Arch Virol (1996) [Suppl] 12: 277–285

Archives of Virology

The molecular biology of astroviruses

M. J. Carter and **M. M. Willcocks**

School of Biological Sciences, University of Surrey, UK

Summary. Astroviruses (genus *Astrovirus*) are assigned to a newly established virus family, the *Astroviridae*. The molecular biology of these agents reveals many features unique amongst the non-enveloped animal viruses and resembles that of members of certain plant virus families. In particular, their possession of a serine protease and use of ribosomal frameshifting to express the RNA polymerase are similar to the luteoviruses. Many aspects of the astrovirus replication strategy are still unclear, but replication may involve a nuclear step and non-structural proteins may influence host cell range.

Introduction

Astroviruses were first identified in 1975 in the stools of infants with diarrhoea [12]. Particles are 28 nm across with a smooth outer margin. In their centre, a proportion of particles display a prominent 5- or 6-pointed star-like motif. This is unique among viruses and at present there is no satisfactory model for how this structure might be produced. However, this feature led to the recognition of these viruses as distinct from other enteric viruses and also offered an obvious name; the astroviruses (astron [Greek]; a star).

The occurrence of similar particles in the stools of symptom-free children implied that virus infection would be generally associated with mild illness, and hospitalization figures show that documented astrovirus illnesses account for only 2–3% of hospital admissions for diarrhoea [2]. However, the detection of astroviruses still rests largely on morphological criteria, not necessarily a reliable guide. Few laboratories are equipped to conduct more detailed tests and it is likely that the incidence of astrovirus-induced illness has probably been underestimated. Recent surveys, using more objective techniques, have found a much higher incidence of astroviruses in symptomatic illness [4, 21]. While infection is often mild, this is by no means always so and astroviruses have been implicated in severe illnesses in adults.

Immune electron microscopy indicated that astroviruses could be grouped into serotypes; 5 were initially recognised [8, 10], although this number was subsequently extended to 7 [11] and now to 8 [Kurtz; unpublished]. Data from a small survey in the UK suggest that infection is widespread. Seropositivity to

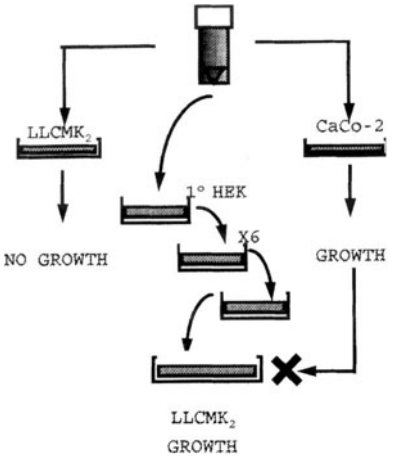

Fig. 1. Diagrammatic representation of the adaptation procedure used to rescue astroviruses in different cell lines. Virus in the stool does not grow in $LLCMK_2$ cells, but can be induced to do so after six passages in primary human embryo kidney cells. Alternatively virus in stool specimens can be grown directly in CaCo-2 cells but does not become adapted to growth in $LLCMK_2$ cells by this process. All techniques require the addition of trypsin to the medium for release of infectious virus to occur

type 1 rises rapidly in childhood, reaching 80% by age 5 [7]. However, infections by other serotypes are not acquired as rapidly and these viruses could cause significant disease in older persons if cross-reactive immunity, induced by a prior infection with type 1 virus, is poor. This may apply particularly to serotype 4, which shows greatest divergence in structural protein sequence from other serotypes.

Astroviruses were first cultivated by adapting them to replicate in cell cultures by serial blind passage in primary human embryo kidney cells with a protease supplement [9]. After six passages, virus released from these cells had acquired the ability to replicate in a continuous monkey kidney cell line, $LLCMK_2$, which cannot support the growth of virus directly from stool specimens (Fig. 1). Each of the five astrovirus serotypes then known was adapted to culture in this way, but all such laboratory-adapted isolates must be host range mutants. In view of the difficulties in performing this adaptation, it was not surprising that the number of such adapted isolates remained few. However, in 1990 astroviruses were cultivated directly from fecal samples in a continuous cell line derived from a human colonic carcinoma (CaCo-2) [20]. Virus grown in CaCo-2 cells does not acquire the ability to grow in $LLCMK_2$ cells (Fig. 1). Other cell lines derived from gut tissue have since been used successfully to cultivate astroviruses, but all require a proteolytic supplement.

