Molecular analysis of human parechovirus type 2 (formerly echovirus 23)

Farideh Ghazi,† Pamela J. Hughes,† Timo Hyypia and Glyn Stanway

1 Department of Biological Sciences, John Tabor Laboratories, University of Essex, Colchester CO4 3SQ, UK
2 Haartman Institute, Department of Virology, P.O.B. 21, FIN-00014 University of Helsinki, Helsinki, Finland

Picornaviruses have been divided into five genera until recently, when a sixth genus, Parechovirus, was defined. Human parechovirus type 1 (HPeV1; formerly echovirus 22) was the first recognized member of this genus and preliminary sequence analysis of echovirus 23 (now renamed human parechovirus type 2 (HPeV2)) suggested that it is also a parechovirus. Here we describe the complete nucleotide and predicted amino acid sequences of HPeV2, which indicate a close relationship to HPeV1 throughout the genome. Sequence covariance in the 5′ untranslated region allows a prediction of the secondary structure, which indicates that these parechoviruses have a type 2 internal ribosome entry site, most closely related to that of cardioviruses. Overall, HPeV2 has 87.9% amino acid identity with HPeV1, most divergence being seen in regions of the capsid proteins that probably define antigenic sites. The N-terminal sequence extension to VP3, seen only in parechoviruses, is highly basic in both viruses, but has a variable sequence, suggesting that it does not have a sequence-specific role. There is an RGD motif near the C terminus of VP1, in an analogous location to that in HPeV1 which is believed to be functionally significant. The results confirm that both viruses are parechoviruses and give insights into the molecular features of this genus.

Introduction

Picornaviruses include a large number of clinically and economically important viruses among over 200 known serotypes (Rueckert, 1996). They consist of a single-stranded, positive-sense RNA genome (around 7100–9000 nucleotides in length), surrounded by 60 copies of each of four capsid proteins, VP1–4. The RNA has a single open reading frame, which is preceded by a 5′ untranslated region (UTR) and followed by a 3′ UTR and a poly(A) tract. A processing cascade, brought about by virus-encoded proteases, gives a number of functionally important precursors and the individual proteins. These include the structural proteins, encoded by the 5′-proximal region of the open reading frame, and non-structural proteins such as proteases and an RNA-dependent RNA polymerase. Until recently, five genera of the family Picornaviridae were recognized (aphtho-, cardio-, entero-, hepato- and rhinoviruses; Minor et al., 1995). Genera have been defined in the past largely on the basis of physico-chemical properties, but molecular genetics is now playing a more significant role and members of different genera, although sharing the same basic structure, show some differences in genome organization which can aid assignment (Rueckert, 1996). These include the higher order structure of the 5′ UTR, presence or absence of a leader protein (L), encoded upstream of the capsid proteins, and the nature of the 2A protein, encoded downstream of the capsid region. Nucleotide sequence analysis of echovirus 22 demonstrated that this virus is distinct from members of the existing genera (Hyypiä et al., 1992; Stanway et al., 1994) and it has recently been classified as the type member of a sixth genus, Parechovirus (Mayo & Pringle, 1998). As a consequence, echovirus 22 has been renamed human parechovirus type 1 (HPeV1). In addition to its overall low degree of sequence identity with members of the other genera, HPeV1 has some unusual characteristics, such as the apparent possession of only three structural proteins,
resulting from the lack of cleavage of VP0 into VP4 and VP2 (Hyypia et al., 1992; Stanway et al., 1994). Partial sequence
analysis of another picornavirus, echovirus 23, suggested that
it is a close relative of HPeV1 (Stanway et al., 1994), and it has
now been renamed human parechovirus type 2 (HPeV2). Here
we describe the complete nucleotide sequence of HPeV2,
together with further analysis which defines the distinctive
features of parechoviruses. The results confirm that HPeV2 is
a close relative of HPeV1 and give further insights into the
molecular characteristics of members of this recently recog-
nized picornavirus genus.

