Identification and sequence determination of the capsid protein gene of human astrovirus serotype 1

Margaret M. Willcocks and Michael J. Carter *

Division of Virology, School of Pathological Sciences, The Medical School, University of Newcastle-upon-Tyne, NE2 4HH, UK

Abstract: We present the sequence of an open reading frame (ORF) at the 3′ end of human astrovirus serotype 1. Primer extension experiments showed that the RNA expressing this gene is shorter than the complete ORF, and could form a protein of M_r 85540. The protein was expressed by recombinant baculovirus and was recognized by anti-virion serum, indicating a structural role. Sequence comparison indicates that astrovirus serotypes 1 and 2 differ markedly in the C-terminal half of the protein but are well conserved towards the N-terminus.

Introduction

Astroviruses are spherical, non-enveloped positive strand RNA viruses first observed in the faeces of human infants with diarrhoea [1]. The 28 nm particles have a smooth margin and a proportion display a characteristic five or six pointed star motif on their surfaces (from which the viruses are named). Astroviruses have since been found in a variety of hosts and are often associated with gastrointestinal symptoms [2]. Early difficulties in the culture of these viruses impeded their study, however several isolates were adapted to growth in tissue culture by passage through primary human embryo cells [3]. This requirement limited the isolations made in the past, but now viruses can be grown directly from diarrhoeal stools [4]. Molecular biological experiments have shown that astroviruses resemble the Caliciviridae in their synthesis of an intracellular subgenomic RNA from the 3′ end of the genome [5,6], but differ from this family in possessing 3–4 structural proteins, a property in which they resemble the Picornaviridae. Furthermore, in contrast to the picornaviruses, it is possible that virus replication may involve the host cell nucleus [2]. These features led to the recognition of the astroviruses as a distinct family, the Astroviridae [7].

* Corresponding author. Fax (091) 261 1182. e-mail m.j.carter@newcastle.ac.uk.

We have reported the sequence of the 3' terminus from a Newcastle serotype 1 astrovirus and found that this region cross-hybridized to a variable extent with other serotypes [8]. This suggested that serotypic variation could reside in this area, and since typical non-structural protein motifs were absent, that gene order could differ between picorna- and astroviruses; astrovirus structural genes being located at the 3' end. Recently Monroe and co-workers [5] described the synthesis of a protein (M_r 90000) in astrovirus serotype 2-infected cells which could be processed by trypsin to yield products resembling virion proteins. Finally the sequence of the astrovirus type 2 subgenomic RNA has been reported [9]. A single open reading frame was present which could encode a protein of M_r 88000 which was assumed to represent the 90000 protein reported previously.

We have sequenced the complete open reading frame at the 3' end of the astrovirus genomic RNA from which this subgenomic RNA is derived. We confirm the general location of the 5' end of the mRNA but find that the open reading frame in the genome extends beyond the 5' terminus of the mRNA.

Materials and Methods

Astrovirus serotype 1, isolate A88/2 (Newcastle), was grown in $CaCO_2$ cells using a stool extract as inoculum [4] and RNA was harvested during this first passage. cDNA was synthesized in a primer extension reaction using the oligonucleotide 5'-GATGGCATACACATCA-3' (termed primer 1) which hybridized 878–893 bases from the 3' end of the genome determined from our previous sequence [8]. cDNA was synthesized by reverse transcription and converted to a double-strand with DNA polymerase 1. Any ragged ends were repaired with T4 DNA polymerase before the cDNA was ligated to BstXI linkers and ligated to the plasmid vector pcDNA II (Invitrogen) cut with the same enzyme. The longest clone obtained was termed p22A8. This and other clones were then sequenced using a Sequenase reaction kit (United States Biochemicals). The 5'

end of the subgenomic RNA was determined by primer extension [10] using a primer derived from the sequence of clone p22A8, 5'-ACTGTGATC-TTGACTG-3' (termed primer 2). This annealed 2322–2337 bases from the 3' end of the assembled sequence.