Virus replication

Studies of virus replication in both $LLCMK_2$ and CaCo-2 cells have shown that two virus-specific RNAs are induced in infected cells [1, 14]. The first is full-length genomic RNA (6.8 kb), whilst the second is smaller (2.8 kb). Both species are polyadenylated at their 3′ termini. An earlier report identified only a single, full-sized RNA in these cells [18], but this work was conducted late in infection; at this point even cellular RNA had been extensively degraded and

only encapsidated RNA had survived. Thus only genomic RNA was observed. This also implies that astroviruses do not encapsidate their subgenomic RNA.

There are few data on astrovirus protein synthesis. Non-structural proteins are made in low amount and are difficult to detect; only the structural protein precursor has been studied in detail. This is a trypsin-labile protein of approximately Mr 90,000 which reacts with antiserum to purified virions [14]. Virus particles contain several proteins but the consensus opinion is that there are two of approximately Mr 30,000 and a smaller moiety (Mr 27,000 in human astrovirus type 1) which varies in size between serotypes [22]. Together these three proteins account for almost all of the putative 90,000 precursor protein, and this can be shown to be processed to smaller molecules, presumably reflecting maturational cleavage [14]. However, at present the positions of cleavage in the polyprotein are not known. N-terminal sequence analysis has suggested a confusing pattern of overlapping proteins [17], but this may derive from partial processing *in vitro* or unsuitability of the protease supplement (trypsin), which may not reflect accurately the host protease normally exploited by the virus during replication in the gut. At present there is no information on the synthesis and maturation of non-structural proteins.

The astrovirus genome

Astroviruses contain a single-stranded RNA genome (approximately 6.8 kb) of positive sense (Fig. 2). The 3′ terminus is polyadenylated and the 5′ end is believed to be linked covalently to a protein, VPg, although direct evidence for this is lacking. Complete genome sequences are known for three human astroviruses, cell-culture adapted serotypes 1 and 2 [5], and CaCo-2 cell-grown serotype 1 [23], and capsid protein sequences have been determined from 7 of the 8 serotypes (serotype 7 has yet to be examined in detail).

Coding analysis

The RNA contains three sequential open reading frames (ORFs), termed 1a, 1b and 2 (Fig. 2). ORF 1a commences at the initiation codon at position 89 and terminates at position 2890. However, this ORF overlaps that termed 1b by 71 bases. ORF 1b lacks a suitable initiation codon at its 5′ end, the first possible

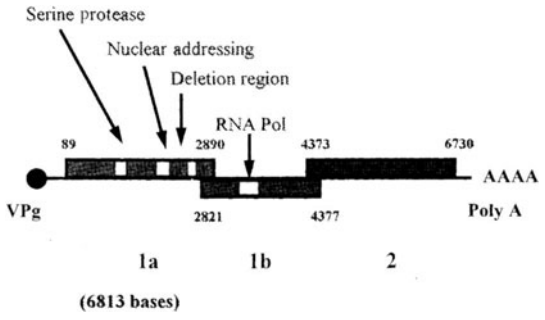

Fig. 2. Astrovirus genome structure. Motifs indicated are described in the text

```
PV-1 VAILPTHASP (26) LEITIITLK (66) GQCGGVITCT-GK-VIGMHVG
RV14 VCVLPTHAQP (26) LELTVLTLD (65) GQCGGVLCAT-GK-IFGIHVG
BWY  ALMTATHVLR (30) GDVTLLRGP (58) GHSGSPYFN--GKTILGVHSG
HAst DIVTAAHVVG (23) KDIAFLTCP (52) GMSGAPVCDKYGR-VLAVHQT
       **  x          x* **         * x*       *  **#*#
```

Fig. 3. Comparison of the astrovirus protease with similar enzymes from other viruses: PV-1 (poliovirus type 1); RV14 (rhinovirus type 14); BWY (Barley western yellows virus, a luteovirus); HAst (human astrovirus serotype 1). Catalytic triad residues are marked (X), Residues conserved in type are marked (*), Binding cleft residues are marked (#)

