Epidemiol. Infect. (1995), **114**, 385–391

Prevalence of human astrovirus serotype 4: capsid protein sequence and comparison with other strains

M. M. WILLCOCKS[1], J. B. KURTZ[2], T. W. LEE[2] AND M. J. CARTER[1]*

[1] *School of Biological Sciences, University of Surrey, Guildford GU2 5XH, UK*
[2] *Virology Laboratory, John Radcliffe Hospital, Headington, Oxford OX3 9DU*

(*Accepted 27 October 1994*)

SUMMARY

Astrovirus serotype 4 has increased in relative prevalence in the Oxford, UK area in 1993. The structural gene of human astrovirus serotype 4 has been sequenced and the results indicate that this protein differs substantially from serotypes 1 and 2. In particular, conservation at the C terminus is greatly reduced. However, amino acid substitutions in this region show a strong conservation in character suggesting that structural or functional constraints operate in this region.

INTRODUCTION

Astroviruses are smooth-edged 28 nm particles first observed in the stools of infants with diarrhoea. A proportion of these virus particles displays a characteristic 5 or 6 pointed star motif on their surface from which the group is named [1, 2]. Subsequently astroviruses have been associated with symptomatic infections, mainly gastroenteritis, in the young of many animal species [3]. In humans the use of diagnostic electron microscopy has established astrovirus as a cause of diarrhoea in both the young and the elderly[4, 5]. Most astrovirus infections are considered to be mild, and astrovirus association with symptomatic illness has been made relatively infrequently [3]. However, this general perception of the importance of astroviruses in symptomatic infection may be an underestimate. Diagnosis by electron microscopy (EM) is complicated because virions may be shed in high numbers for only a short time; thus most samples sent for diagnosis contain few particles which may not be above the threshold for detection by EM. Secondly, the major diagnostic feature, the characteristic star motif, is present on only *c.* 10% of particles. Until recently lack of diagnostic reagents has hampered epidemiological studies on these viruses, but molecular biological techniques and improved culture methods have now reduced these difficulties. Findings from the use of these methods suggest that some astrovirus infections may have been overlooked owing to co-infection with parvovirus-like agents or lack of clear structural features [6, 7]. A restrospective survey indicated that 26% of diarrhoeal viruses, classified as parvoviruses by electron microscopy, possessed biophysical properties indicative of astroviruses [8]. Further, an ELISA-based survey in Thailand has indicated that astroviruses may actually be

* Author for correspondence.

the second most common cause of symptomatic illness in children after rotaviruses, and more common than adenoviruses in this respect [9].

Most surveys have concentrated on infection amongst children, because these are thought to be at greatest risk, and because seroprevalence, measured by seropositivity to the type 1 virus, rises rapidly in the first 5 years of life, eventually achieving levels of approximately 87% between 5 and 10 years of age [10]. Illness amongst adults is still generally considered relatively rare. Seven serotypes of human astrovirus are now recognized [11], but the prevalence of each serotype is different; in Europe and the USA serotype 1 is most common, whilst in China serotype 4 is most prevalent (W. D. Cubitt, personal communication). Further, the frequency of the other serotypes varies from year to year, with type 4 being in second or third place and accounting for *c.* 12% of identifications per year. Thus, although most adults have been exposed to serotype 1 in childhood, exposure to the other serotypes may be delayed, and in this instance cross-serotype protection from prior exposure may not be sufficient to prevent infection or to eliminate symptoms. For example, some 4700 young adults were infected in a 5 day period in a recent food-borne outbreak of type 6 astrovirus in Japan; 10% were sufficiently ill as to be unable to work [12]. Furthermore, serotype 4 astrovirus has been associated with severe infections in adults, necessitating several days absence from work (W. D. Cubitt, personal communication). Recently serological evidence of exposure to this serotype has been found in 93% of adult coastal windsurfers tested, compared with only 22% in non sea-bathing controls; all these seropositive surfers had reported diarrhoeal illness in the previous 12 months [13]. This suggests that serotype 4 is currently present in coastal waters. The probable source of this virus is sewage effluent, but this in turn implies that serotype 4 infections generally should be more common in recent times. In view of this we wished to re-assess the prevalence of astrovirus serotype 4 in the UK, and assess its relatedness to the serotype 1 virus, to which adults may be assumed to have had previous exposure, but which failed to prevent serotype 4 infection.