In order to express the entire 3' ORF in baculoviruses it was necessary to join the new clone p22A8 to that containing the genome 3' terminus p21T43. This was achieved using the Pst I site present in the area of overlap of the two clones. The region expressed in the subgenomic RNA was then selectively amplified by using a synthetic primer, 5'-GAAGTGTGATGGCTAG-CAAGTC-3' (termed primer 3) which annealed 2427–2448 bases from the virus 3' terminus and primed DNA synthesis towards the 3' end. The reverse primer was annealed to vector sequences beyond the 3' end of the inserted virus sequence. Primer 3 was chosen to preserve the Kozak environment of the initiation codon present in the subgenomic mRNA [11]. The positions of the primers used in the three aspects of this work: cloning, primer extension and amplification, are illustrated in Fig. 1 with reference to the virus-specific RNAs.
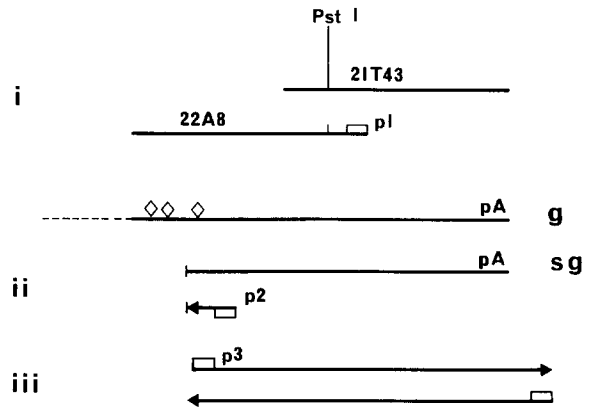

Fig. 1. Diagrammatic representation of the primers (open boxes) used in: (i) cloning; (ii) subgenomic RNA 5' end mapping and (iii) DNA amplification for protein expression. Genomic (g) and subgenomic (sg) RNAs are shown co-terminal and polyadenylated at their 3' ends and the positions of the three potential initiation codons in ORF1 are indicated (◇). Primers are numbered P1, P2 and P3 as referred to in the text.

The amplification product was cloned into the vector pTZ18R at the *Sma*I site and sequenced, before transfer to the baculovirus shuttle vector pVL1392. The construct was combined with *Sau*I linearized baculovirus (Lac-z) DNA for co-transfection into Sf21 cells using the methods described by King and Possee [12]. Virus yield was harvested, plaqued on Sf21 monolayers and stained with X-gal. White plaques were picked and tested for the presence of astrovirus sequences by hybridization.

