Review article

Structure, function and evolution of picornaviruses

Glyn Stanway

Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K.

Introduction

Picornaviruses are a diverse family of small, RNA viruses which include several pathogens of man and animals. They are divided, on the basis of physicochemical properties such as acid stability and buoyant density, into four main genera: enteroviruses, rhinoviruses, cardioviruses and aphthoviruses. The enteroviruses are the most important in terms of human pathogenicity and include the polioviruses, coxsackieviruses and Echoviruses, together with enterovirus types 68 to 72. There are also several bovine and porcine serotypes. Rhinoviruses include 100 human serotypes (the major cause of the common cold) and two bovine serotypes. The cardioviruses, the best studied being encephalomyocarditis virus (EMCV), all belong to one serotype and are generally considered to be murine viruses. The seven serotypes of foot-and-mouth disease virus (FMDV) constitute the aphthoviruses. Altogether there are over 230 serotypes and their medical and economic importance have ensured that they have been the subject of a great deal of research (see the review by Rueckert, 1985). Progress has been facilitated by the fact that the viruses contain a relatively small amount of genetic information and can thus be studied readily at the molecular level, it is necessary by complete nucleotide sequence determination. The ability to crystallize picornaviruses and the application of X-ray crystallography to complete virus particles has generated further important data. The amount of information obtained by these two approaches is considerable and since this now includes nucleotide sequences and high resolution three-dimensional structures for representatives of each of the genera, this is at present a moment to review progress in the area. Here I will consider how this information is giving insights into picornavirus conservation and divergence and how, aided by genetic manipulation, immunological and other techniques, our understanding of the basic biological properties of the viruses is being advanced.

Structure of picornaviruses

Picornaviruses are non-enveloped and have a 25 to 35 nm capsid of icosahedral symmetry, made up of 60 copies of four proteins, VP1 to VP4 (Rueckert, 1985). Capsid assembly is via pentameric intermediates composed of copies of VP1, VP3 and a precursor protein, VP0, in which VP4 and VP2 are covalently linked. Cleavage of VP0, one of the final steps of capsid assembly, may be related to encapsidation of the nucleic acid and to stabilization of the mature particle. One or two copies of VP0 remain uncleaved in each capsid, the functional significance of this observation being unclear. Picornavirus proteins are non-glycosylated but VP4 (and VP0) is modified at its N terminus by linkage to myristic acid (Chow et al., 1987). This fatty acid has been shown to be associated with viruses of several different families, and an involvement in receptor binding, uncoating or capsid assembly has been suggested (Chow et al., 1987). In addition to this modification, the potential for VP2 and VP4 to be phosphorylated by cellular kinases, thereby destabilizing the capsid, may play a role in uncoating of the nucleic acid (Ratka et al., 1989).

The capsid proteins surround a single-stranded, positive-sense RNA genome of 7200 to 8500 nucleotides, the 5'-terminal nucleotide of which is linked covalently to a small virus-encoded protein, VPg. Upon entry into the cell and following uncoating, this RNA is translated by a cap-independent mechanism into one long polyprotein which is nascently cleaved by at least two virus proteases, 2A and 3C, to give the structural and non-structural proteins. Protease 2A performs the initial cleavage, liberating the precursor of the capsid proteins, whereas 3C is responsible for the majority of the processing of the polyprotein. The non-structural proteins include a virus-encoded RNA-dependent RNA polymerase which functions, in conjunction with cell proteins, to replicate the genomic RNA via a negative-sense intermediate (Rueckert, 1985).
The organization of the picornavirus genome is shown in Fig. 1. A 5' untranslated region (5' UTR), varying in length from 600 (rhinoviruses) to 1250 (aphthoviruses) nucleotides, precedes the open reading frame (ORF), which extends in excess of 2000 codons. This is followed by a 3' UTR and a poly(A) tract. Structural proteins, VP4, VP2, VP3, VP1 [1A, 1B, 1C, 1D according to the nomenclature of Rueckert & Wimmer (1984)] are encoded toward the 5' end of the ORF, whereas the other proteins, involved in RNA replication and protein processing, are encoded downstream. There are a few major differences in genome organization between the picornavirus groups (Palmenberg, 1987). The 5' UTRs of cardio- and aphthoviruses contain a long poly(C) tract which is absent in the entero- and rhinoviruses. No poly(C) tract is found in Theiler's murine encephalomyelitis virus (TMEV) which is elsewhere more similar to the cardioviruses than to the other picornaviruses (Pevear et al., 1987). Cardio- and aphthoviruses (together with TMEV) are further characterized by the presence of a 'leader' protein encoded prior to the capsid proteins. Aphthoviruses are unique in having three similar, but not identical, VPg-encoding sequences in tandem and they also possess an extremely short 2A-encoding region (see the review by Palmenberg, 1987).

