which is encoded by gene segment 10 in most strains, functions as an intracellular receptor that mediates the acquisition of a transient envelope as subviral particles bud into the lumen of the endoplasmic

reticulum (Estes, 1996). NSP4 attracted broader attention when Ball *et al.* (1996) provided evidence that NSP4 acts as a viral enterotoxin in the mouse model system. The NSP4 of the SA11 strain, more specifically a 22 peptide region spanning amino acid residues 114–135, was shown to trigger a signal transduction pathway, inducing an increase of intracellular  $Ca^{2+}$ . This results in decreased absorption of  $Na^+$  and water, thus causing secretory diarrhoea. Ball *et al.* (1996) further identified tyrosine at residue 131 as a key amino acid that is associated with the toxigenic activity of NSP4.

This discovery led us to hypothesize that, if NSP4 is an enterotoxin, then the protein, and particularly the enterotoxigenic peptide region, might be conserved between strains from diverse host species. We also wanted to know to what degree the structural features that were involved in rotavirus morphogenesis were conserved among strains and through evolution. We therefore sequenced seven NSP4 genes from various human and animal rotavirus strains and compared them with six published and nine unpublished NSP4 sequences currently available in the DNA databases.

**Author for correspondence:** Osamu Nakagomi.

Fax +81 188 36 2607. e-mail onakagom@ipc.akita-u.ac.jp

The GenBank, EMBL and DDBJ accession numbers of the sequences reported in this paper are D88829, D88830, D88831, D88832, D88833, D89873 and D89874.

## Methods

■ **Virus strains.** The rotavirus strains used in this study are summarized in Table 1. Of those, AU32, KUN, AU-1, FRV-1, RS15, FRV64 and OSU were grown in MA104 cells in the presence of 0.5  $\mu$ g

**Table 1.** Rotavirus strains used in this study

Strain	Host species	G type	P type	Subgroup	Accession no.	Reference
Wa	Human	1	1A[8]	II	K02032	Okada <i>et al.</i> (1984)
ST3	Human	4	2A[6]	II	U59110	–
RV3	Human	3	2A[6]	II	U42628	–
RV4	Human	1	1A[8]	II	U59108	–
AU32	Human	9	1A[8]	II	D88830	This study
M37	Human	1	2A[6]	II	U59109	–
OSU	Porcine	5	9[7]	I	D88831	This study
YM	Porcine	11	9[7]	I	X69485	Lopez & Arias (1993)
KUN	Human	2	1B[4]	I	D88829	This study
RV5	Human	2	1B[4]	I	U59103	–
S2	Human	2	1B[4]	I	U59104	–
1076	Human	2	2A[6]	I	U59105	–
E210	Human	2	1B[4]	II	U59107	–
A28	Human	10	ND	I	D01145	Ballard <i>et al.</i> (1992)
UK	Bovine	6	7[5]	I	K03384	Baybutt & McCrae (1984)
NCDV	Bovine	6	6[1]	I	X06806	Powell <i>et al.</i> (1988)
SA11	Simian	3	[2]	I	K01138	Both <i>et al.</i> (1983)
AU-1	Human	3	3[9]	I	D89873	This study
FRV-1	Feline	3	3[9]	I	D89874	This study
RS15	Canine	3	5[3]	I	D88832	This study
FRV64	Feline	3	5[3]	I	D88833	This study
RRV	Simian	3	5[3]	I	L41247	–

ND, Not described.

–, Unpublished data.

of trypsin/ml (type IX, Sigma). Virus particles were purified from virus-infected MA104 cell cultures by centrifugation at 36 000 r.p.m. for 3 h in a Beckman type 45Ti rotor, and the pellet was then centrifuged through 30% (w/v) sucrose at 38 000 r.p.m. for 3 h in a Beckman type SW41Ti rotor.

**■ RT-PCR, cloning and sequencing.** Genomic RNA was extracted with phenol–chloroform from purified virions and reverse transcribed with an NSP4 gene-specific 3'-terminal primer (EndG10: 5' GGTCACATTAAGACCATTC 3'). The cDNA was amplified by PCR with the same 3'-terminal primer and the 5'-terminal primer (BegG10: 5' GGCTTTTAAAAGTTCTGTTC 3'), cloned into pCRII vector (Invitrogen), and sequenced on Applied Biosystems 373 or ABI PRISM 377 automated DNA sequencers by the DyeDeoxy terminator method (Perkin Elmer).