Expressed proteins were analysed by immuno-precipitation. Sf21 cells were labelled with 0.5 ml of Tris-buffered saline (pH 6.6) containing 20 μCi/ml of ^{35}S-Trans label (ICN-Flow) for 1 h. Cells were lysed in RIPA buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.4, 0.1% SDS, 1.0% Triton X-100, 1.0% sodium deoxycholate) containing protease inhibitors, and the nuclei were removed by sedimentation. Virus-specific proteins were immunoprecipitated [13] with rabbit serum raised to purified virus (kindly supplied by Mr. T. Lee, John Radcliffe Hospital, Oxford) and analysed by electrophoresis.

```
ATGTCCTTTTGTGGCTTCACTGTTAATGAAAACCTTGAGCCTGTGCCAACTTCTCCTGA - 60
 L  S  F  C  G  F  T  V  N  E  N  L  E  P  V  P  T  S  P  E

AAAGTTAATGGCATCACTGTTAAAACCTTACAAAATATTACCTGATCTTGAATCACTCCA - 120
 K  L  M  A  S  L  L  K  P  Y  K  I  L  P  D  L  E  S  L  H

TGGGAAACTCCTATGCTATCAGTTGCTTGCTGCGTTCATGGCAGAGGACCACCCCTTTAA - 180
 G  K  L  L  C  Y  Q  L  L  A  A  F  M  A  E  D  H  P  F  K

GGTGTATGTGGAGCACTGCCTCTCGCGGACTGCAAAGCAGCTTCGTGACTCTGGCCTTCC - 240
 V  Y  V  E  H  C  L  S  R  T  A  K  Q  L  R  D  S  G  L  P
                              *     *      *     *
AGCCAGACTCACAGAAGAGCAACTCCATCGGCATTTGGAGGGGAGGACCAAAGAAGTGTGA - 300
 A  R  L  T  E  E  Q  L  H  R  I  W  R  G  G  P  K  K  C  D

TGGCTAGCAAGTCCAACAAGCAAGTAACTGTTGAGGTCAGTAATAATGGCCGCAGCAGGA - 360
 G  -
 M  A  S  K  S  N  K  Q  V  T  V  E  V  S  N  N  G  R  S  R

GTAAATCAAGGGCCCGTTCACAATCTAGGGGCCGAGACAAATCAGTCAAGATCACAGTCA - 420
 S  K  S  R  A  R  S  Q  S  R  G  R  D  K  S  V  K  I  T  V

ATTCAAGAAACAGAGCCAGGAGACAGCCCGGACGCGACAAACGTCAATCTTCTCAACGTG - 480
 N  S  R  N  R  A  R  R  Q  P  G  R  D  K  R  Q  S  S  Q  R

TCCGTAACATTGTCAATAAGCAACTCAGGAAACAGGGTGTCACAGGACCAAAACCTGCAA - 540
 V  R  N  I  V  N  K  Q  L  R  K  Q  G  V  T  G  P  K  P  A

TATGTCAGAGAGCAACAGCCAACCCTTGGGACAGTCGGGTCAAACACCAGTGGCACCACTG - 600
 I  C  Q  R  A  T  A  T  L  G  T  V  G  S  N  T  S  G  T  T

AGATTGAGGCGTGTATTCTCCTCAACCCTGTCCTCGTTAAGGACGCTACTGGAAGCACTC - 660
 E  I  E  A  C  I  L  L  N  P  V  L  V  K  D  A  T  G  S  T

AGTTTGGCCCTGTGCAGGCGCTAGGTGCACAGTATTCCATGTGGAAGTTGAAGTATTTGA - 720
 Q  F  G  P  V  Q  A  L  G  A  Q  Y  S  M  W  K  L  K  Y  L

ATGTCAAATTGACCTCTATGGTTGGTGCATCTGCTGTTAATGGTACTGTCTCAGGGGTCT - 780
 N  V  K  L  T  S  M  V  G  A  S  A  V  N  G  T  V  S  G  V
```

```
CACTTAACCCCACTACCACGCCTACTTCTACTAGTTGGTCAGGATTAGGTGCACGTAAAC - 840
 S  L  N  P  T  T  T  P  T  S  T  S  W  S  G  L  G  A  R  K

ATCTTGATGTTACGGTTGGTAAGAATGCAACATTTAAATTGAAACCTTCTGACCTTGGTG - 900
 H  L  D  V  T  V  G  K  N  A  T  F  K  L  K  P  S  D  L  G

GACCTAGGGATGGCTGGTGGCTCACAAAACACCAACGACAATGCATCTGACACCCTGGGAC - 960
 G  P  R  D  G  W  W  L  T  N  T  N  D  N  A  S  D  T  L  G

CCTCCATTGAAATACACACACTCGGACGAACAATGTCCTCATACAAGAATGAGCAGTTCA - 1020
 P  S  I  E  I  H  T  L  G  R  T  M  S  S  Y  K  N  E  Q  F

CAGGAGGGCTATTTTTGGTTGAACTTGCTTCAGAGTGGTGCTTCACAGGTTATGCTGCCA - 1080
 T  G  G  L  F  L  V  E  L  A  S  E  W  C  F  T  G  Y  A  A

ACCCAAACCTTGTTAACTTGGTCAAGTCAACAGATAATCAAGTATCTGTGACATTTGAAG - 1140
 N  P  N  L  V  N  L  V  K  S  T  D  N  Q  V  S  V  T  F  E

GAAGTGCAGGATCACCATTAATAATGAATGTGCCAGAGGGGAGCCATTTTGCACGAACAG - 1200
 G  S  A  G  S  P  L  I  M  N  V  P  E  G  S  H  F  A  R  T

TTCTTGCACGCTCAACAACACCCACCACTCTAGCGCGTGCAGGAGAGAGAACCACCTCAG - 1260
 V  L  A  R  S  T  T  P  T  T  L  A  R  A  G  E  R  T  T  S

ACACAGTATGGCAGGTGCTCAATACAGCTGTATCTGCTGCTGAGCTTGTCACGCCTCCTC - 1320
 D  T  V  W  Q  V  L  N  T  A  V  S  A  A  E  L  V  T  P  P

CATTCAATTGGTTAGTTAAAGGAGGTTGGTGGTTTGTGAAACTCATTGCCGGGAGAACCA - 1380
 P  F  N  W  L  V  K  G  G  W  W  F  V  K  L  I  A  G  R  T

GAACTGGCTCTCGTAGCTTTTATGTGTATCCCAGCTACCAAGATGCTTTGTCCAACAAGC - 1440
 R  T  G  S  R  S  F  Y  V  Y  P  S  Y  Q  D  A  L  S  N  K

CAGCACTCTGCACCGGCAGCACTCCAGGAGGCATGAGGACGCGGAATCCTGTGACAACCA - 1500
 P  A  L  C  T  G  S  T  P  G  G  M  R  T  R  N  P  V  T  T
```

Fig. 2. Sequence of the astrovirus serotype 1 capsid gene. The sequence determined from clone 22A8 described here is shown assembled to that of clone 21T43 reported previously (EMBL Ac. No. Z11682). Putative translation products are shown below. The incomplete non-structural proteins extend from the 5′ end to terminate at position 305. Structural gene products are shown commencing at the first ATG in the subgenomic message, position 300. Primer extension products defining the 5′ end of the mRNA are illustrated (*). The major termination product is that at position 286.