Since 1981, when the Mahoney strain of poliovirus type 1 became the first picornavirus to be sequenced (Kitamura et al., 1981), many of the viruses have been analysed in molecular detail giving a wealth of information. Representatives of all the genera have now been studied and over 30 complete sequences have been published (see Table 1). Several partial sequences have also been determined. The data have transformed our knowledge of the fine structure of the virus genome and, together with three-dimensional structure studies on the virus capsid, have given a real insight into virus replication.

The 5' untranslated region

The 5' UTRs of picornaviruses are long compared with those of many cellular or other viral RNAs; indeed, they form 8 to 12% of the total virus genetic information. Between closely related viruses, the 5' UTR is generally the most similar part of the genome and contains several blocks of nucleotides which are perfectly or almost perfectly conserved (Stanway et al., 1984a). These presumably have sequence-specific, critical roles in virus replication although these have not yet been elucidated. Conserved sequences of this kind are at present finding use for broad-range virus detection systems based on the polymerase chain reaction (PCR) (Gama et al., 1988, 1989; Hyypiä et al., 1989; Torgersen et al., 1989) or oligonucleotide hybridization (Bruce et al., 1988).
An interesting aspect of the 5' UTR is that in the closely related entero- and rhinoviruses the regions align for the first 600 nucleotides, showing 60% homology; then the enteroviruses have a relative insertion of 100 to 140 nucleotides. This is one of the few distinguishing features at the molecular level between these viruses but its significance remains obscure. Viable poliovirus type 1 mutants lacking the extra sequences can be produced by in vitro manipulation (Kuge & Nomoto, 1987). Most of these mutants do not differ significantly in growth properties from the original strain in tissue culture cells and thus it is not clear why all the enteroviruses maintain the extra nucleotides, despite the additional replication cost involved, when rhinoviruses find them dispensable.

<table>
<thead>
<tr>
<th>Genera</th>
<th>Virus</th>
<th>Serotype/strain</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphthoviruses</td>
<td>FMDV</td>
<td>A10</td>
<td>Carroll et al. (1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1</td>
<td>Beck et al. (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>O1K</td>
<td>Forss et al. (1984)</td>
</tr>
<tr>
<td>Cardioviruses</td>
<td>EMCV</td>
<td>Rueckert, D (diabetogenic)</td>
<td>Palmenberg et al. (1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B (non-diabetogenic)</td>
<td>Bae et al. (1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M (medium plaque)</td>
<td>Luo et al. (1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BeAn 8386</td>
<td>Pevear et al. (1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DA</td>
<td>Ohara et al. (1988)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GD7</td>
<td>Pevear et al. (1988)</td>
</tr>
<tr>
<td>Mengovirus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMEV*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterovirus</td>
<td>Poliovirus</td>
<td>1 Mahoney</td>
<td>Kitamura et al. (1981)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Sabin</td>
<td>Toyoda et al. (1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Lansing</td>
<td>La Monica et al. (1986)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Sabin</td>
<td>Toyoda et al. (1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 117 (revertant)</td>
<td>Pollard et al. (1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 W-2</td>
<td>Pevear et al. (1990)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Leon/37</td>
<td>Stanway et al. (1984a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Sabin</td>
<td>Stanway et al. (1983)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 119 (revertant)</td>
<td>Cann et al. (1984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Finland</td>
<td>Hughes et al. (1986)</td>
</tr>
<tr>
<td>Coxsackievirus</td>
<td>A2</td>
<td>Horsnell et al. (unpublished)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A9</td>
<td>Chang et al. (1989)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A21</td>
<td>Hughes et al. (1989)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B1</td>
<td>Iizuka et al. (1987)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>Lindberg et al. (1987)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B4</td>
<td>Jenkins et al. (1987)</td>
<td></td>
</tr>
<tr>
<td>Echovirus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterovirus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>Auvinen &amp; Hyypiä (1990)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70</td>
<td>Ryan et al. (1990)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H/376</td>
<td>Inoue et al. (1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UKG/27/72</td>
<td>Knowles et al. (unpublished)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vg/5/27</td>
<td>Earle et al. (1988)</td>
</tr>
<tr>
<td></td>
<td>HAV</td>
<td>Chiron</td>
<td>Najarian et al. (1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MBB</td>
<td>Paul et al. (1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hm175</td>
<td>Cohen et al. (1986)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR326</td>
<td>Linemeyer et al. (1985)</td>
</tr>
<tr>
<td></td>
<td>(Hepatitis A)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HAV</td>
<td>Chiron</td>
<td>Najarian et al. (1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MBB</td>
<td>Paul et al. (1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hm175</td>
<td>Cohen et al. (1986)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR326</td>
<td>Linemeyer et al. (1985)</td>
</tr>
<tr>
<td></td>
<td>Rhinoviruses</td>
<td>Human rhinovirus (HRV)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1A</td>
<td>Palmenberg (1989)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1B</td>
<td>Hughes et al. (1988)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Skern et al. (1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>Leckie (1988)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>Stanway et al. (1984a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>85</td>
<td>Callahan et al. (1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>89</td>
<td>Horsnell et al. (unpublished)</td>
</tr>
</tbody>
</table>