MATERIALS AND METHODS

Serotyping of viruses in stool samples

The prevalence of astrovirus serotypes in Oxford, UK during 1993 was determined by solid phase immune electron microscopy of stool specimens submitted for diagnosis, using type-specific rabbit antiserum prepared against astrovirus serotypes 1–5 [14].

Sequence determination

The human astrovirus serotype 4 (HAst-4) used in the sequence studies was isolated in Oxford, UK in 1971 [15]. This virus had been isolated in primary human embryo cells and adapted to culture in continuous cell lines [16]. HAst-4 was grown for these studies in CaCo-2 cells [17] and RNA was extracted for cloning from infected cell cytoplasm [18]. cDNA was synthesized by conventional procedures using the reverse primer given below as primer for reverse transcription. The structural gene was then amplified by PCR using an additional primer derived from the 5′ end of the gene. Both primers were chosen from a

Table 1. *Astrovirus serotype isolations in the Oxford area 1993–4*

*Serotype**	January–June 1993	July–December 1993	Total
1	10	1	11
2	0	0	0
3	9 (6)	0	9 (6)
4	3	11 (8)	14 (11)
5	0	0	0

* Astrovirus isolations made in Oxford UK were serotyped by IEM. Serotype designations shown in italics. Numbers in parentheses discount hospital contacts who may have been cross-infected with the astrovirus whilst on the ward.

comparison of the sequences of serotype 1 and 2 to lie in conserved regions flanking the structural gene [19]. Sequences of the primers used are given below:

Reverse transcription primer 5′-TTTGCTTCTGATTAAATCAATTT-3′
Forward primer 5′-GAAGTGTGATGGCTAGCAAGTC-3′

Amplification was performed using Taq DNA polymerase (Boehringer Mannheim) by 35 cycles of the following conditions: 94 °C, 45 s; 55 °C, 45 s; 72 °C, 4 min. A final extension time of 72 °C for 5 min was added at the end of the amplification. Amplification products were analysed on agarose gels to assess the specificity of the reaction. A prominent single band was usually obtained which could be sequenced directly after isopropanol precipitation [20]. Sequencing commenced using the original PCR primers and was continued using synthetic oligonucleotides synthesized from the sequence previously determined. The sequence presented here was thus determined in both orientations using uncloned, PCR-amplified DNA. All sequencing was performed on an Applied Biosystems model 373A automatic DNA sequencer using Taq DNA polymerase.

RESULTS AND DISCUSSION

The prevalence of astrovirus serotypes in the Oxford UK area has been determined by immune electron microscopy [11]. New data obtained over the last year are presented in Table 1. Isolations of serotype 4 were second only to those of serotype 1 during 1993–4, and were actually the most common astrovirus serotype detected during one of the 6-month periods. Relatively few symptomatic astrovirus infections were detected resulting in a small sample size. Few laboratories in the UK conduct this sort of typing analysis but preliminary data support similar observations elsewhere in the UK. These observations agree with the predictions made from the data of Myint and colleagues [13], that serotype 4 infections have recently risen in relative prevalence and that this has been reflected in increased exposure of adults to this serotype through contamination of coastal waters.