Methods

- **Purification of HPeV2 RNA.** HPeV2 (strain Williamson), originally
  obtained from the ATCC, was grown in A-Vero cells. When the
cytopathic effect was complete, virus particles were released by freezing
and thawing three times. Cell debris was removed by centrifugation and
the virions were purified by polyethylene glycol–NaCl precipitation and
sucrose-gradient centrifugation as described previously (Auvinen &
Hyypia, 1990). RNA was then isolated by phenol extraction and ethanol
precipitation.

- **Cloning of HPeV2 cDNA.** One µg HPeV2 RNA was reverse
  transcribed using an oligo(dT)$_12$ primer and aliquots were subjected to
  PCR amplification, with primers based either on the limited HPeV2
  genome or on the genomic HPeV1 sequence (Hyypia et al.,
  1992; Stanway et al., 1994). The procedure was as previously described
  (Gama et al., 1989). PCR products were cloned into the plasmid
  vector (Stratagene), using restriction enzyme sites built into the primers, after
  gel electrophoresis and electroelution. The products, together with their sense ( + ,
  same sense as genomic RNA; − , complementary to genomic RNA), the genomic locations of their
  binding sites in the HPeV2 genome and their sequence, were:

  - OL250 (+, 2764–2786) GGTGTGTAATAACGGTACNCNGG;
  - OL251 (−, 4423–4442) GATTTATTAACCATGTGATCACAACATCTCCTATGCC;
  - OL315 (+, 637–660) AAACGCTAGAGGGCCAACCCAG;
  - OL316 (−, 2959–2977) GTTGTGCAATATCTGCTGTGTA;
  - OL328 (+, 4462–4480) GATATGGTACCTATGGCCC;
  - OL327 (−, 6824–6846) TGAATGATCACATCATCNCCTA;
  - OL367 (−, 6379–6778) TCCAGATCGACATATGCGCAGAAGATG;
  - OL129 (−, poly(A)) ACGGCATGCACGGC(T)$_{15}$.

  To allow cloning of the first 70 nucleotides of both HPeV2 and
  HPeV1, the RNA was heated to 95 °C for 5 min and snap-cooled, prior
  to reverse transcription using OL538 (−, position 237–254, GCAGC-
  GGATCTGTACATAATACAGGT) as primer. PCR was then carried
  out with OL538 and OL537 (−, based on positions 1–16 of HPeV1
  RNA, CAGCCGAGCTCTTTGAAAGGGGTCTCC) as primers. The
  products were sequenced both directly and after cloning into the vector
  plasmid, with identical results in each case.

- **Sequencing and sequence analysis.** Sequencing was performed
  by the dideoxynucleotide method, after the generation of nested
deletions (Henikoff, 1984). Data were assembled and analysed using the
  Staden-Plus (Amersham Life Science) suite of programs. RNA secondary
  structure predictions were made on the basis of previously published
  structures for encephalomyocarditis virus (EMCV) (Pilipenko et al., 1989;
  Duke et al., 1992) and with the program PCFOLD (Zuker & Stiegler,
  1981). HPeV1 and HPeV2 amino acid sequences were aligned using an
  EBI Fasta3 database search (http://www2.ebi.ac.uk/fasta3/). Dendro-
  grams, based on identities derived using the ALIGN query at the
  GENESTREAM SEARCH network server CRBM Montpellier, France
  (http://genome.eerie.fr/bin/align-guess.cgi), were produced using the
  PHYLIP software (Felsenstein, 1989).