At least two independent cDNA clones for each strain were sequenced at least twice, mostly on both strands, with an average sequence-run length for 'unambiguous reading' from each primer of approximately 200 nt for the 373 sequencer and approximately 350 nt for the 377 sequencer. Six primers were used to sequence almost all strains. M13 forward and reverse primers, BegG10 and EndG10, and BegIG10 (5' TCCAACCATGAAAATAGC 3'; based on the sequence nt 194–211) and EndIG10 (5' GCAGCTCAACCTCTCACA 3'; based on the sequence nt 583–566). Several additional unique primers were made wherever sequencing with those common primers was difficult or ambiguous.

**■ Sequence and phylogenetic analyses.** For sequence and phylogenetic analyses, we used only the amino acid sequences that were deduced from the open reading frame spanning nt 42–566, thus avoiding

the effect of incorporated terminal primer sequences. The hydrophobicity profiles of the NSP4 amino acid sequences were analysed with a program in the GeneWorks version 2.5 software package (IntelliGenetics) that utilized the Kyte and Doolittle algorithm; the NSP4 sequences themselves were aligned with the CLUSTAL W software package (Thompson *et al.*, 1994). Genetic distances between every pair of the NSP4 amino acid sequences were calculated with the ProtDist program in the Phylip 3.5c software package (Felsenstein, 1993). A phylogenetic tree based on the neighbour-joining method of Saitou & Nei (1987) was drawn with the N-J plot program in the CLUSTAL W package (Thompson *et al.*, 1994).

## Results

We determined NSP4 sequences of seven human and animal rotavirus strains and analysed them together with six published and nine unpublished NSP4 sequences in the DNA databases (Table 1). The overall amino acid conservation among strains was high, with observed amino acid identities being greater than 79% (Table 2). To infer phylogenetic relationships more correctly, the genetic distances between strains (the percentage of the expected numbers of amino acids changed) were computed for each pair of the NSP4 amino acid sequences using the Dayhoff PAM 001 matrix (Dayhoff *et al.*, 1978). The genetic distances ranged from 0.5% (between AU-1 and FRV-1, between KUN and S2 and between KUN and E210) to 22.8% (between 1076 and RS15) (Table 2). A closer look at the genetic distances allowed classification of these

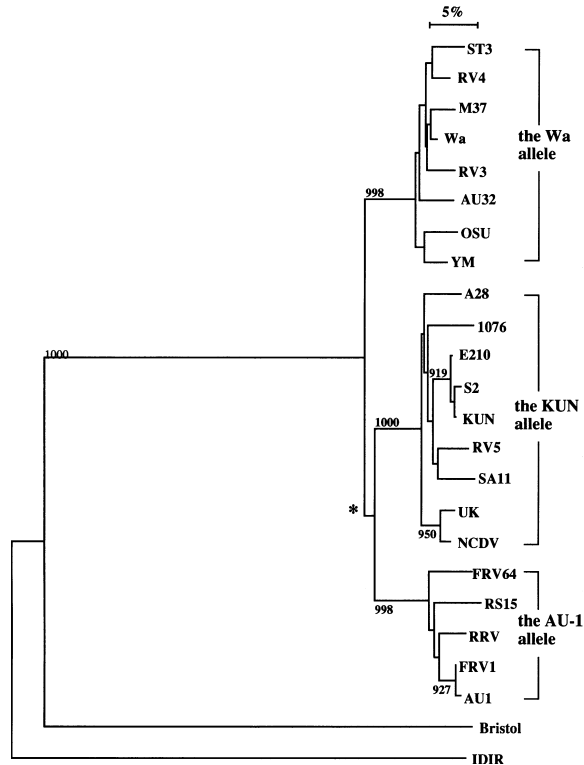


Fig. 1. Phylogenetic tree for NSP4s constructed by the neighbour-joining method of Saitou & Nei (1987). The horizontal distance connecting two NSP4s is proportional to the genetic distance between these two NSP4s. The distance is expressed as the number of amino acid substitutions per 100 sites (% divergence). Bootstrap probabilities equal or greater than 900 (out of 1000 trials) are indicated. The bootstrap probability at the asterisk node (\*) is 58.3%. This node is referred to in the text.