Recently several astroviruses have been cloned, and the complete genomic sequences of types 1 and 2 have been determined [21, 22]. The virus capsid protein is encoded in a separate open reading frame at the 3′ end [19, 23]. Translation yields an unstable precursor of 90 kDa [24] which is processed to form the 3 [17], or 4 [25] structural proteins of the virus. Comparison of this gene between the

```
HAST1CAPS   MASKSNKQVTVEVSNNGRSRSKSRARSQSRGRDKSVKITVNSRNRARRQPGRDKR    55
HAST2CAPS   MASKSDKQVTVEVNNNGRNRSKSRARSQSRGRGRSVKITVNSHNKGRRQNGRNKY    55
HAST4CAPS   MASKSHKEVTVEVSSTGRSRRKSRARSQSRGRGRSSNITGSPHNKGRRQNGRTKY    55
            *****s*s*****sss**s*s**********ss* s** sss*ss*** **s*

HAST1CAPS   QSSQRVRNIVNKQLRKQGVTGPKPAICQRATATLGTVGSNTSGTTEIEACILLNP   110
HAST2CAPS   QSNQRVRKIVNKQLRKQGVTGPKPAICQRATATLGTIGSNTTGATEIEACILLNP   110
HAST4CAPS   QSNQRVRKIVNKQLRKQGVTGPKPAICQTATATLGTIGSNTTGATEIEACILLNP   110
            **s****s****************** *******s****s*s**********

HAST1CAPS   VLVKDATGSTQFGPVQALGAQYSMWKLKYLNVKLTSMVGASAVNGTVSGVSLNPT   165
HAST2CAPS   VLVKDATGSTQFGPVQALGAQYSMWKLKYLNVKLTSMVGASAVNGTVLRISLNPT   165
HAST4CAPS   VLVKDATGSTQFGPVPTLGAQYSIWKLKYLNVRLTSMVGASAVNGTVVRISLNPT   165
            ***************ss******s********s**************  s*****

HAST1CAPS   TTPTSTSWSGLGARKHLDVTVGKNATFKLKPSDLGGPRDGWWLTNTNDNASDTLG   220
HAST2CAPS   STPSSTSWSGLGARKHMDVTVGRNAVFKLRPSDLGGPRDGWWLTNTNDNASDTLG   220
HAST4CAPS   STPSSTSWSGLGARKHLDVTVGKNAAFKLKPSDLGGPRDGWWLTNTNDNASDTLG   220
            s**s************s*****s**s***s************************

HAST1CAPS   PSIEIHTLGRTMSSYKNEQFTGGLFLVELASEWCFTGYAANPNLVNLVKSTDNQV   275
HAST2CAPS   PSIEIHTLGKTMSSYKNEQFTGGLFLVELASEWCFTGYAANPNLVNLVKSTDHEV   275
HAST4CAPS   PSIEIHTLGQTMSSYQNTQFTGGLFLVELSSAWCFTGYAANPNLVNLVKSTARSV   275
            *********s*****s*s***********s*s*******************ss *

HAST1CAPS   SVTFEGSAGSPLIMNVPEGSHFARTVLARSTTPTTLARAGERTTSDTVWQVLNTA   330
HAST2CAPS   NVTFEGSKGTPLIMNVAEHSHFARMAEQHSSISTTFSRAGGDATSDTVWQVLNTA   330
HAST4CAPS   DVTFEGSAGTPLIMNVPEHSHFARMAVEHSSLSTSLSRAGGESSSDNCLQVLNTA   330
            s****** *s******s* ***** s ss*s s*sss***s ss**s  ******

HAST1CAPS   VSAAELVTPPPFNWLVKGGWWFVKLIAGRTRTGSRSFYVYPSYQDALSNKPALCT   385
HAST2CAPS   VSAAELVAPPPFNWLIKGGWWFVKLIAGRTRTGTKQFYVYPSYQDALSNKPALCT   385
HAST4CAPS   VSAAELVTPPPFNWLVKGGWWFVKLIAGRARTGARRFYVYPSYQDALSNKPALCT   385