### Results

#### HPeV2 nucleotide sequence

Nucleotide sequence analysis of HPeV2 was performed following PCR amplification of fragments of cDNA. The primers used were based on previously obtained sequences from HPeV2 cDNA clones covering the VP1 region (Stanway et al., 1994), part of the 5’ UTR and part of 2BC, and on the published HPeV1 sequence (Hyypia et al., 1992). Previous attempts to obtain cDNA clones spanning the first 72 nucleotides of the HPeV1 and HPeV2 genomes had proved unsuccessful and the published HPeV1 sequence for this region was obtained by sequencing the RNA genome directly using reverse transcriptase (Hyypia et al., 1992). This region from both HPeV1 and HPeV2 was successfully cloned and sequenced by heating the RNA to 95 °C (to denature secondary structure) prior to reverse transcription, then amplifying by PCR with a primer based on nucleotides 1–16 of HPeV1 RNA. The 5’-terminal sequence obtained for HPeV1 shows eight differences from the published sequence, presumably due to technical difficulties during direct RNA sequencing. Excluding the 3’ poly(A) tract, the HPeV2 genome is 7328 nucleotides long and the nucleotide and predicted amino acid sequences show a high degree of identity to those of HPeV1 (Table 1). Both viruses exhibit a preponderance of

### Table 1. Identities between the sequences of HPeV1 and HPeV2

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number of amino acids</th>
<th>Amino acid differences</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’ UTR</td>
<td>—</td>
<td>—</td>
<td>86</td>
</tr>
<tr>
<td>3’ UTR</td>
<td>—</td>
<td>—</td>
<td>86</td>
</tr>
<tr>
<td>Protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP0</td>
<td>289</td>
<td>52</td>
<td>82</td>
</tr>
<tr>
<td>VP3</td>
<td>253</td>
<td>42</td>
<td>82</td>
</tr>
<tr>
<td>VP1</td>
<td>230</td>
<td>51</td>
<td>78</td>
</tr>
<tr>
<td>2A</td>
<td>150</td>
<td>21</td>
<td>86</td>
</tr>
<tr>
<td>2B</td>
<td>122</td>
<td>6</td>
<td>95</td>
</tr>
<tr>
<td>2C</td>
<td>339</td>
<td>43</td>
<td>87</td>
</tr>
<tr>
<td>3A</td>
<td>117</td>
<td>12</td>
<td>90</td>
</tr>
<tr>
<td>3B</td>
<td>19</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>3C</td>
<td>200</td>
<td>3</td>
<td>99</td>
</tr>
<tr>
<td>3D</td>
<td>404</td>
<td>22</td>
<td>95</td>
</tr>
</tbody>
</table>
Fig. 1. Comparison of the predicted polyprotein sequences of HPeV1 and HPeV2, showing the likely position of capsid protein secondary structural features and characteristic motifs. Identical amino acids are indicated by vertical lines. β-strands (underlined) and α-helices (underlined with wavy lines) are as predicted by Stanway et al., 1994. Hydrophobic domains in the 2A and 3A proteins are in black highlighted in grey. The VP3 N-terminal extension, seen only in parechoviruses, and the VP1 RGD motif, are in white highlighted in black. Functional motifs in several proteins, discussed in the text, are in white highlighted in grey.

Analysis of human parechovirus 2

2649

2025
codons ending in A or U (63.5% in HPeV2). The predicted amino acid sequence is compared to that of HPeV1 in Fig. 1. There are 2179 amino acids, of which 1916 (87.9%) are identical to HPeV1 residues.

**Molecular relationships with other picornaviruses**

Dendrograms based on amino acid sequence identities of the proteins VP1 and 2C, illustrating the molecular genetic relationships between representative picornaviruses, are shown in Fig. 2(a, b). A similar result was obtained whichever protein was analysed, and the results demonstrate that the genetic grouping of picornaviruses is largely consistent with recognized genera. There are some exceptions, namely the overlap between entero- and rhinoviruses in the dendrogram based on the 2C protein (Fig. 2b) and the failure of equine rhinovirus 1 (ERV1) to cluster with the other aphthoviruses in the 5′ UTR dendrogram (Fig. 2c). The results show that parechoviruses constitute a clear genetic lineage. Although the rest of the parechovirus genome is divergent from those of other picornaviruses in primary sequence terms, the 5′ UTR shares several blocks of nucleotide identity with cardioviruses and, to a lesser extent, with aphthoviruses. The overall degree of identity between picornavirus 5′ UTRs can be seen in Fig. 2(c), which shows that on this basis entero- and rhinoviruses cluster together and there is clustering of aphtho-, cardio- and parechoviruses. Hepatoviruses are more distantly related to other picornaviruses.