* Reclassified virus.
† Proposed genus.
It is unlikely that the extra sequences themselves play a key role in replication as this is always the most diverse part of the 5' UTR in enteroviruses (Toyoda et al., 1984). However, the possibility remains that the length and general nature of the region is somehow important in defining the properties of the virus in the natural host. Indeed, recent work has indicated that mutants lacking the region have reduced neurovirulence (Iizuka et al., 1989). Mengovirus, a cardiovirus, is another case where 5' UTR sequences, i.e. the poly(C) tract, influence virulence (Duke et al., 1990). Shortening of this tract (from C_{50}UC_{10} to C_{15}UC_{10} or C_9) does not affect viability in tissue culture but has a marked effect on the pathogenicity of the virus in mice, its natural host (Duke & Palmenberg, 1989; Duke et al., 1990). The ability of these attenuated derivatives to stimulate immunity does not seem to be impaired. If FMDV pathogenicity in cattle can be attenuated by a similar truncation of its poly(C) tract, this could be an important route to a new generation of FMDV vaccines (Duke et al., 1990).

With the exception of the variability between enteroviruses and rhinoviruses just prior to the ORF, there are no other major differences in the 5' UTR between these closely related viruses except a degree of length and sequence heterogeneity around position 100. This is most marked in bovine enterovirus, where there is a large insertion, but whether this contributes to the distinct pathogenicity or host range of this virus is not clear (Earle et al., 1988). The conservation of the 5' UTRs between related viruses strongly implies a critical function and mutation experiments indicate that, in general, few changes can be made without compromising some aspect of the virus growth characteristics (Kuge & Nomoto, 1987; Trono et al., 1988).

In all picornaviruses the 5' UTR has a somewhat different nucleotide composition from the rest of the genome (Stanway et al., 1984a; Palmenberg, 1987), being lower in A and higher in C residues. The increased G + C content may reflect the fact that secondary structure plays an important role in the function of the 5' UTR, and stable structures can be derived by computer predictions. One folding model for the aphthoviruses predicts a highly ordered structure, much of the region consisting of a linear, double-stranded arrangement (Clarke et al., 1987). A second model, for which there are biochemical data derived from experiments with single- and double-strand-specific probes, is more complex and defines five main structural domains (Pilipenko et al., 1989a). A similar model seems to be applicable to the cardioviruses and to TMEV and this, together with the biochemical data, suggests that these structures do exist in vivo and have functional significance. In the case of enteroviruses and rhinoviruses, computer predictions, again augmented by experimental analysis, have produced a compelling model applicable to all the viruses studied but different from the predicted folding of the cardio- and aphthoviruses (Rivera et al., 1988; Pilipenko et al., 1989b; Skinner et al., 1989). There are three main predicted structural domains located at positions 236 to 443, 451 to 559 and 581 to 620 (numbers refer to poliovirus type 3) and all are supported by observed covariance. 'Covariance' is the retention of similar predicted secondary structure features between different viruses, despite variations in the nucleotide sequences. The structures are maintained by substitutions which retain a base-pairing capability with the corresponding nucleotide at the other side of the stem (e.g. C-G to G-U) or by double, compensating mutations (e.g. A-U changing to C-G). Conservation of these features suggests that they have some functional significance.