AUG is in a poor context according to the Kozak consensus, and is not reached until position 3 274, 383 bases downstream from the stop codon of ORF 1a. Sequence analysis in these two ORFs has identified motifs indicative of enzymic functions, as expected for virus non-structural proteins. Astroviruses may be expected to encode a protease for polyprotein maturation, a RNA-dependent RNA polymerase, and a helicase. Analysis has identified both the protease and RNA-dependent RNA polymerase, but there appears to be no helicase-like sequence. This may be a consequence of the relatively small size of the astrovirus RNA; genomes of approximately 6 000 nucleotides may not require this activity [6].

The virus protease has been identified in ORF 1a, centered on the catalytic triad histidine (461), serine (551) aspartic acid (489). A fourth residue (histidine 566) is also conserved and thought to lie in the substrate binding pocket of the enzyme (Fig. 3). This sequence resembles most closely the picornavirus proteases, which are chymotrypsin-like, with the exception that picornaviruses have replaced the active serine residue with cysteine. A potential bipartite nuclear addressing signal has also been detected in ORF1a, and although virus proteins do appear to be moved to the nucleus during replication, the significance of this signal is not known [1]. Sequences towards the carboxy terminal of ORF1a may also determine host cell range (see below). The RNA-dependent RNA polymerase motif has been detected in ORF1b. The sequence YGDD which is thought to lie at the active site of the enzyme is encoded at position 3 940. However, this frame is not expressed from its own initiation codon (see below).

ORF 2 encodes a protein of approximately Mr 90 000 [15, 19, 25]. This is in agreement with a presumptive precursor for the virion proteins identified in the infected cell [14]. Sequence comparisons between human astroviruses show that this protein has a well-conserved amino terminus, which is relatively basic. This is reminiscent of some plant viruses in which such a sequence is probably involved in RNA binding. Similar basic regions have been observed in some coronavirus capsids and in hepatitis B core antigen, in which they may perform a similar role.

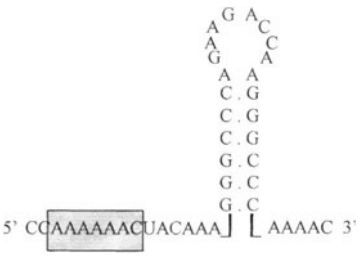

Fig. 4. Predicted secondary structure of the astrovirus frameshifting signal. Residues of the heptanucleotide "slippery" sequence are boxed

Genome expression

Only two species of RNA have been detected intracellularly, the largest, which corresponds to genomic RNA, could act as mRNA for the proteins encoded in ORF 1a. The smaller, subgenomic RNA, is approximately the same size as ORF 2, but there is no candidate mRNA for ORF 1b. However, ORF 1a overlaps ORF 1b by 71 nucleotides, and examination of the region of overlap reveals a heptanucleotide sequence which agrees with the consensus deduced for sites of ribosomal translation frameshift; XXXYYYN, (where X is A, U or G; Y is A or U and N is A, U or C). Such frame-shifts have been detected in a variety of viruses such as corona-, arteri- and retroviruses, but these so called "slippery" sequences act in conjunction with elements of secondary structure; a minimum of a stem-loop and, in many instances, a pseudoknot is also required. Similar features can be identified in this region of the astrovirus RNA (Fig. 4) and mapping studies support the existence of the predicted secondary structure [13]. This region of the RNA has been tested by translation *in vitro* and shown to direct -1 ribosome translational frameshifting at the ORF1a/1b junction. Approximately 5% of ribosomes translating this RNA would slip frame and thus avoid termination at the end of ORF 1a. This would form a 1a/1b fusion protein including sequences from the RNA-dependent RNA polymerase. This system was used to investigate the minimum requirements for ribosomal frameshifting by site-directed mutagenesis; the heptanucleotide sequence is essential; deletion or modification from the XXXYYYN consensus abolishes frame shifting entirely. A stable base-paired stem is also required. Mutations that weaken this secondary structure reduce the efficiency of the frameshift, but those that maintain the strength of the base pairs are tolerated. Thus, a stable stem is needed but its sequence is not critical. Finally, the sequence of the loop is unimportant [13]. This suggests that pseudoknot formation is not involved because these sequences would be required to maintain downstream base pairing interactions. Frameshifting in the astroviruses is thus similar to that seen in animal viruses such as in HTLV-1, but is unique among the non-enveloped animal viruses. The closest similarities in terms of genome organization, presence of a serine protease and expression of the polymerase by means of a translational frameshift are found in non-enveloped plant viruses, luteoviruses.