            *******s*******s**************s***ss ******************

HAST1CAPS   GSTPGGMRTRNPVTTTLQFTQMNQPSLGHGEAPAAFGRSIPAPGEEFKVVLTFGA   440
HAST2CAPS   GGVTGGVLRTTPVTT-LQFTQMNQPSLGHGEHTATIGSIVQDPSGELRVLLTVGS   439
HAST4CAPS   GGVSAYTRQSNPVRTTLQFTQMNQPSLGHGTAPATLGRSVPEPGDQFKVIMTVGA   440
            *ssss     s** * ***************s s*ss*s sss*ssss*ss* *s

HAST1CAPS   PMSPNANNKQTWVNKPLDAPSGHYN---VKIAKDVDHYLTMQGFTSIASVDWYTI   492
HAST2CAPS   IMSPNSADRQVWLNKTLTAPGTNSNDNLVKIAHDLGHYLIMQGFMHIKTVEWYTP   494
HAST4CAPS   LVQPNRSDTQNWLFTTVTPPTGH---DAARVGWNTQHYLTIQGFLLIDSLEWLTP   492
             s ** ss * *s sssss*sss     ssss s  ***ss***  * sss* *

HAST1CAPS   DFQPSEAPAPIQGLQVLVNSSKKADVYAIKQFVTAQTNNKHQVTSLFLVKVTTGF   547
HAST2CAPS   DFQPSRDPTPIAGMSVMVNITKKADVYFMKQFKNSYTNNRHQITSIFLIKPLADF   549
HAST4CAPS   NLQESQEPPSIPELGVYVGIHKKALVYFMQQYVNPHTNNKHQVSSIFLIKPTENF   547
            ss* * s*ss*sss * *s  *** ** ss*s ss ***s**ss*s**s*  ss*

HAST1CAPS   QVNNYLSYFYRASATGDATTNLLVRGDTYTAGISFTQGGWYLLTNTSIVDGAMPP   602
HAST2CAPS   KVQCYMSYFKRESHDNDGVANLTVRSMTSPETIRFQVGEWYLLTSTTLKENNLPE   604
HAST4CAPS   SVTNYMSYFFRESQSGQNVANLKIRPQTWQQTVNFQRGKWYLVTNTAIRNGPPPS   602
             *  *s*** *s* ssssss** s*  *  ssss*  * ***s*s*ss ss  *

HAST1CAPS   GWVWNNVELKTNTAYHMDKGLVHLIMPLPESTQMCYEMLTSIPRSRASGHG----   653
HAST2CAPS   GWVWDRVELKSDTPYYADQALTYFITPPPVDSQILFEGNTTLPRISSPPDNPSGR   659
HAST4CAPS   GWVWDNIELTNESIYYADQVLAHFINPPPQNSKIYFEVHTTMPQSSV-------R   650
            ****sss**ssss *s *s *sss* * * ssss s*  *ss*s s

HAST1CAPS   Y---ESDNTEYLDAPDSADQFKEDIETDTDIESTEDEDEADRFDIID--TSDEED   703
HAST2CAPS   YMESHQQDCDSSDDEDDCENVSEETETEDE-EDEDEDDEADRFDLHSPYSSEPED   713
HAST4CAPS   NIGLEEDQTDNWQEPD------EDVQTSTE----ESDYETDSLE-----GESDDE   690
                s ss s  ss *      *sss*sss    sss *s*sss     sss ss

HAST1CAPS   ENETDRVTLLSTLVNQGMTMTRATRIARRAFPTL----SDRIKRGVYMDLLVSGA   754
HAST2CAPS   SDENNRVTLLSTLINQGMTVERATRITKRAFPTC----AEKLKRSVYMDLLASGA   764
HAST4CAPS   DSNTCRELVINTLVNQGISRERATYIGMSAYPNVEWGSGEQSTSQHIQEISSDDV   745
            ssss *  sss**s***ss s*** *sss*s*s     sss ss    ss  sss

HAST1CAPS   SPGNAWSHACEEARKAAGEINPCTSGSRGHAE   786
HAST2CAPS   SPSSAWSNACDEARNVGSNQLAKLSGDRGHAE   796
HAST4CAPS   GAGAHYSCVCERKQQSLNQ------GSRGHAE   771
            ssss s* s*s  ss  ss       *s*****
```