The functions of the 5' UTR are by no means fully elucidated but this part of the genome is being studied actively. It is probable that the 5' UTR contains sequences involved in RNA packaging and replication, and secondary structure may be important in RNA stability. Molecular analysis of the 5' UTR is, however, most advanced in the area of translation of the virus polyprotein and detailed studies have been performed on two picornaviruses, poliovirus type 1 (Pelletier et al., 1988a; Pelletier & Sonenberg, 1988; Bienkowska-Szewczyk & Ehrenfeld, 1988) and EMCV (Shih et al., 1987; Jang et al., 1988, 1989). These studies have shown that picornavirus polyprotein synthesis follows internal binding of ribosomes to sequences within the 5' UTR, termed the internal ribosome entry site (IRES) or ribosome landing pad (RLP). This is in contrast to the 'scanning model', believed to be applicable to the majority of cell mRNAs, in which the initial interaction is held to be between ribosomes and the 5' terminus of the mRNA (Kozak, 1989). The ribosomes then scan along the cellular mRNA and initiate translation at the first AUG or, in some cases, a later AUG which lies in a more favourable context. Analysis of several hundred mRNAs indicates that the consensus sequence for initiation codons is: 5', GCC(A/G)CCAUGG...3', where the underlined residues (-3 and +4) are particularly important. Despite the difference from cellular mRNAs in the point of ribosomal entry, a sequence which conforms well to this motif is found around the start of the long ORF in most picornaviruses.

It seems likely that in EMCV and poliovirus type 1, key elements of the IRES or RLP are located throughout much of the 5' UTR (Jang et al., 1989; Trono et al., 1988; Pelletier & Sonenberg, 1988; Pestova et al., 1989). Analysis of EMCV shows nucleotides 260 to 484 to be necessary for efficient translation and that at least part of
the IRES must be located within this region (Jang et al., 1988, 1989). In poliovirus type 1, the 320 to 630 region seems to be important in cap-independent translation and internal initiation (Pelletier et al., 1988a; Pelletier & Sonenberg, 1988) and elements toward the 3' side of this range seem to be of particular significance (Bienkowska-Szewczyk & Ehrenfeld, 1988). The corresponding region has been shown to play an important role in the translation of human rhinovirus 14 (HRV-14) RNA (AISaadi et al., 1989). Analysis of this region shows a C + U-rich sequence followed by the third conserved structural domain described above (a stem–loop), one or both of which may be involved in internal ribosome binding. Similar C + U-rich regions are found in FMDV RNA, located immediately prior to both the initiation codons used in this virus, and here homology with the 3' terminus of 18S RNA has been noted, suggesting some role in potentiating ribosome binding (Beck et al., 1983). Complementarity to ribosomal RNA has also been observed in enteroviruses (Iizuka et al., 1989). In enteroviruses the region may not be of critical importance to virus replication since viable poliovirus mutants have been produced which lack both the C + U and stem–loop features (Kuge & Nomoto, 1987). Interestingly, the stem–loop is located precisely before the 100 to 140 extra nucleotides seen in enteroviruses and, in all rhinoviruses sequenced to date, actually contains the AUG that initiates the ORF (AISaadi et al., 1989). This juxtaposition is both suggestive and puzzling as it may indicate that the highly similar entero- and rhinoviruses use different strategies for translation. In rhinoviruses, the ribosome-binding site may be closely followed by the AUG, whereas in enteroviruses the ribosomes would have to track the extra nucleotides before reaching the AUG. Tracking from a site around position 550 to 620 would explain the observation that in poliovirus an extra AUG introduced after this point, but before the authentic initiation codon, gives small-plaque mutants (Kuge et al., 1989). This is probably mediated at the translation level and reflects the reduced authentic initiation caused by the presence of the upstream AUG, as predicted by the scanning model. In contrast, the presence or absence of AUGs upstream of the IRES does not seem to affect poliovirus translation, at least in vitro (Pelletier et al., 1988b).