4

CCCTACAATTCACACAGATGAACCAACCTAGTCTAGGGCACGGTGAAGCACCAGCTGCGT - 1560
 T   L   Q   F   T   Q   M   N   Q   P   S   L   G   H   G   E   A   P   A   A

TTGGTAGATCCATTCCAGCACCTGGTGAGGAGTTTAAAGTCGTCCTCACGTTTGGAGCCC - 1620
 F   G   R   S   I   P   A   P   G   E   E   P   K   V   V   L   T   F   G   A

CAATGAGCCCTAATGCAAATAACAAACAGACTTGGGTTAATAAACCTCTTGATGCGCCTT - 1680
 P   M   S   P   N   A   N   N   K   Q   T   W   V   N   K   P   L   D   A   P

CAGGCCATTACAATGTGAAGATTGCAAAGGATGTTGACCACTATCTAACCATGCAGGGTT - 1740
 S   G   H   Y   N   V   K   I   A   K   D   V   D   H   Y   L   T   M   Q   G

TCACTTCTATAGCTTCTGTTGACTGGTACACCATAGATTTTCAACCATCTGAGGCCCCTG - 1800
 F   T   S   I   A   S   V   D   W   Y   T   I   D   F   Q   P   S   E   A   P

CCCCAATACAAGGCCTGCAGGTACTTGTGAACAGCTCGAAAAAAGCTGATGTGTATGCCA - 1860
 A   P   I   Q   G   L   Q   V   L   V   N   S   S   K   K   A   D   V   Y   A

TCAAACAATTTGTTACAGCGCAGACCAACAACAAGCACCAGGTTACAAGCCTGTTCCTAG - 1920
 I   K   Q   F   V   T   A   Q   T   N   N   K   H   Q   V   T   S   L   F   L

TAAAAGTGACAACTGGTTTTCAGGTGAACAACTACCTGAGTTACTTTTACAGGGCGTCTG - 1980
 V   K   V   T   T   G   F   Q   V   N   N   Y   L   S   Y   F   Y   R   A   S

CCACTGGGGATGCCACAACTAACCTGTTGGTTAGAGGAGACACATACACAGCAGGGATAA - 2040
 A   T   G   D   A   T   T   N   L   L   V   R   G   D   T   Y   T   A   G   I

GTTTTACCCAGGGTGGATGGTATTTGTTGACAAATACATCTATTGTTGATGGGCTATGC - 2100
 S   F   T   Q   G   G   W   Y   L   L   T   N   T   S   I   V   D   G   A   M

CACCTGGTTGGGTCTGGAATAACGTGGAACTTAAAACTAACACAGCATATCACATGGACA - 2160
 P   P   G   W   V   W   N   N   V   E   L   K   T   N   T   A   Y   H   M   D

AAGGCTTGGTCCATTTAATAATGCCTTTGCCTGAGTCCACGCAAATGTGTTATGAGATGC - 2220
 K   G   L   V   H   L   I   M   P   L   P   E   S   T   Q   M   C   Y   E   M

TGACATCTATACCACGCTCCAGGGCATCTGGTCATGGTTATGAGAGTGACAACACCGAAT - 2280
 L   T   S   I   P   R   S   R   A   S   G   H   G   Y   E   S   D   N   T   E

ACTTGGATGCTCCAGACTCTGCTGACCAGTTTAAAGAAGACATAGAGACAGATACAGACA - 2340
 Y   L   D   A   P   D   S   A   D   Q   F   K   E   D   I   E   T   D   T   D

TTGAGAGTACAGAGGACGAAGACGAAGCGGACAGGTTTGACATCATAGACACTTCTGATG - 2400
 I   E   S   T   E   D   E   D   E   A   D   R   F   D   I   I   D   T   S   D

AAGAAGATGAAAATGAGACAGACCGTGTAACCCTCCTCTCAACTCTCGTAAACCAAGGAA - 2460
 E   E   D   E   N   E   T   D   R   V   T   L   L   S   T   L   V   N   Q   G