NSP4 genes into three distinct alleles with an average distance within alleles of 5.5% (95% confidence interval 5.3–5.7) and with an average distance between alleles of 18.0% (95% confidence interval 17.9–18.1). Among the first allele, represented by human strain Wa (G1P1A[8]), were five other human strains and two porcine strains. The G and P types of the Wa allele strains span G types 1, 3, 4, 5, 9 and 11, and P types 1A[8], 2[6] and 9[7]. Among the second allele, represented by human strain KUN (G2P1B[4]), were five other human strains, two bovine strains and the simian SA11 strain. The G and P types of the KUN allele strains span G types 2, 3, 6 and 10 and P types 1B[4], 2A[6], 6[1], 7[5] and [2]. The third allele included human AU-1 strain (G3P3[9]), feline FRV-1 and FRV64, canine RS15 and rotavirus vaccine candidate RRV. The G and P types of the AU-1 allele strains span G type 3 and P types 3[9] and 5[3]. The average genetic distance within each allele is 5.7% (range expressed by  $2 \times$  standard deviation: 2.7–8.7%) for the Wa allele, 5.2% (0.8–9.6%) for the KUN allele and 5.5% (1.1–9.9%) for the AU-1 allele, suggesting that each allele contains similarly micro-heterogeneous NSP4 genes.

A phylogenetic tree was constructed by the neighbour-joining method using NSP4 from group C rotavirus (the Bristol

strain) and group B rotavirus (the IDIR strain) as outgroups, and confirmed three major lineages corresponding to three distinct NSP4 alleles (Fig. 1). Because the horizontal distance connecting the node containing the Wa lineage and another node that contained the other lineages (asterisk in Fig. 1) was short (0.6%) and the bootstrap probability at this node was low (58.3%), the branching order into the three lineages was inconclusive. This in turn suggests that divergence into three distinct lineages occurred at about the same time during evolution.

Previous studies identified a number of structural features along the entire NSP4 protein that are of biological importance for rotavirus morphogenesis (Au *et al.*, 1993; Taylor *et al.*, 1996; Tian *et al.*, 1996) other than the enterotoxigenic peptide (Fig. 2). To ascertain whether such structural features were conserved among diverse rotavirus strains, a variability index was calculated at each amino acid position in the sequence according to the formula proposed by Wu & Kabat (1970) who had used it to identify the variable regions in the light chains of immunoglobulin. This formula defines a variability index as the number of different amino acids observed at a given position divided by frequency of occurrence of the most common amino acid at that residue. By definition, an invariant residue would have a value of 1 while the theoretical upper limit for 20 amino acids randomly occurring at a given position of 22 strains would be 220. A highly variable amino acid position was defined in this paper as the position possessing a variability index greater than 6.73 which was significantly ( $P = 0.01$ ) greater than the average variability index (2.10). Sequence variation increased immediately downstream from the toxic peptide region and remained high toward the carboxy terminus of the NSP4, with the greatest sequence variation occurring between residues 135 and 141 (Fig. 2).

The amino acid sequences corresponding to the enterotoxigenic peptide region (114–135) are aligned in Fig. 3 in order to examine whether there is any specific amino acid change that is different between strains isolated from asymptomatic neonates ('asymptomatic' strains) and those from infants and young children with diarrhoea ('symptomatic' strains). Among four asymptomatic strains and seven symptomatic strains a total of seven amino acid substitutions was found but none was segregated according to whether the strain was asymptomatic or not.