**Secondary structure of the 5′ UTR**

Cardio- and aphthoviruses have been shown to share the same overall 5′ UTR secondary structure, which is different from that of entero- and rhinoviruses (Pilipenko et al., 1989; Skinner et al., 1989). The predicted folding of this region in parechoviruses is very similar to that of cardio- and aphthoviruses (Fig. 3). Covariance between HPeV1 and HPeV2 sequences adds considerable weight to the structure proposed. Twelve pairs of nucleotide substitutions and a further four individual substitutions (involving GU and GC interactions) maintain the stability of the structure. Many of the other differences are concentrated in a small number of the predicted loops, while most loops are well-conserved. Parts of the 5′ UTR likely to be involved in internal ribosome entry site (IRES) function, by analogy with other picornaviruses, are structurally closely related to cardioviruses (Jang & Wimmer, 1990). The sequences predicted to make up the top of the largest stem–loop structure (l) and the whole of stem–loops J and K show considerable identity to the cardiovirus EMCV, and evidently share the structure proposed for this virus (Pilipenko et al., 1989; Duke et al., 1992). In HPeV2, the initiation codon is located at position 703 and is in an optimal Kozak context (ANNAUGG). It is preceded by an oligopyrimidine tract, which is similar in sequence and in spacing from the AUG to that of cardio- and aphthoviruses (Meerovitch & Sonenberg, 1993). Parechoviruses lack the poly(C) tract seen in some cardio- and aphthoviruses.

The 5′-terminal region of parechoviruses is predicted to fold into a long, stable hairpin structure, which is analogous to that seen in cardioviruses (Pilipenko et al., 1990). A shorter structure occurs in hepatoviruses, while the terminal nucleotides of aphthoviruses appear to interact with a downstream region, forming a similar structure (Pilipenko et al., 1990). Rather than a long stem–loop, entero- and rhinoviruses have a cloverleaf structure at their 5′ terminus (Andino et al., 1990).

**Parechovirus cleavage sites**

In HPeV1, two of the polyprotein cleavage sites (VP0/VP3, VP3/VP1) have been identified directly by amino acid sequence determination of purified capsid proteins, and the others were predicted by protein alignments (Hyypia et al., 1992; Stanway et al., 1994). The sequences flanking the probable cleavage sites of HPeV1 and HPeV2 are shown in Table 2. It can be seen that there is a marked preference for glutamine (Q) in the P1′ position, while both P2′ and P3′ tend to be polar amino acids and P4′ is frequently a small, non-polar amino acid. Similar P1′ and P4′ preferences are seen in several picornaviruses and are functionally significant (Ryan & Flint, 1997).

**Capsid proteins**

It is expected that parechovirus capsid proteins share the basic core structure, a β-barrel, seen in other picornaviruses (Rossmann et al., 1985; Hogle et al., 1985; Luo et al., 1987; Acharya et al., 1989); detectable sequence conservation allowed a prediction of the amino acids making up the β-strands in HPeV1 proteins (Stanway et al., 1994). The HPeV2 sequence largely supports these predictions, since differences from HPeV1 are mainly outside the putative β-strands (Fig. 1).

Several of the most variable regions are at the termini, particularly the N terminus of VP3 and C terminus of VP1. The latter contains an RGD motif, which has been shown to be functionally significant in HPeV1 and is also present in HPeV2 (Stanway et al., 1994). In VP0, there is a particularly variable region between amino acids 15–40, which is within the portion of this protein corresponding to VP4 in other picornaviruses.