It now seems clear that internal ribosome binding and initiation occur in picornaviruses and there is growing evidence that some other viruses use a similar mechanism (Herman, 1989). Among the advantages are the ability to circumvent cap-dependence, and to be translated selectively when, as is the case with many picornavirus infections, the cap-binding complex is inactivated and host cell protein synthesis is shut off. It is possible that some cellular mRNAs exploit internal initiation to allow translation under conditions where cap-dependent translation is not efficient, for example during heat-shock (Sonenberg & Pelletier, 1989). Thus the principles established for picornaviruses may have wide applicability.

In addition to internal ribosome binding, deletion experiments have demonstrated the existence of elements which inhibit translation in vitro. In HRV-14 an inhibitory element seems to be located in the 490 to 550 region, and in poliovirus type 1 a similar element has been mapped to nucleotides 70 to 381 (AISaadi et al., 1989; Pelletier et al., 1988c). It is not clear whether this has any significance in vivo but if so, such an element, which may be based on secondary structure, could be involved in some hitherto unsuspected translational control. Such control could be mediated by cellular and/or virus proteins and it is interesting that interactions between cellular proteins and parts of the polio- and cardiovirus 5' UTRs have been demonstrated (del Angel et al., 1989; Sonenberg & Pelletier, 1989; Borovjagin et al., 1990). It has been suggested that variations in the levels of these proteins between cells may in part determine the tissue tropism of poliovirus and presumably of other picornaviruses.

The 5' UTR has, for several years, been the focus of attempts to understand the neurovirulence and attenuation of poliovirus. Comparative sequence analysis of three poliovirus type 3 strains differing in neurovirulence revealed that a single nucleotide in the 5' UTR (position 472) correlated with the virulent phenotype (Stanway et al., 1983, 1984b; Cann et al., 1984). This work was extended by partial sequence analysis of other strains and subsequently by in vitro manipulation of the viruses which indicated that the C to U change observed between the virulent and attenuated strains was a major determinant of attenuation (Evans et al., 1985; Westrop et al., 1989). The basis is not obvious but changes in the predicted secondary structure of the 5' UTR have been reported (Skinner et al., 1989) and their effect may be exerted at the translational level (Svitkin et al., 1985, 1990). Base changes near to position 472 are also believed to be important in the attenuation of the poliovirus type 1 and type 2 vaccine strains (Kawamura et al., 1989; Pollard et al., 1989). In addition to the extensive work on polioviruses, it has been shown that the 5' UTR from TMEV can influence virulence in mice (Calenoff et al., 1990). However differences in the 5' UTR of hepatitis A virus (HAV) contribute to cell culture adaptation but are not involved in attenuation (Cohen et al., 1989). Taking the results together, it is clear that the 5' UTR plays a major role in determining the biological properties of picornaviruses, and nucleotide sequencing, together with other forms of molecular analysis, is beginning to give an insight into the nature of that role.
Capsid proteins

The four capsid proteins, 1A, 1B, 1C and 1D, more commonly called VP4, VP2, VP3 and VP1 respectively, are encoded at the 5' end of the ORF. Three of the proteins, VP1, VP2 and VP3 are of similar size to one another in different picornaviruses [209 to 302 (VP1), 218 to 272 (VP2) and 221 to 246 (VP3) amino acids], whereas VP4 is much smaller. The HAV VP4 is only 17 amino acids in length and is significantly shorter than that of other picornaviruses (68 to 85 amino acids). As might be expected from their major role in defining antigenic and other characteristics such as receptor binding, the capsid proteins are diverse, this being particularly evident for FMDV where VP1 to VP3 are shorter than the other picornaviruses (Carroll et al., 1984; Palmenberg, 1989).