Fig. 1. For legend see opposite.

https://doi.org/10.1017/S0950268800058015 Published online by Cambridge University Press

Table 2. *Percentage homology between astrovirus structural proteins of different serotypes**

	Residues 1–415	Residues 416–707	Residues 649–707	Residue 708–terminus†
Serotype 1 v 2	85	42	17	62 (74)
Serotype 1 v 4	81	39	24	31 (72)
Serotype 2 v 4	82	38	16	24 (65)

* In order to compensate for slightly differing lengths between the proteins analysed, all 3 were first aligned using the multiple alignment program *CLUSTAL* in the PC Gene software package (Intelligenetics). The regions described in [19] were then delineated with reference to their positions in the serotype 1 protein presented in that reference. The proteins were then aligned pair-wise using the *PALIGN* program, and residues identified above, at the boundaries of the conserved/variable regions, were marked. Homology scores were then determined within each region and expressed as a percentage of the total number of residues in the longer sequence of that area.

† Figures in parentheses indicate percentage residue similarity (identical residues plus those conserved in character).

serotypes has revealed a remarkable degree of variation in the structural protein sequences; sections in the centre of the gene are almost totally different between serotypes, and whilst the termini are well conserved [19]. Cross-hybridization studies have indicated that of the first 5 serotypes, this variation is most marked in serotype 4 [18]. An immuno-reactive epitope has been located to the variable region suggesting that this area is exposed on the surface of the virion [26].

The sequence of the structural protein of HAst-4 is shown in Fig. 1, aligned with the 2 other serotypes for which this information is available. Overall, the serotype 4 capsid protein is shorter than the others; 771 residues compared with 786 and 790 for types 1 and 2. Serotype 4 capsid protein shares 60% homology with HAst-1 [27], and 62% with HAst-2 [23]. In all cases the N terminus of the protein is very well conserved, and as reported previously the central region is more variable. However, this comparison of type 1 and 2 viruses also identified high homology towards the C terminus [19]. In that report we considered the type 1 virus protein in 3 sections: a conserved N terminus (amino acids 1–415); a variable central region (416–707) which includes a hypervariable section from 649–707; and a conserved C terminus (708–786). We have maintained this division in the analysis of serotype 4 and Table 2 presents data which compare each of the proteins in these regions relative to serotype 1. The most striking feature is the excellent conservation of the N terminus section of the protein. All serotypes have > 80% identical residues in this region. The central section is indeed less well conserved between the viruses with similar scores of identical residues between any combination of two serotypes. This section of the protein also shows greatest divergence in the hydropathic profile, but in each case, the peak of maximum hydrophilicity for the protein occurs in this region: residues 685–690 (serotype 1),

Fig. 1. The sequence of human astrovirus serotype 4 capsid protein is presented aligned with those of serotypes 1 [21] and 2 [22]. HAST1CAPS, serotype 1; HAST2CAPS, serotype 2; HAST4CAPS. serotype 4. Conserved residues are indicated (*), positions at which the character of the amino acid is maintained are indicated (s). The nucleic acid sequence from which this protein was determined has been lodged in the EMBL Database Accession No. Z33883.

https://doi.org/10.1017/S0950268800058015 Published online by Cambridge University Press

687–697 (serotype 2) and 686–691 (serotype 4). The combination of sequence variability between antigenically distinct virus proteins and maximum hydrophilicity combine to suggest that this region is probably on the surface of the virus and subject to antibody attack, as observed in the case of serotypes 1 and 2 [26]. The most distinctive feature of the serotype 4 protein is the decrease in identical residues at the C terminus compared with the other two serotypes; this region has been previously described as 'conserved' [19], and the data reported here thus prompt a re-assessment of this conclusion. Although the 5 amino acids at the exact terminus are conserved, this identity is rapidly lost. However, despite this sequence divergence, most substitutions maintain the character of the amino acids in this region. In fact the totals of identical and similar residues in this section are not significantly different between any of the three viruses (Table 2, figures in parentheses). This implies that some structural constraint is still imposed on this region despite the apparent variability. However, it is not known whether this region is exposed on the surface of the virion, and thus the effect of these changes on the antigenic cross-reactivity of the serotypes cannot be predicted.

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council of the UK. We thank Mr S. Dyer for technical assistance.