**Non-structural proteins**

Some of the non-structural proteins of parechoviruses, particularly 2C, 3C and 3D, contain motifs typical of well-understood proteins in other picornaviruses and this facilitates their assignment and the definition of their boundaries (Fig. 1). These motifs include the 2C sequences CGXCG(S/T) and DDLXQ, which are seen in all picornaviruses and are probably indicative of helicase function. In the 3C protease (3Cpro), the
active-site cysteine is in the context GXCG as in other picornaviruses. The active site of the 3D polymerase contains the sequence YGDD, and this protein also exhibits the well-conserved motifs PSG and FLKR. In other cases, the assignment of the boundaries is more difficult. This is particularly so for 3B (VPg), as there are no consensus sequences for proteolytic cleavage, as defined above, in the appropriate region (Table 2). The sequence for HPeV2 in this region is almost identical to that of HPeV1 and in both there is a long hydrophobic stretch, reminiscent of that seen close to the C terminus of the 3A protein in other picornaviruses (Fig. 1). This, together with the requirement for tyrosine at the third amino acid position for linkage to the RNA, means that the position of 3B is probably as shown in Fig. 1. The predicted protein shows some sequence identity with the 3B protein from other picornaviruses, notably in possessing the dipeptide KP, seen in the majority of 3Bs (Fig. 4). If this is the location of 3B, the N- and C-terminal cleavages (E/R and Q/R, respectively) are not typical of the others known or predicted for parechoviruses. However, there are similarly atypical cleavages at the predicted 3B N terminus of ERV1 and equine rhinovirus 2 (ERV2) (Wutz et al., 1996).

The parechovirus 2A lacks the motifs in the trypsin-like 2A of enterovirus- and rhinoviruses associated with protein cleavage or with the unusual 2A/2B cleavage activity of the cardio- and aphthoviruses. Studies using in vitro translation suggest that it has no proteolytic activity (Schultheiss et al., 1995; F. Ghazi and others, unpublished results). Re-examination of the probable boundaries revealed sequences similar to the 3C consensus defined above at both N and C termini, and these are shown in Fig. 1. Comparison of the parechovirus 2A with protein databases revealed no significant similarities to other proteins and its function remains unknown. It has a stretch of hydrophobic amino acids close to the C terminus which is possibly a transmembrane domain (Fig. 1).

3′ UTRs

Picornavirus 3′ UTRs vary between 40 and 165 nucleotides in length, those of HPeV1 and HPeV2 both being 90 nucleotides long. In several picornaviruses, the 3′ UTR has functionally important secondary or tertiary structure domains (Pilipenko et al., 1992; Pöyry et al., 1996; Mirmomeni et al., 1997), but the low degree of covariance between HPeV1 and HPeV2 does not enable the actual structure to be determined from a range of possible structures of comparable stability.

---

Fig. 2. Dendrograms, based on amino acid sequence identities in the VP1 (a) and 2C (b) proteins and nucleotide sequence identities in the 5′ UTR (c), expressing the relationship between representative picornaviruses. The sources of the sequences used are given in Stanway (1990) except equine rhinoviruses 1 and 2 (ERV1 and ERV2) which are from Wutz et al. (1996). HRV, human rhinovirus; CAV and CBV, coxsackieviruses A and B; PV, poliovirus; TMEV, Theiler’s murine encephalomyelitis virus; EMCV, encephalomyocarditis virus; FMDV, foot-and-mouth disease virus; and HAV, hepatitis A virus. The genus to which each virus belongs is also indicated; ERV2 is currently unassigned to a genus.
Discussion

We present evidence that HPeV2 is closely related to HPeV1 and therefore confirm that it is a second member of the newly defined parechovirus genus of the family Picornaviridae. HPeV1 and HPeV2 were originally considered to be enteroviruses. Some sequence data are now available for virtually all known virus serotypes that were originally isolated from cases of human disease and designated enteroviruses. These are the echoviruses, polioviruses, coxsackie viruses A and B (CAV and
Table 2. Proposed protein cleavage sites in parechoviruses

For each cleavage site, the top line represents the sequence from HPeV1, and the bottom line represents the sequence from HPeV2. Amino acid residues (one-letter code) are labelled from the fifth residue N-terminal of the cleavage site (P5') to the fifth residue C-terminal of the cleavage site (P5).