TGACAATGACGCGTGCCACAAGGATAGCAAGGCGCGCATTCCCCACGCTTTCCGATAGGA - 2520
 M   T   M   T   R   A   T   R   I   A   R   R   A   F   P   T   L   S   D   R

TCAAGCGTGGAGTATACATGGACCTGCTTGTCTCGGGGGCAAGCCCAGGCAATGCATGGT - 2580
 I   K   R   G   V   Y   M   D   L   L   V   S   G   A   S   P   G   N   A   W

CCCATGCGTGTGAAGAAGCACGCAAAGCAGCAGGGGAAATCAATCCCTGCACATCTGGAA - 2640
 S   H   A   C   E   E   A   R   K   A   A   G   E   I   N   P   C   T   S   G

GCCGCGGCCACGCCGAGTAGGAACGAGGGTACAGCTTCCTTCTTTTCTGTCTCTGTTTAG - 2700
 S   R   G   H   A   E   -

ATTATTTTAATCACCATTTAAAATTGATTTAATCAGAAGC(An) - 2740

Fig. 2 (continued).

Results and Discussion

cDNA cloning

Primer extension cloning was performed using the sequence of the 3' end of astrovirus A88/2 Newcastle, reported previously as clone p21T43 [8] (Fig. 1). This led to the production of clone p22A8 (1862 bp). This was sequenced in both orientations and overlapped clone p21T43 as expected. The coding structure of the assembled sequences revealed two overlapping ORFs, one of which was incomplete and extended towards the 5' terminal of the virus. The sequence of this region, and possible translation products are presented (Fig. 2). The open reading frame at the 3' end is complete and could encode a protein of M_r 92 180. Astroviruses produce a subgenomic RNA [5] which is derived from the 3' end [6] and is of the size expected if it were to contain this open reading frame. We had earlier been unable to demonstrate the synthesis of this RNA [8] but have since found it to be synthesized earlier in infection than previously examined [14]. This general structure and strategy are reminiscent of the caliciviruses [15].

Determination of the 5' end of the subgenomic RNA

The open reading frame identified above is longer than the subgenomic RNA reported for a type 2 human astrovirus [5]. Since this ORF contains two potential initiation codons which are both in frame and precede that identified by Monroe [5,9], it was possible that the type 1 virus could synthesize a larger protein if all this information were to be represented in the mRNA. We

therefore mapped the 5' end of the virus subgenomic RNA by primer extension (Fig. 3). A prominent termination product (135n) was observed, co-migrating with base 286 in the sequence. Fainter terminations were obtained 5' to
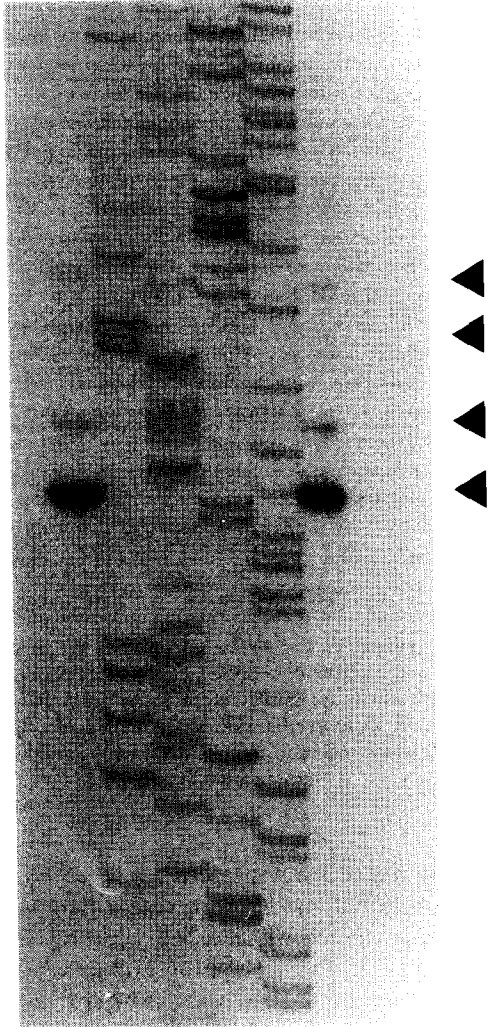