Three hydrophobic domains, designated H1, H2 and H3, previously identified in SA11 and NCDV NSP4 proteins (Chan *et al.*, 1988; Bergmann *et al.*, 1989) were also predicted by the Kyte and Doolittle algorithm for all NSP4 sequences examined in this study (data not shown) (Fig. 2). The oligomerization-associated domain (Taylor *et al.*, 1996), the amino-terminal half of the VP4-binding domain (Au *et al.*, 1993), the amino-terminal three-quarters of the enterotoxigenic peptide (Ball *et al.*, 1996) as well as the membrane-destabilization activity domain (MDA) (Tian *et al.*, 1996) were composed of well-conserved amino acids. By sharp contrast, the carboxy-terminal half of the

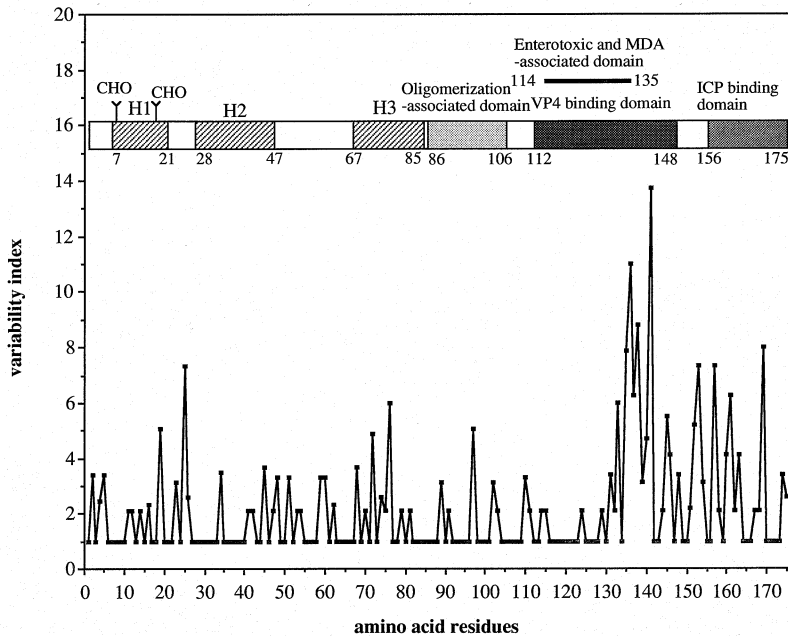


Fig. 2. Variability plot according to Wu & Kabat (1970) in relation to various functional domains that were previously identified (Chan *et al.*, 1988; Bergmann *et al.*, 1989; Au *et al.*, 1993; Ball *et al.*, 1996; Taylor *et al.*, 1996; Tian *et al.*, 1996). Three hydrophobic domains are indicated as H1 (aa 7–21), H2 (aa 28–47) and H3 (aa 67–85). Amino acid residues for H1, H2 and H3 are according to Chan *et al.* (1988). Carbohydrate side chains are indicated as CHO. The oligomerization-associated domain (aa 86–106), the VP4 binding domain (aa 112–148), and the inner-capsid particle (ICP) binding domain (aa 156–175) are boxed. Solid bar indicates the enterotoxigenic peptide and the MDA domain, both spanning from aa 114 to aa 135.

Strain	114	131	135
Asymptomatic strains	↓	↓	↓
ST3	DKLTTREIEIQ	IELLKRIHDN	L
RV3	.....	V.....C.Y..	..
M37	.....	V.....	..
1076	.R.....	V.....K	..
Wa	.....	V.....	..
AU32	.....	V.....N.	..
Symptomatic strains			
RV4	.....	V.....	..
A28 ***	.....	V.....Y.K	..
E210	.....	V.....Y.K	..
S2	.....	V.....Y.K	..
KUN	.....	V.....Y.K	..
RV5	.....	V.....Y.K	..

Fig. 3. Comparison of amino acid sequences of the enterotoxigenic peptide region (aa 114–135) between strains isolated from asymptomatic neonates ('asymptomatic' strains) and those from infants and young children with diarrhoea ('symptomatic' strains): \* indicates rotavirus strains isolated from asymptomatic neonates; \*\* indicates rotavirus strains isolated from infants and young children with diarrhoea; and \*\*\* indicates a rotavirus strain isolated from a child with severe combined immunodeficiency.

VP4-binding domain and the inner-capsid particle-binding domain (Taylor *et al.*, 1996) showed high sequence variation (Fig. 2).