REFERENCES

1. Madeley CR, Cosgrove BP. Viruses in infantile gastroenteritis. Lancet 1975; ii: 124.
2. Madeley CR, Cosgrove BP. 28 nm particles in faeces in infantile gastroenteritis. Lancet 1975; ii: 451–2.
3. Willcocks MM, Carter MJ, Madeley CR. Astroviruses. Rev Med Virol 1992; **2**: 97–106.
4. Madeley CR, Cosgrove BP, Bell EJ, Fallon RJ. Stool viruses in babies in Glasgow. 1. Hospital admissions with diarrhoea. J Hyg 1977; **78**: 261–73.
5. Oshiro LS, Haley CE, Roberts RR, *et al*. A 27 nanometre virus isolated during an outbreak of acute infectious nonbacterial gastroenteritis in a convalescent hospital: a possible new serotype. J Infect Dis 1981; **143**: 791–5.
6. Willcocks MM, Carter MJ, Silcock JG, Madeley CR. A dot-blot hybridization procedure for the detection of astrovirus in stool samples. Epidemiol Infect 1991; **107**: 405–10.
7. Silcock JG, Willcocks MM, Codd A, Carter MJ, Record CO. Virus involvement in community acquired diarrhoea examined by stool dot-blot hybridization and electron microscopy. Europ J Gastroent Hepatol 1993; **5**: 601–6.
8. Oliver AR, Phillips AD. An electron microscopical investigation of faecal small round viruses. J Med Virol 1988; **24**: 211–18.
9. Herrmann JE, Taylor DN, Echeverria P, Blacklow NR. Astroviruses as a cause of gastroenteritis in children. N Eng J Med 1991; **324**: 1757–60.
10. Kurtz JB, Lee TW. Astroviruses: human and animal. In: Bock G, Whelan J, eds. Novel diarrhoea viruses. CIBA Foundation Symposium 128. Chichester: John Wiley and Sons, 1987: 92–107.
11. Lee TW, Kurtz JB. Prevalence of human astrovirus serotypes in the Oxford region 1976–1992, with evidence for two new serotypes. Epidemiol Infect 1994; **112**: 187–93.
12. Oishi I, Yamazaki K, Kimoto T, *et al*. A large outbreak of acute gastroenteritis associated with astrovirus among students and teachers in Osaka, Japan. J Infect Dis 1994; **170**: 439–43.
13. Myint S, Manley R, Cubitt WD. Viruses in bathing waters. Lancet 1994; **343**: 1640–1.
14. Lee TW, Kurtz JB. Human astrovirus serotypes. J Hyg 1982; **89**: 539–40.
15. Kurtz JB, Lee TW. Human astrovirus serotypes. Lancet 1984; ii: 1405.

16. Lee TW, Kurtz JB. Serial propagation of astrovirus in tissue culture with the aid of trypsin. J Gen Virol 1981; **57**: 421–4.
17. Willcocks MM, Carter MJ, Laidler FR, Madeley CR. Growth and characterization of human faecal astrovirus in a continuous cell line. Arch Virol 1990; **113**: 73–82.
18. Willcocks MM, Carter MJ. The 3′ terminal sequence of a human astrovirus. Arch Virol 1992; **124**: 279–89.
19. Willcocks MM, Carter MJ. Identification and sequence determination of the capsid protein gene of human astrovirus serotype 1. FEMS Micro Lett 1993; **114**: 1–8.
20. Brow MAD. Sequencing with Taq DNA polymerase. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. PCR protocols. A guide to methods and applications. San Diego: Academic Press, 1990: 189–96.
21. Willcocks MM, Madeley CR, Brown TDK, Carter MJ. The complete sequence of a human astrovirus. J Gen Virol 1994; **75**: 1785–8.
22. Jiang B, Monroe SS, Koonin EV, Stine SE, Glass RI. RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frame-shifting signal that directs the viral replicase synthesis. Proc Natl Acad Sci USA 1993; **90**: 10539–43.
23. Monroe SS, Jiang B, Stine SE, Koopmans M, Glass RI. Subgenomic RNA sequence of human astrovirus supports classification of Astroviridae as a new family of RNA viruses. J Virol 1993; **67**: 3611–14.
24. Monroe SS, Stine SE, Gorelkin L, Herrman JE, Blacklow NR, Glass RI. Temporal synthesis of proteins and RNAs during human astrovirus infection of cultured cells. J Virol 1991; **65**: 641–48.
25. Kurtz JB. Astroviruses. In: Farthing MJG, ed. Viruses and the gut. Proceedings of the ninth BSG, Smith Kline and French International Workshop, 1989: 84–7.
26. Sanchez-Fauquier A, Carrascosa AL, Carrascosa JL, *et al.* Characterisation of a human astrovirus serotype 2 structural protein (VP26) that contains an epitope involved in virus neutralization. Virology 1994; **201**: 312–20.
27. Lewis TL, Greenberg HB, Herrmann JE, Smith LS, Matsui SM. Analysis of astrovirus serotype 1 RNA, identification of the viral RNA-dependent RNA polymerase motif, and expression of a viral structural protein. J Virol 1994; **68**: 77–93.