<table>
<thead>
<tr>
<th>Cleavage site</th>
<th>Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP0/VP3</td>
<td>N I Y D N A P N G K</td>
</tr>
<tr>
<td></td>
<td>D I Y D T A P S K P</td>
</tr>
<tr>
<td>VP3/VP1</td>
<td>V V T F Q N S W G S</td>
</tr>
<tr>
<td></td>
<td>L V T F Q N S W G S</td>
</tr>
<tr>
<td>VP1/2A</td>
<td>N L T N Q S P Y G Q</td>
</tr>
<tr>
<td></td>
<td>T W S D Q S P Y G R</td>
</tr>
<tr>
<td>2A/2B</td>
<td>D G D E Q G L S L L</td>
</tr>
<tr>
<td></td>
<td>E G D E Q G L S L L</td>
</tr>
<tr>
<td>2B/2C</td>
<td>I L S N Q G P F K G</td>
</tr>
<tr>
<td></td>
<td>I L S N Q G P F K G</td>
</tr>
<tr>
<td>2C/3A</td>
<td>Q L E N Q T L D L D L</td>
</tr>
<tr>
<td></td>
<td>Q L E N Q T L D L D L</td>
</tr>
<tr>
<td>3A/3B</td>
<td>E S K D E R A Y N P</td>
</tr>
<tr>
<td></td>
<td>E S K D E R A Y N P</td>
</tr>
<tr>
<td>3B/3C</td>
<td>F P V S Q R E F K N</td>
</tr>
<tr>
<td></td>
<td>F P V S Q R E F K N</td>
</tr>
<tr>
<td>3C/3D</td>
<td>D M S D Q G I V T E</td>
</tr>
<tr>
<td></td>
<td>D M S D Q G I V T E</td>
</tr>
</tbody>
</table>

CBV) and enteroviruses 67–71 (Pulli et al., 1995; Huttunen et al., 1996; Hyypia et al., 1997). It is clear that all except HPeV1 and HPeV2 are typical enteroviruses and thus these viruses are the only currently known members of the parechoviruses which infect humans. However, a previously unknown virus has recently been isolated from bank voles (Clethrionomys glareolus) and shown, by partial sequence analysis, to be related to HPeV1 and 2 (Niklasson et al., 1998).

HPeV1 and HPeV2 were first isolated in 1956 during an epidemic of summer diarrhoea and, although classified as enteroviruses, were shown to have atypical growth characteristics and cytopathology (Wigand & Sabin, 1961; Shaver et al., 1961; Jamison, 1974). Subsequent biological and molecular characterization, culminating in the determination of the HPeV1 nucleotide sequence, has confirmed the distinctive nature of these viruses. The work presented here adds to our understanding of the molecular features of parechoviruses.

In general, all picornaviruses have the same basic genomic organization, but the different genera exhibit specific characteristics, particularly with regard to 5′ UTR structure, the L and 2A proteins and the 3′ UTR. Sequence comparisons reveal that although parechoviruses are relatively distinct from other picornavirus genera in terms of their proteins, the 5′ UTR shares considerable identity with those of cardio- and aphthoviruses (Fig. 2). This may reflect a recombination event in the evolution of parechoviruses. Alternatively, it could be that obligatory, cis-acting sequences or interactions with cellular factors have limited the divergence of this region.