### Discussion

By comparing 22 NSP4 amino acid sequences from human and animal group A rotavirus strains, this study has identified three distinct NSP4 alleles, i.e. the Wa, KUN and AU-1 alleles, that appear to have been diverged from a common ancestral allele at about the same time during evolution. However, the evolutionary event that formed the basis for the currently

observed NSP4 polymorphism is beyond our speculation. As far as human rotavirus strains are concerned, classification into three NSP4 alleles agrees with three human rotavirus genogroups, i.e. the Wa, DS-1 and AU-1 genogroups, previously identified by RNA–RNA hybridization (Nakagomi *et al.*, 1989). For example, the strains possessing the Wa NSP4 allele such as human strains AU32, ST3 and M37, all of which have long RNA patterns and subgroup II specificity, were previously shown to belong to the Wa genogroup (Nakagomi *et al.*, 1989, 1990; Nakagomi & Nakagomi, 1991). Of the strains possessing the KUN NSP4 allele, human strains S2, RV5 and 1076, all of which have short RNA patterns, G2 and subgroup I specificity, belong to the DS-1 genogroup of which KUN is a prototype (Nakagomi *et al.*, 1989; Nakagomi & Nakagomi, 1991). The AU-1 strain which possesses a long RNA pattern yet subgroup I specificity represents its own NSP4 allele distinct from the Wa and KUN alleles. It should be noted, however, that animal rotaviruses from at least one non-human host-species share the same allele with human rotaviruses irrespective of their genogroup relationships with human rotaviruses. These inter-species relationships suggest acquisition of NSP4 alleles beyond host-species barriers during the course of evolution.

While we failed to identify any specific amino acid substitution indicative of virulence or avirulence in the region of the enterotoxigenic peptide (Fig. 3), Kirkwood *et al.* (1996) compared asymptomatic and symptomatic G3P2A[6] strains isolated from neonates in Melbourne, Australia and made an interesting observation. They found that amino acid residue 135 of NSP4 was associated with virulence in humans because only amino acid substitution at this residue in NSP4, together with some substitutions in VP4 and VP7, invariably correlated with whether the strain was derived from asymptomatic or

**Table 2.** Amino acid identities (above diagonal) and genetic distances (below diagonal) between NSP4s of various human and animal rotavirus strains