The subgenomic RNA corresponds to the 3′-terminus of the RNA and includes ORF2 in its entirety [15, 19]. That this RNA could encode the virion structural proteins was confirmed by N-terminal sequencing of virion proteolytic fragments and by expression of this gene in recombinant baculoviruses [15, 19]. ORF 2 from the serotype 1 virus was found to be 300 bases longer than that of the serotype 2 virus sequenced. This raised the possibility that the structural protein made by type 1 virus could be larger than that produced by other serotypes. However, 5′ end mapping of the subgenomic RNA revealed that the mRNAs formed by these two viruses are in fact the same, and the potential extra information is not represented in the mRNA [19]. The actual proteins formed are thus of very similar size between these two serotypes of human astrovirus. Baculovirus-expressed ORF2 protein reacted strongly with antiserum to purified astrovirions [19]. However, although astroviruses can form empty particles [3, 20], virus-like particle assembly was not detected in insect cells. In contrast, most of the protein formed a high molecular weight aggregate. This presumably implies that unlike Norwalk virus, post-translational processing is required before particles can be assembled.

Comparative sequence analysis

Three complete astrovirus genomic sequences have now been reported: two strains of serotype 1 have been sequenced (one isolated by laboratory adaptation in $LLCMK_2$ cells, and the other several years later in CaCo-2 cells), and a single strain of serotype 2 (isolated by laboratory adaptation in $LLCMK_2$ cells). Non-structural genes from all three were very similar except in one region: both isolates adapted to growth in $LLCMK_2$ cells lacked 15 amino acids that were present in virus isolated in CaCo-2 cells. This suggests that the process of host range adaptation could involve selection of deletion mutants from the viruses originally present in stool. This would be best tested by repetition of the adaptation procedure, but this has not been feasible due to the difficulties in working with human fecal material. An alternative would have been to examine the original stool material from which the type 1 laboratory-adapted virus had been derived. These experiments had taken place in Oxford (UK) some 15 years ago, and no such original material remained. However, the stool in question had also been used for volunteer studies. One such volunteer had become ill and a sample of this stool, containing virus removed by a single passage in the human gut from that used for cell culture adaptation, was still available. This was tested by PCR and was found not to contain deletion mutations, even though the adapted virus derived from it did (Fig. 5). Subsequent sequence studies showed that all new isolates, whether grown in CaCo-2 cells or sequenced directly from stool samples, possess 15 amino acids not present in the five laboratory adapted strains. Furthermore, deletion is not selected even after 10 serial passages in CaCo-2 cells (Fig. 6) [24]. It is therefore highly probable that the process of laboratory adaptation to a different host cell can select for this mutation and thus