The picornavirus 5′ UTR has at least two functional domains; the 5′-terminal portion (about 90 nucleotides) seems to be involved in RNA replication, while the 3′-proximal portion makes up the IRES, involved in translation (Rohli et al., 1994). Several picornaviruses also seem to contain regions of distinctive nucleotide composition [e.g. the poly(C) or poly-pyrimidine tracts of aphtho-, cardio- and hepatoviruses] between these functional domains. Covariance between the HPeV1 and HPeV2 sequences, together with these similarities to other picornaviruses, enabled the secondary structure of the 5′ UTR to be predicted with some confidence, and this confirms the close relationship with cardio- and aphthoviruses (Fig. 3). The long, thermodynamically stable stem–loop (domain A) at the 5′ terminus is very similar in size to that of cardioviruses, and presumably functions in an analogous manner in RNA replication (Pilipenko et al., 1990). This stem–loop probably prevented read-through by reverse transcriptase during the cloning of both the HPeV1 and HPeV2 genomes, explaining the fact that all cDNA clones initially obtained terminated just before this feature. This also probably explains the difficulty in obtaining completely accurate results when the RNA genome was sequenced using reverse transcriptase. The revised HPeV1 sequence of this stem–loop, derived by RT–PCR following heating of the RNA to melt secondary structures, is presented in Fig. 3 and shows eight differences from the published sequence and differs at two positions from the HPeV2 sequence, determined in the same way. Since the strategy used required a primer corresponding to the first 16 nucleotides, it is possible that the sequence presented for positions 1–16 is not correct. However, it should be noted that a complete cDNA clone of HPeV1 incorporating the sequence presented here gives a virus of wild-type phenotype upon transfection (F. Ghazi and others, unpublished results).
Domains A to C are followed by the part of the 5′ UTR that varies most between HPeV1 and HPeV2. This seems to lack base-pairing and is noticeably rich in A and C residues. It may correspond functionally to the poly(C) tract of cardioviruses, since it is located in an analogous position. Downstream of this variable region are several stem–loops (D to L), some of which probably compose the IRES. The 3′-proximal part of the parechovirus IRES is over 60% identical to that of cardioviruses and has an oligopyrimidine tract immediately upstream of the AUG that initiates the open reading frame. It is therefore probable that parechoviruses initiate translation in the same way as cardio- and aphthoviruses, i.e. directly from the AUG, rather than by scanning from the initial ribosome binding site as in entero- and rhinoviruses (Meerovitch & Sonenberg, 1993).

A notable feature of the 5′ UTR is the structure of stem–loop D (Fig. 3). Although this stem–loop is situated in a region which is rather dissimilar to that in EMCV, and is itself somewhat different from its EMCV counterpart, it has two opposite, asymmetrical bulges, which are identical in sequence (GAAG and AAUUA) to those seen in a corresponding structure in EMCV. This suggests that they may have an important role in the formation of pseudo-knots or as a protein binding site. Analysis of the sequences of other picornaviruses indicates that a structure with identical bulge sequences can be predicted to occur in the presently unclassified ERV2, but not in the aphthoviruses ERV1 or foot-and-mouth disease virus (FMDV).

In aphtho- and cardioviruses, an L protein gene lies at the 5′ end of the open reading frame. No L protein appears to exist in the parechoviruses, as sequences close to the predicted N terminus of the polyprotein have been observed in the VP0 protein purified from virus particles (Stanway et al., 1994). This contradicts the original prediction of a short L, based on the presence in HPeV1 of a consensus sequence for myristylation 12 amino acids downstream of the initiation codon and the observation of myristylation of VP0 in most picornaviruses (Chow et al., 1987; Hyypia et al., 1992). A potential myristylation site also exists in HPeV2, but this has the sequence GXXXT, while the sequence in other picornaviruses that show this modification is GXXXX. This, together with the protein results from HPeV1, makes it unlikely that the HPeV2 VP0 is myristylated. It is probable that the hepatovirus VP0 is also not myristylated (Tesar et al., 1993).