Fig. 3. Primer extension mapping of the subgenomic mRNA 5' end. cDNA Primer extension products were electrophoresed alongside the sequence of clone 22A8 initiating at the same primer. Four potential RNA start sites were observed (marked ▲). The prominent product is 135 bases long and co-migrates with a T in the gel shown and corresponds to base 286 (A) in the positive sense.
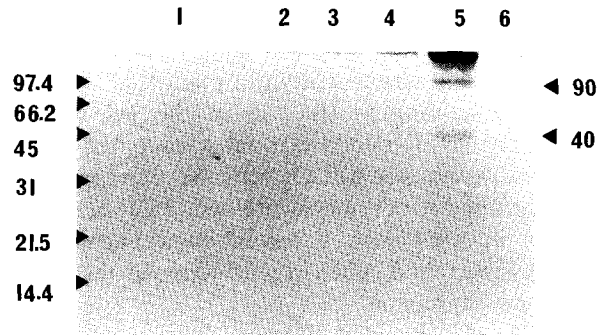

Fig. 4. Immunoprecipitation of astrovirus-specific proteins from insect cells infected with recombinant baculoviruses. Sf21 cells were infected with various baculovirus recombinants as follows. Tracks 2–5, astrovirus recombinant described in the text; track 6, Feline calicivirus recombinant; track 1, mock-infected. Cells were pulse-labelled with [^{35}S]methionine at various times post-infection, processed for lysate preparation as described and immunoprecipitated [18] using anti-astrovirus rabbit serum. Samples were processed at the following times post-infection: track 2, 1 day; track 3, 3 days; track 4, 4 days; tracks 1, 5, 6, 5 days. Resulting samples were separated on a 15% polyacrylamide gel. The positions of molecular mass markers are indicated.

this strong stop; all of these are indicated in Fig. 2. This analysis indicated that the open reading frame present in the genome is longer than the region expressed as subgenomic RNA, and that the first AUG present in the mRNA is actually the third in the open reading frame at position 300 in the sequence presented. This is also the first potential initiation codon in a favourable context as determined by Kozak [11].

Identification of the protein product

The coding region identified above was expressed in baculovirus (above). Figure 4 demonstrates the induction of immunoprecipitable astrovirus-specific protein following Sf21 cell infection with this recombinant. A protein (M_r approx. 90 000) was observed in increasing amounts from 4 days post-infection; a smaller protein (M_r 40 000) was also precipitated concurrently. The 90 000 protein identified by Monroe [5] is thought to be cleaved to yield the structural proteins of the virion and we attribute this smaller species to partial cleavage in this unnatural host cell. Both proteins were selectively immunoprecipitated with

6

rabbit antiserum raised to astrovirus type 1 particles, confirming a relationship between these products and virion structural peptides. The material remaining at the top of the gel is also observed in immunoprecipitates prepared from astrovirus-infected $CaCO_2$ cells. No proteins were precipitated from mock-infected Sf21 cells, or those infected with a recombinant baculovirus expressing feline calicivirus proteins, at the same time post-infection.

Strain comparison

Figure 5 compares the proteins predicted from astrovirus serotype 1 with those from serotype 2. This indicates that the extent of similarity varies in different regions of the proteins. The serotype 2 virus protein is 10 residues longer than that of serotype 1 but both proteins are clearly related at the N terminus. Between residues 1 and 415 (numbers relative to serotype 2 virus) amino acids are 85.6% conserved. This number rises to 90.6% similarity if maintenance of amino acid character is taken into account. Between residues 416 and 707 the sequences are only 41.7% conserved (56.2% if allowance is made for substitutions which maintain the character of the residue). This region includes a highly variant area 649–707 in which only 16.9% of amino acids are conserved and only 28.8% are similar; 8 of the 10 extra residues present in serotype 2 are deleted from serotype 1 in this region. Finally, towards the C terminal conservation and similarity once again rise reaching 61.8% and 74.2%, respectively. Despite these changes in sequence, the hydropathy profiles of the two serotype proteins are very similar (not shown). In particular, a stretch of basic residues is present in both serotypes between residues 18 and 34. A similar sequence is present in some coronavirus capsid proteins [14]. It will be of great interest to compare the sequences of proteins from other astrovirus serotype proteins with that presented here.