	G type	P type	Wa	AU32	RV3	RV4	ST3	M37	OSU	YM	KUN	RV5	S2	1076	E210	A28	UK	NCDV	SA11	AU1	FRV1	FRV64	RS15	RRV
Wa	1	1A[8]	—	95	96	96	95	97	94	95	84	82	83	82	84	83	84	85	82	84	85	82	83	84
AU32	9	1A[8]	4.4	—	94	94	92	93	93	93	83	82	82	82	83	82	84	85	81	85	86	82	84	85
RV3	3	2A[6]	3.3	6.1	—	95	94	95	92	94	84	82	83	81	84	83	82	83	82	82	83	81	81	82
RV4	1	1A[8]	3.3	5.5	4.9	—	95	95	93	95	84	82	83	82	85	84	84	85	83	84	85	82	83	84
ST3	4	2A[6]	4.4	7.9	6.1	4.4	—	94	91	93	82	81	82	81	82	81	82	82	80	82	83	80	81	82
M37	1	2A[6]	2.7	6.7	4.9	4.4	5.5	—	92	93	84	82	83	82	85	83	85	85	82	83	83	81	82	83
OSU	5	9[7]	5.5	7.3	7.2	6.6	8.4	7.2	—	95	85	84	84	83	85	83	85	86	83	82	83	82	81	82
YM	11	9[7]	4.9	7.2	6.6	4.9	7.8	7.2	5.4	—	83	82	82	82	83	83	83	84	82	82	83	80	82	82
KUN	2	1B[4]	16.4	17.6	16.4	16.3	18.4	16.4	15	18.2	—	95	99	94	99	94	94	94	94	83	84	84	81	83
RV5	2	1B[4]	18.2	18.8	18.2	18.1	20.2	18.3	16.1	19.5	4.5	—	95	93	95	93	94	94	94	83	83	83	80	82
S2	2	1B[4]	17	18.2	17	16.9	19.1	17.1	15.7	18.9	0.5	5	—	93	99	94	93	94	94	83	83	83	80	82
1076	2	2A[6]	18.3	18.9	19.6	18.2	20	18.4	16.8	18.9	6.2	7.2	6.7	—	93	91	92	91	91	82	83	81	79	82
E210	2	1B[4]	15.7	16.7	16.3	15.6	18.3	15.8	14.4	17.5	0.5	5	0.8	6.7	—	95	94	95	94	84	85	85	81	83
A28	10	ND*	17	18.2	17.6	16.3	19.7	17.1	16.8	17.7	5.6	6.8	6.2	8.6	5	—	93	94	92	84	85	84	81	83
UK	6	7[5]	16.4	16.3	18.2	16.3	19	15.7	15	17	6.1	6.2	6.7	7.9	5.5	6.7	—	98	92	86	86	84	83	85
NCDV	6	6[1]	15.8	15.7	17.6	15.7	18.4	15.8	14.5	17	5.6	5.7	6.2	8.5	5	5.6	2.2	—	92	86	86	84	83	85
SA11	3	[2]	18.8	19.4	18.8	17.4	20.9	18.2	17.2	18.8	5.5	6.2	6.1	9.1	6.1	7.9	7.8	8	—	82	82	83	81	81
AU1	3	3[9]	17.1	15.8	19	17.1	19.2	18.6	18.5	19	17.6	18.3	18.3	19.6	16.9	17.2	15.7	15.2	19.5	—	99	93	94	95
FRV1	3	3[9]	16.5	15.2	18.4	16.5	18.6	17.9	17.8	18.4	17	17.6	17.7	18.9	16.3	16.6	15.1	14.5	18.9	0.5	—	93	94	96
FRV64	3	5[3]	19.9	19.8	20.5	19.8	22	20.1	19.2	21.8	17.2	18.7	17.9	21.2	16.5	17.5	17.9	17.4	18.3	6.7	7.3	—	92	92
RS15	3	5[3]	18.3	17	20.1	17.6	20.3	19.7	19.6	19.6	20.8	21.4	21.4	22.8	20	20.4	18.8	18.3	20.9	5.5	6.1	7.9	—	93
RRV	3	5[3]	17.1	15.8	19	17	19.2	18.5	18.4	19	18.2	18.8	18.9	20.2	17.5	17.8	16.3	15.7	20.1	4.9	4.4	7.9	7.2	—

\* Not described

symptomatic neonates. Kirkwood *et al.* (1996) found isoleucine at aa 135 only in 'asymptomatic' G3P2A[6] strains, but it was shared by both 'asymptomatic' (ST3, M37 and RV3) and 'symptomatic' strains (Wa, RV4, AU32, E210, S2 and KUN). In addition, methionine at aa 135 was shared by 'asymptomatic' 1076 and 'symptomatic' RV5 strains. Thus, as Kirkwood *et al.* (1996) speculated, the specific substitutions they found may be specific for G3P2A[6] strains from Melbourne. Thus only comparison of strains of similar genetic composition but possessing different capacity to induce symptoms, may yield reliable results.

While the carboxy-terminal half of the VP4-binding domain contains the region of the greatest sequence variation, its amino-terminal half is well conserved. Similarly, the amino-terminal three-quarters of the MDA domain is well conserved. Thus, we speculate that these conserved regions are more important in their functions than the highly variable carboxy terminus. It remains to be determined, however, whether the region of the highest sequence variation is the result of a lack of functional constraint or the reflection of immune selective pressure to which the ligand of NSP4, i.e. VP4, may be subjected.

This study was supported in part by a grant-in-aid for a general scientific study from the Ministry of Education, Science, Sport and Culture, Japan.