Prior to the determination to high resolution of the three-dimensional structures of several picornaviruses, the main information gleaned was that the capsid proteins were generally the most diverse of the virus proteins and that differences between serotypes were largely concentrated in certain parts of the three major proteins, often in hydrophilic regions. Direct analysis using monoclonal antibody escape mutants showed that some of these regions corresponded to neutralizing antigenic sites and this has allowed the production of synthetic peptides as potential vaccines and for probing the antigenic structure (Minor et al., 1983, 1986a; Ferguson et al., 1985; DiMarchi et al., 1986; Sherry et al., 1986). Alignment of sequences also enables the corresponding antigenic regions in viruses not studied directly to be identified. The N terminus of VP1 is another variable region but is hydrophobic in character. Subsequent analysis has suggested that it may interact with endosomal membranes to facilitate cell entry (Fricks & Hogle, 1990).

Our understanding of the structure of the capsid proteins and their role in virus replication improved dramatically with the determination of the three-dimensional structures of HRV-14 and poliovirus type 1 (Rossmann et al., 1985; Hogle et al., 1985; Arnold & Rossmann, 1990). These showed that the viruses share a great deal of structural homology and that VP1, VP2 and VP3 are all structurally similar to one another, each being composed of an eight-stranded $\beta$-barrel and differing primarily in the size of the loops and other elaborations which join or project from the $\beta$-sheets making up the core structure. VP1 and VP3 each have a relatively unstructured N terminus and VP4, whose cleavage from the precursor VP0 generating VP2 is the final step of maturation of the virus particle, seems to be effectively the N-terminal extension of VP2. Of interest is the fact that the $\beta$-barrel structure is found in the plant virus, Southern bean mosaic virus (SBMV) suggesting that it may be common to many RNA viruses with an icosahedral capsid (Rossmann et al., 1985). SBMV has a single coat protein and this, together with the common structures of VP1, VP2 and VP3, implies that these picornavirus proteins have arisen by gene duplication. Although in the capsid region there is little amino acid homology between HRV-14 or poliovirus type 1 and the more distantly related EMCV and FMDV, the ubiquity of the $\beta$-barrel structure was confirmed by the elucidation of the structures of the latter viruses (Luo et al., 1987; Acharya et al., 1989; Krishnaswamy & Rossmann, 1990). In addition, the arrangement of the proteins in the capsid, including the location and relative orientation, is essentially equivalent. The basic building block of the icosahedral capsid is the pentamer, made up of five copies of VP1 to VP4. The analyses show that, at the surface, the area surrounding the fivefold axis of the pentamer is composed primarily of VP1 while VP2 and VP3 are arranged side by side at a greater distance from the fivefold axis. VP4 has an extended configuration and is located internally underneath the other proteins. The pentamer appears to be stabilized by interactions involving the N and C termini of VP1 and VP3 together with VP4. Adjacent pentamers are held together by hydrogen bonds between parts of VP2 and VP3 and the relative weakness of these interactions may be important in the uncoating of the viruses. For FMDV, it has been suggested that the concentration of histidine residues (with a pK around 7) lining the pentamer interfaces is the basis of the instability of the virus below pH 7, and acid lability of rhinoviruses may have a similar origin (Acharya et al., 1989; Warwicker, 1989). An amino acid change at the pentamer interface in poliovirus type 3 vaccine strain gives a virus with a temperature-sensitive (ts) phenotype and this may be involved in attenuation of the type 3 vaccine (Filman et al., 1989; Westrop et al., 1989).

One of the most interesting observations to come from the first structure studied, that of HRV-14, was the presence of a deep canyon, cleft or pit running at a constant radius around the fivefold axis and lined by residues from VP1 and VP3 (Rossmann et al., 1985). It was suggested that this feature is involved in receptor binding, the rationale being that it would exclude antibodies which are bulky but could accommodate smaller, specific components projecting from the cell receptor, thus allowing attachment. This interaction probably requires the virus receptor-binding site to be highly conserved and this makes it vulnerable to the immune system, compromising the continuing existence of the viruses. Locating the key residues in an inaccessible position may be a strategy for preventing immune intervention. The virus surface can then be composed of
non-critical, potentially variable sequences, allowing a measure of antigenic diversity and enhancing the prospects for long-term picornavirus survival. Direct evidence for involvement of the canyon in HRV-14 attachment to cells was obtained when appropriate mutagenesis was shown to give viruses with an altered affinity for the cell receptor (Colonno et al., 1988). In addition, drugs known to bind inside a hydrophobic pocket underneath the canyon, resulting in a distortion of the canyon floor, prevent attachment of this rhinovirus serotype (Smith et al., 1986; Pevear et al., 1989).