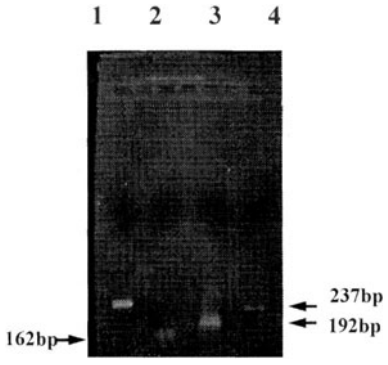

Fig. 5. PCR analysis of deletion in serotype 1 viruses: Track 1, Newcastle CaCo-2 cell isolate A2/88; track2 HPV PCR size marker 162 bp; Track 3, HAst-1 lab adapted; Track 4, HAst-1 from volunteer stool (ie related by one passage in the human gut from the progenitor virus used to obtain the sample analysed in track 3)

```
PEPEVESQPLDLSQKKEKQSEYEQQVVKSTKPQQLEHEQQVVKPIKPQKSEPQPYSQ
HAst-1 (A2/88 Newcastle)
PEPEVESQPLDLSQKKEKQSEY----------------EQQVVKSIKPQKSEPQPYSQ
HAst-1 (Lab adapted)

PEPEVESQPLDLCQKKEKQSEYEQQVVKSIKPQQLEHEQQVVKSIKPQKSEPQPYSQ
HAst-1(Volunteer)

PEPETETQPLDLSQKKEKQPE----------------HEQQVVKSTKPQKNEPQPYSQ
HAst-2 (Lab adapted)
PELEAEAQPLDLSQKKEKQPE----------------HEQQVMKPTKPQKSEPQPYSQ
HAst-3 (Lab adapted)
PEPEAEAQPLDLSQKKEKQPE----------------HEQQVMKPTKPQKSEPQPYSQ
HAst-4 (Lab adapted)
PEPEAETQPLDLSQKKEKQPE----------------HEQQVVKSTKPQKNDPQPYSH
HAst-5 (Lab adapted)

PEPEAESQPLDLSQKKEKQSEYEQQVVKSIKPQQLEHEQQVVKPTKPQKSEPQPYSQ
Serotype 4 (CaCo-2 cellp4)
PEPEAESQPLDLSQKKEKQSEYEQQVVKSIKPQQLEHEQQVVKPTKPQKSEPQPYSQ
Serotype 4 (CaCo-2 cellp10)
```

Fig. 6. Sequence analysis of laboratory adapted and CaCo-2 cell grown astrovirus isolates. Isolate designation is given in the Figure. The deletion is indicated as amino acid sequence and removes one copy of a partially repetitious sequence EQQVVK ··· KPQ

this region of the ORF1a gene may influence the host cell range of the virus. The basis of this effect is not known.

Structural proteins must be responsible for the antigenic variation observed between different astrovirus serotypes. The available sequences were compared, and a variation plot was constructed (Fig. 7). Sequences towards the 5′ end appear well conserved between serotypes, but a region between residues 649 to 707 (positions refer to HAstV-1) is highly variable, and this could contribute to antigenic differences between the viruses. There are numerous insertions and deletions between strains in this area. These could be responsible for the size variation observed between serotypes in the smallest virion polypeptide. This small protein is also thought to be located on the outside of the virus, since it is stripped from the virus by detergent treatment [20]. These observations indicate that the smallest virion protein may be derived from this region and could be responsible for most, if not all, of the antigenic variation between serotypes.

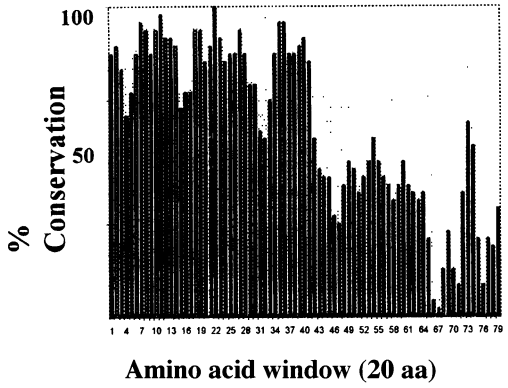

Fig. 7. Comparison of astrovirus structural protein sequences. All available sequences for proteins encoded in ORF2 were aligned using the CLUSTAL program of the PC Gene software package. Alignments were divided into windows of 20 residues, overlapping by 10. Each was scored with respect to positions where substitutions/deletions were permissible. If all amino acids at that position are identical score = 1; if maintained in character score = 0.5; if differing in character or deleted in some viruses score = 0. Maximum score per window = 20, minimum score = 0

Conclusion

Astrovirus is the only genus in a recently recognized family of animal viruses, the *Astroviridae* [16]. Although much remains to be elucidated, concerning their replication strategy, these agents of disease have already revealed features that make them unique among the non-enveloped animal viruses. Similarly progress in the molecular biology of these agents has allowed the production of novel reagents for diagnosis and has challenged our view of the significance of these agents in human disease.