Comparisons between HPeV1 and HPeV2 capsid proteins allow regions of antigenic importance to be predicted, since these are likely to be variable between serotypically distinct viruses. As in several other picornaviruses, these are mostly located in loops connecting β-strands. The βBC loop of VP1 is probably antigenically important as are the VP1 βGH, VP0 βBC and VP3 βEF loops. Interestingly, the N terminus of VP3, which alignments with other picornaviruses suggest is extended by around 20 amino acids in parechoviruses, shows a number of substitutions between HPeV1 and HPeV2, while retaining its highly basic character. It is not clear whether this region is exposed on the surface of the virus or is internal, possibly interacting with the RNA. The RGD motif is also in a position where extreme variability is tolerated between HPeV1 and HPeV2, but is itself conserved. This is strong evidence that it is functional, a conclusion supported by its immediate context since, as in the functionally active RGD of strains of coxsackievirus A9 (CAV9), the motif is followed by a methionine (M; HPeV1) or leucine (L; HPeV2) (Chang et al., 1989, 1992; Roivainen et al., 1991; Hughes et al., 1995). The same is true of the Barty strain of echovirus 9, which contains a C-terminal extension relative to the Hill strain (Zimmermann et al., 1995, 1996). In parechoviruses, CAV9 and echovirus 9, the RGD motif is exactly the same distance from the VP1 C terminus, but the significance of this is unclear. The significance of the parechovirus RGD has been further shown by virus competition using CAV9, by blocking with specific peptides and by phage-display approaches (Roivainen et al., 1994; Stanway et al., 1994; Pulli et al., 1997).

The non-structural proteins of parechoviruses seem to have similar features to those of other picornaviruses. It is interesting that the predicted boundaries of VPg conform very poorly to the consensus for 3Cpro cleavage (Table 2), since the N-terminal boundary lacks a P1′ glutamine (Q) and both boundaries have a basic amino acid at the P1 position. A similar N-terminal cleavage is predicted in ERV2 (Wutz et al., 1996). It is possible that such ‘non-optimal’ cleavages are needed to prolong the life of precursors which may be needed for alternative activities.

The only parechovirus protein which differs radically from its counterpart in other picornaviruses is 2A. This is the least conserved non-structural protein between HPeV1 and HPeV2 and the differences are concentrated at the N terminus of the protein. The structural and functional properties of 2A gene products are variable among members of the different genera of picornaviruses. 2A serves a known proteolytic function in entero- and rhinoviruses by performing the primary cleavage event at its own N terminus, and is similar in structure to 3Cpro. In aphtho- and cardioviruses, an early cleavage occurs at the C terminus of 2A by a mechanism which is still not well-understood (Ryan & Flint, 1997). The aphthovirus 2A is very short (16 amino acids) and there is a homologous, similar-sized, region in the cardiovirus 2A. In both cases, these short sequences are sufficient for processing. The processing of hepatoviruses is distinct, since 2A does not appear to be involved in its own cleavage at either of its boundaries and VP1–2A accumulates as a major protein intermediate, cleavage occurring late in replication (Jia et al., 1993). Although 2A in parechoviruses is similar to that in hepatoviruses in that it does not appear to have a proteolytic function, it is structurally quite different (Schultheiss et al., 1995; F. Ghazi and others, unpublished results). Four distinct forms of 2A therefore exist among the picornaviruses. The absence of an L protein (a protease in aphthoviruses) and the probable lack of protease
activity in 2A, means that all cleavages in parechoviruses are likely to be mediated by 3C\textsuperscript{pro}. Together with characteristics such as the lack of cleavage of VP0, the nature of the 5’ UTR and the absence of myristoylation, it is clear that parechoviruses have a number of distinct features, which is consistent with their overall degree of genetic diversity from other genera.

This work was supported by the Wellcome Trust, the Academy of Finland and the Sigrid Juselius Foundation. F.G. is grateful for the award of a PhD studentship from the Ministry of Health and Medical Education, Islamic Republic of Iran.

References


Received 27 May 1998; Accepted 14 July 1998