The coding organization at the 3′ end of human astrovirus is similar to that of the caliciviruses, in that two partially overlapping open reading frames are present, the most 3′ of which encodes the virion structural polypeptides. The incomplete 5′ open reading frame presumably

```
SEROTYPE 1- MASKSNKQVTVEVSNNGRSRSKSRARSQSRGRDKSVKITVNSRNRARRQP  -50
            ***** ******* **** ************* s******** *s ***
SEROTYPE 2- MASKSDKQVTVEVNNNGRNRSKSRARSQSRGRGRSVKITVNSHNKGRRQN  -50

SEROTYPE 1- GRDKRQSSQRVRNIVNKQLRKQGVTGPKPAICQRATATLGTVGSNTSGTT  -100
            ** * ** **** ************************s*s**s*s*
SEROTYPE 2- GRNKYQSNQRVRKIVNKQLRKQGVTGPKPAICQRATATLGTIGTNTTGAT  -100

SEROTYPE 1- EIEACILLNPVLVKDATGSTQFGPVQALGAQYSMWKLKYLNVKLTSMVGA  -150
            **************************************************
SEROTYPE 2- EIEACILLNPVLVKDATGSTQFGPVQALGAQYSMWKLKYLNVKLTSMVGA  -150

SEROTYPE 1- SAVNGTVSGVSLNPTTTPTSTSWSGLGARKHLDVTVGKNATFKLKPSDLG  -200
            ******* s*****s**s***********s*****s** ***s*****
SEROTYPE 2- SAVNGTVLRISLNPTSTPSSTSWSGLGARKHMDVTVGRNAVFKLRPSDLG  -200

SEROTYPE 1- GPRDGWWLTNTNDNASDTLGPSIEIHTLGRTMSSYKNEQFTGGLFLVELA  -250
            **********************************s*************
SEROTYPE 2- GPRDGWWLTNTNDNASDTLGPSIEIHTLGKTMSSYKNEQFTGGLFLVELA  -250

SEROTYPE 1- SEWCFTGYAANPNLVNLVKSTDNQVSVTFEGSAGSPLIMNVPEGSHFART  -300
            ********************* * ****** *g****** * *****
SEROTYPE 2- SEWCFTGYAANPNLVNLVKSTDHEVNVTFEGSKGTPLIMNVAEHSHFARM  -300

SEROTYPE 1- VLARSTTPTTLARAGERTTSDTVWQVLNTAVSAAELVTPPPFNWLVKGGW  -350
            *s ** s*** g***************************s*******s****
SEROTYPE 2- AEQHSSISTTFSRAGGDATSDTVWQVLNTAVSAAELVAPPPFNWLIKGGW  -350

SEROTYPE 1- WFVKLIAGRTRTGSRSFYVYPSYQDALSNKPALCTGSTPGGMRTRNPVTT  -400
            **************ss ****************** **s ****
SEROTYPE 2- WFVKLIAGRTRTGTKQFYVYPSYQDALSNKPALCTGGVTGGVLRTTPVTT  -400

SEROTYPE 1- TLQFTQMNQPSLGHGEAPAAFGRSIPAPGEEFKVVLTFGAPMSPNANNKQ  -450
            *************** *s * s * * s*s** *s ****s s*
SEROTYPE 2- -LQFTQMNQPSLGHGEHTATIGSIVQDPSGELRVLLTVGSIMSPNSADRQ  -449

SEROTYPE 1- TWVNKPLDAPSGHYN---VKIAKDVDHYLTMQGFTSIASVDWYTIDFQPS  -497
            *s** * ** * **** *s *** **** * s*s*** *****
SEROTYPE 2- VWLNKTLTAPGTNSNDNLVKIAHDLGHYLIMQGFMHIKTVEWYTPDFQPS  -499

SEROTYPE 1- EAPAPIQGLQVLVNSSKKADVYAIKQFVTAQTNNKHQVTSLFLVKVTTGF  -547
            *s** *s *s** s****** s*** s ***s**s**s**s* s *
SEROTYPE 2- RDPTPIAGMSVMVNITKKADVYFMKQFKNSYTNNRHQITSIFLIKPLADF  -549

SEROTYPE 1- QVNNYLSYFYRASATGDATTNLLVRGDTYTAGISFTQGGWYLLTNTSIVD  -597
            *s *s*** * *  * s** ** *  * * ***** *ss s
SEROTYPE 2- KVQCYMSYFKRESHDNDGVANLTVRSMTSPETIRFQVGEWYLLTSTTLKE  -599

SEROTYPE 1- GAMPPGWVWNNVELKTNTAYHMDKGLVHLIMPLPESTQMCYEMLTSIPRS  -647
            s* **** ****s * *  *   * * * s*s s* *ss**
SEROTYPE 2- NNLPEGWVWDRVELKSDTPYYADQALTYFITPPPVDSQILFEGNTTLPRI  -649

SEROTYPE 1- RASGHGYESDNTEYLDAPDSADQFKEDIETDTDIESTEDEDEADRFDIID  -697
            s        *    s   s* *  ss  ****ss *  * *
SEROTYPE 2- SSPPDNPSGRYMESHQQDCDSSDDEDDCENVSEETETEDEEDEDEDDEAD  -699

SEROTYPE 1- --------TSDEEDENETDRVTLLSTLVNQGMTMTRATRIARRAFPTLSD  -739
                    s*s ** * *******s****s *****ss***** ss
SEROTYPE 2- RFDLHSPYSSSEPEDSDENNRVTLLSTLINQGMTVERATRITKRAFPTCAE -749

SEROTYPE 1- RIKRGVYMDLLVSGASPGNAWSHACEEARKAAGEINPCTSGSRGHAE  -786
            ss** ****** ***** *** *s*** ** ***** **
SEROTYPE 2- KLKRSVYMDLLASGASPSSAWSNACDEARNVGSNQLAKLSGDRGHAE  -796
```