A similar canyon is observed in the structures of poliovirus types 1 and 3 and Mengo virus and thus there is strong direct and comparative evidence that it is functionally active in a range of picornaviruses (Rossmann & Palmenberg, 1988). However FMDV shows several differences from the other picornaviruses, including the apparent absence of a canyon (Acharya et al., 1989). In FMDV, as already discussed, the capsid proteins are smaller than their counterparts in the other viruses, mainly due to the loops which link the β-sheets in the core structure being shorter. The overall effect is to make the protein shell thinner and to give it a smoother appearance on the surface, with no canyon apparent. It seems probable, therefore, that the FMDV determinants of receptor binding are different from those of the other viruses and the evidence suggests that binding is mediated by elements exposed on the outside of the particle rather than hidden within the canyon. These elements are a loop in VP1 (positions 133 to 158), which projects to the surface, and the C terminus of the same protein which is located near to this loop (Cavanagh et al., 1977; Acharya et al., 1989). Antibodies known to bind to the loop inhibit cell attachment and proteolytic cleavage of either of these structures also abolishes binding. Furthermore, short peptides including the sequence RGD (Arg-Gly-Asp) are able to inhibit attachment (Fox et al., 1989). The possession of this sequence is a feature of several proteins which attach to cells via a group of receptors named integrins (Ruoslahti & Pierschbacher, 1987). These proteins include fibronectin, vitronectin and type I collagen. An RGD tripeptide is found in the 133 to 158 loop of virtually all FMDV strains studied to date (Makoff et al., 1982; Beck & Strohmaier, 1987), strongly suggesting a functional role. Interestingly, an RGD sequence has recently been identified in the VP1 protein of the enterovirus, coxsackievirus A9 (Chang et al., 1989). The motif appears within an apparent insertion of 17 amino acids relative to other enteroviruses, located at the C-terminal end of the protein. This part of the coxsackievirus A9 protein is predicted to be hydrophilic and so may be located on the surface, although it is not yet clear whether it is involved in attachment.

The location of the attachment site to a prominent part of VP1 in FMDV apparently contradicts the argument above that protection from the immune system is required for regions where maintenance of sequence is imperative to maintenance of function. However, it has been suggested that a short region of fixed sequence can be camouflaged by hypervariability of the flanking residues (Acharya et al., 1989) and such variability exists in these regions of FMDV strains (Makoff et al., 1982; Beck & Strohmaier, 1987). This model is supported by results from an analysis of coxsackievirus A9 strains isolated over several years, all of which were found to possess the VP1 extension (K. H. Chang et al., unpublished). In each case an RGD sequence is present but the amino acids immediately flanking this motif exhibit a high degree of variability. This suggests that RGD is involved in a critical function, presumably receptor binding, but that mutations around the motif are not detrimental to this function. By allowing antigenic change, these mutations may, indeed, be advantageous to the long-term survival of the virus.

Since FMDV has no canyon, it might be expected that receptor binding occurs by a different mechanism from that used by other picornaviruses. However recent results suggest that even in picornaviruses where the presence of a canyon has been observed, the situation may be more complex than previously suggested. Replacement of an antigenic surface loop in poliovirus type 1 with the corresponding region from a mouse virulent type 2 strain made the mutant pathogenic for mice (Murray et al., 1988a). If this effect is mediated at the receptor level, it may suggest that regions outside the canyon are involved and even that several parts of the capsid can be used, possibly with the potential for attaching to more than one receptor.

Picornaviruses can be divided into several groups on the basis of the receptors which they recognize (reviewed by Colonno, 1987). FMDVs form one group and cardioviruses another, but the situation among the entero- and rhinoviruses is more complicated. A small group of rhinoviruses (10 to 15%) recognize a receptor as yet unknown (Abraham & Colonno, 1984). The majority of rhinoviruses, together with some enteroviruses (coxsackieviruses 13, 18 and 21), attach to a second receptor. This has been identified as intercellular adhesion molecule 1 (ICAM-1), a protein found on the surfaces of many kinds of cells (Greve et al., 1989; Staunton et al., 1989). The identification of this rhinovirus receptor has enabled the in vitro expression of a truncated, non-membrane bound derivative (Marlin et al., 1990). Addition to culture medium prevents rhinovirus infections from infecting cells in vitro, presumably by blocking receptor-binding sites on the virus capsid. The antiviral activity indicated by these preliminary experiments...
suggests that this approach may contribute significantly to the future control of picornavirus infections.