## References

- Au, K.-S., Mattion, N. M. & Estes, M. K. (1993). A subviral particle binding domain on the rotavirus nonstructural glycoprotein NS28. *Virology* **194**, 665–673.
- Ball, J. M., Tian, P., Zeng, C. Q.-Y., Morris, A. P. & Estes, M. K. (1996). Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein. *Science* **272**, 101–104.
- Ballard, A., McCrae, M. A. & Desselberger, U. (1992). Nucleotide sequences of normal and rearranged RNA segments 10 of human rotaviruses. *Journal of General Virology* **73**, 633–638.
- Baybutt, H. N. & McCrae, M. A. (1984). The molecular biology of rotaviruses. VII. Detailed structural analysis of gene 10 of bovine rotavirus. *Virus Research* **1**, 533–541.
- Bergmann, C. C., Maass, D., Poruchynsky, M. S., Atkinson, P. H. & Bellamy, A. R. (1989). Topology of the non-structural rotavirus receptor glycoprotein NS28 in the rough endoplasmic reticulum. *EMBO Journal* **8**, 1695–1703.
- Both, G. W., Siegman, L. J., Bellamy, A. R. & Atkinson, P. H. (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. *Journal of Virology* **48**, 335–339.
- Chan, W.-K., Au, K.-S. & Estes, M. K. (1988). Topography of the simian rotavirus nonstructural glycoprotein (NS28) in the endoplasmic reticulum membrane. *Virology* **164**, 435–442.
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, pp. 345–352. Edited by M. O. Dayhoff. Silver Spring, Md.: National Biomedical Research Foundation.

- Estes, M. K. (1996).** Rotaviruses and their replication. In *Fields Virology*, 3rd edn, pp. 1625–1655. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.
- Felsenstein, J. (1993).** PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, Wash., USA.
- Kapikian, A. Z. & Chanock, R. M. (1996).** Rotaviruses. In *Fields Virology*, 3rd edn, pp. 1657–1708. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.
- Kirkwood, C. D., Coulson, B. S. & Bishop, R. F. (1996).** G3P2 rotaviruses causing diarrhoeal disease in neonates differ in VP4, VP7 and NSP4 sequence from G3P2 strains causing asymptomatic neonatal infection. *Archives of Virology* **141**, 1661–1676.
- Lopez, S. & Arias, C. F. (1993).** Sequence analysis of rotavirus YM VP6 and NS28 proteins. *Journal of General Virology* **74**, 1223–1226.
- Mattion, N. M., Mitchell, D. B., Both, G. W. & Estes, M. K. (1991).** Expression of rotavirus proteins encoded by alternative open reading frames of genome segment 11. *Virology* **181**, 295–304.
- Nakagomi, O. & Nakagomi, T. (1991).** Genetic diversity and similarity among mammalian rotaviruses in relation to interspecies transmission of rotavirus. *Archives of Virology* **120**, 43–55.
- Nakagomi, O., Nakagomi, T., Akatani, K. & Ikegami, N. (1989).** Identification of rotavirus genogroups by RNA–RNA hybridization. *Molecular and Cellular Probes* **3**, 251–261.
- Nakagomi, T., Ohshima, A., Akatani, K., Ikegami, N., Katsushima, N. & Nakagomi, O. (1990).** Isolation and molecular characterization of a serotype 9 human rotavirus strain. *Microbiology and Immunology* **34**, 77–82.
- Okada, Y., Richardson, M. A., Ikegami, N., Nomoto, A. & Furuichi, Y. (1984).** Nucleotide sequence of human rotavirus genome segment 10, an RNA encoding a glycosylated virus protein. *Journal of Virology* **51**, 856–859.
- Powell, K. F. H., Gunn, P. R. & Bellamy, A. R. (1988).** Nucleotide sequence of bovine rotavirus genomic 10: an RNA encoding the viral non-structural glycoprotein. *Nucleic Acids Research* **16**, 763.
- Saitou, N. & Nei, M. (1987).** The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.
- Taylor, J. A., O'Brien, J. A. & Yeager, M. (1996).** The cytoplasmic tail of NSP4, the endoplasmic reticulum-localized non-structural glycoprotein of rotavirus, contains distinct virus binding and coiled coil domains. *EMBO Journal* **15**, 4469–4476.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.
- Tian, P., Ball, J. M., Zeng, C. Q.-Y. & Estes, M. K. (1996).** The rotavirus nonstructural glycoprotein NSP4 possesses membrane destabilization activity. *Journal of Virology* **70**, 6973–6981.
- Wu, T. T. & Kabat, E. A. (1970).** An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *Journal of Experimental Medicine* **132**, 211–250.

---

Received 27 February 1997; Accepted 22 April 1997