Fig. 5. Comparison of the structural proteins of astrovirus serotypes 1 and 2. Proteins reported here and by Monroe [5] were aligned using the PALIGN program from PC-Gene (Intelligenetics). Conserved residues are indicated (*), and substitutions which maintain the character of the residue are designated (s).

encodes non-structural proteins, but no characteristic functional motifs were identified within this short sequence. This open reading frame overlaps that of the structural proteins by 186 bases. This is larger than such overlaps seen in the caliciviruses [14]. However, primer extension studies indicate that not all of the potential open reading frame is actually expressed as mRNA, and the 5′ gene thus overlaps the expressed region of the structural protein open reading frame by only six bases. This presumptive mRNA start position is three bases upstream from that identi-

fied by Monroe [9]. The observation of other primer extension products suggests that the subgenomic RNA may be heterogeneous at its 5′ end as suggested for caliciviruses [16,17].

Acknowledgements

References

1 Madeley, C.R. and Cosgrove, B.P. (1975) Lancet ii, 124.
2 Willcocks, M.M., Carter, M.J. and Madeley, C.R. (1992) Rev. Med. Virol. 2, 97–106.
3 Lee, T.W. and Kurtz, J.B. (1981) J. Gen. Virol. 57, 421–424.
4 Willcocks, M.M., Carter, M.J., Laidler, F.R. and Madeley, C.R. (1990) Arch. Virol. 113, 73–82.
5 Monroe, S.S., Stine, S.E., Gorelkin, L., Herrmann, J.E. Blacklow, N.R. and Glass, R.I. (1991) J. Virol. 65, 641–648.
6 Matsui, S.M., Kim, J.P., Greenberg, H.B., Young, La V.M., Smith, L.S., Lewis, T.L., Herrmann, J.E., Blacklow, N.R., Dupuis, K. and Reyes, G.R. (1993) J. Virol. 67, 1712–1715.
7 Monroe, S.S., Kurtz, J.B., Carter M.J., Herrmann J., Greenberg H. and Glass, R.I. (1993) Astroviridae. In: Classification and Nomenclature of Viruses. Report of the International Committee on Taxonomy of Viruses. S. Karger, in press.
8 Willcocks, M.M. and Carter, M.J. (1992) Arch. Virol. 124, 279–289.
9 Monroe, S.S., Jiang, B., Stine, S.E., Koopmans, M. and Glass, R.I. (1993) J. Virol. 67, 3611–3614.
10 Carter, M.J. Milton, I.D., Meanger, J., Bennett, M., Gaskell, R.M. and Turner, P.C. (1992) Virology 190, 443–448.
11 Kozak, M. (1989) J. Cell Biol. 108, 229–241.
12 King, L.A. and Possee, R.D. (1992) The Baculovirus Expression System, A Laboratory Guide. Chapman and Hall, London.
13 Laemmli, U.K. (1970) Nature 227, 680–685.
14 Carter, M.J. (1993) In: Proceedings of the Third International Symposium on Positive Strand RNA Viruses. Elsevier, Amsterdam, in press.
15 Carter, M.J., Milton, I.D. and Madeley, C.R. (1991) Rev. Med. Virol. 1, 177–186.
16 Neill, J.D., Reardon, I.M. and Heinrikson, R.L. (1991) J. Virol. 65, 5440–5447.
17 Meyers, G., Wirblich, C. and Thiel, H.-J. (1991) Virology 184, 664–676.
18 Carter, M.J., Routledge, E.G. and Toms, G.L. (1989) J. Gen. Virol. 70, 2197–2200.