References

1. Carter MJ (1994) Genomic organization and expression of astroviruses and caliciviruses. In: Proceedings of the Third International Symposium on Positive Strand RNA Viruses. Arch Virol [Suppl 9] 429–439
2. Centers for Disease Control (1990) Viral agents of gastroenteritis: public health importance and outbreak management. Morbidity and Mortality Weekly Report 39 (No RR-5): 5–6
3. Herring AJ, Gray EW, Snodgrass DR (1981) Purification and characterisation of ovine astrovirus. J Gen Virol 53: 47–55
4. Herrmann JE, Taylor DN, Echeverria P, Blacklow NR (1991) Astroviruses as a cause of gastroenteritis in children. N Engl J Med 324: 1757–1760
5. Jiang BM, Monroe SS, Koonin EV, Stine SE, Glass RI (1993) RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. Proc Natl Acad Sci USA 90: 10539–10543
6. Koonin EV (1991) Similarities in RNA helicases. Nature 352: 290
7. Kurtz JB, Lee TW (1978) Astrovirus gastroenteritis: Age distribution of antibody. Med Microbiol Immunol 166: 227–230

8. Kurtz JB, Lee TW (1984) Human astrovirus serotypes. Lancet ii: 1405
9. Lee TW, Kurtz JB (1981) Serial propagation of astrovirus in tissue culture with the aid of trypsin. J Gen Virol 57: 421–424
10. Lee TW, Kurtz JB (1982) Human astrovirus serotypes. J Hyg 89: 539–540
11. Lee TW, Kurtz JB (1994) Prevalence of human astrovirus serotypes in the Oxford region 1976–1992, with evidence for two new serotypes. Epidemiol Infect 112: 187–193
12. Madeley CR, Cosgrove BP (1975) 28 nm particles in faeces in infantile gastroenteritis. Lancet ii: 451–452
13. Marczinke B, Bloys AJ, Brown TDK, Willcocks MM, Carter MJ, Brierley I (1994) The human astrovirus RNA-dependent RNA polymerase coding region is expressed by ribosomal frameshifting. J Virol 68: 5588–5595
14. Monroe SS, Stine SE, Gorelkin L, Herrman JE, Blacklow NR, Glass RI (1991) Temporal synthesis of proteins and RNAs during human astrovirus infection in cultured cells. J Virol 65: 641–648
15. Monroe SS, Jiang B, Stine SE, Koopmans M, Glass RI (1993) Subgenomic RNA sequence of human astrovirus supports classification of *Astroviridae* as a new family of RNA viruses. J Virol 65: 67: 3611–3614
16. Monroe SS, Carter MJ, Herrmann JE, Kurtz JB, Matsui SM (1995) Astroviridae. In: Virus Taxonomy. Springer, Wien New York, pp 364–367
17. Sanchez-Fauquier A, Carrasoca AL, Carascosa JL, Overo A, Glass RI, Lopez JA, San Martin C, and Melero JA (1994) Characterisation of a human astrovirus serotype 2 structural protein (VP26) that contains an epitope involved in virus neutralization. Virology 201: 312–320
18. Willcocks MM, Carter MJ (1992) The 3′ terminal sequence of a human astrovirus. Arch Virol 124: 279–289
19. Willcocks MM, Carter MJ (1993) Identification and sequence determination of the capsid protein gene of a human astrovirus serotype 1. FEMS Micro Lett 114: 1–8
20. Willcocks MM, Carter MJ, Laidler F, Madeley CR (1990) Growth and characterisation of a human astrovirus in a continuous cell line. Arch Virol 113: 1–8
21. Willcocks MM, Carter MJ, Silcock JG, Madeley CR (1991) A dot-blot hybridization procedure for the detection of astrovirus in stool samples. Epidemiol Infect 107: 405–410
22. Willcocks MM, Carter MJ, Madeley CR (1992) Astroviruses. Rev Med Virol 2: 97–106
23. Willcocks MM, Brown TDK, Madeley CR, Carter MJ (1994) The complete sequence of a human astrovirus. J Gen Virol 75: 1785–1788
24. Willcocks MM, Ashton N, Kurtz JB, Cubitt WD, Carter MJ (1994) Cell culture adaptation of astrovirus involves a deletion. J Virol 68: 6057–6058
25. Willcocks MM, Kurtz JB, Lee TW, Carter MJ (1995) Prevalence of human astrovirus serotype 4: capsid protein sequence and comparison with other strains. Epidemiol Infect 114: 385–391

Authors' address: Dr. M. J. Carter, School of Biological Sciences, University of Surrey, Surrey GU2 5XH, UK.