ICAM-1 is a member of the immunoglobulin superfamily, proteins with structural similarities to immunoglobulin molecules. The poliovirus receptor, used by all poliovirus serotypes, is a previously unknown member of the same superfamily (Mendelsohn et al., 1989). Members of the coxsackievirus B group seem to recognize yet another receptor, probably shared with some of the Echoviruses, including type 6 (Crowell et al., 1987). Coxsackievirus A9, closely related to the coxsackievirus B viruses in terms of nucleotide homology, does not belong to this group which is perhaps not surprising in view of the possibly unique occurrence of the RGD motif in this enterovirus (Crowell et al., 1987). There is little information on the remainder of the enteroviruses.

To date, with the exception of the known involvement of the RGD motif in FMDV and the implication of an antigenic loop in poliovirus type 2 receptor binding, there is no definitive indication of the molecular basis of distinct receptor specificities, despite the growing number of structures which have been solved to near atomic resolution and the large number of capsid sequences available for comparison. Some of the most promising viruses for detailed analysis are the rhinoviruses, since members of different receptor specificity groups can share considerable overall homology at the structural and sequence level and it should be more readily possible to correlate differences with distinct receptor-binding properties. The solution of a second rhinovirus structure, that of HRV-1A, may be an important step in understanding the determinants of receptor specificity in these viruses since it allows a structural comparison of members of the major (HRV-14) and minor (HRV-1A) receptor groups (Kim et al., 1989). The basis of the distinct receptor properties is not immediately obvious but it has been suggested that differences in charge around the rim of the rhinovirus canyon are involved in determining receptor characteristics. Sequence comparisons between the four minor group (HRV-1A, -1B, -2 and -49) and three major group (HRV-14, -39 and -89) viruses for which capsid protein sequences were available identified five amino acids, all located in the canyon rim region, which correlated with 'receptor grouping'. Of these five amino acids, four [positions 81 and 228 in VP1 and 86 and 182 in VP3 (HRV-1A numbering)] are negatively charged in the minor group viruses and positively charged or neutral in the major group, whereas for one (position 279 in VP1) the trend is reversed. These observations are suggestive but more sequence data are required to establish the general nature of these charge differences among other rhinoviruses. Indeed, the sequence of HRV-85 (subsequently obtained), a member of the major receptor group, correlates poorly with these trends possibly indicating that the determinants of receptor specificity lie elsewhere (P. R. Horsnell et al., unpublished results).

Structural determinations have given a great deal of information about another important area, virus antigenicity. Sites of importance for the induction of neutralizing antibodies have been mapped on the capsid proteins and are found to be concentrated in unstructured, hypervariable regions located in the loops linking the β-sheets and projecting onto the surface (Rossmann et al., 1985; Hogle et al., 1985; Acharya et al., 1989). In this way the most prominent features of the surface are made up of potentially variable residues, presumably masking more critical residues from immune intervention. The degree of flexibility of these antigenic loops has recently been confirmed and exploited by the production of chimeric polioviruses in which a variable loop making up the major antigenic site of VP1 has been replaced by the corresponding loop from poliovirus type 3, giving a virus expressing type 1 and type 3 antigenicity (Burke et al., 1988; Murray et al., 1988b). The generality of this approach has been shown by the insertion of other antigenic domains, including regions from human immunodeficiency virus type 1 and human papillomavirus type 16 (Evans et al., 1989; Jenkins et al., 1990) and it has proved possible to replace another region of known antigenic importance, located in VP2 (Murdin & Wimmer, 1989). These experiments open up the exciting prospect that the extensive structural and molecular studies performed on picornaviruses may lead to the development of a new generation of live vaccines. The work on the receptor properties of poliovirus type 1–type 2 recombinants described above shows that chimeric picornaviruses may also prove to be powerful experimental tools.

Non-structural proteins

The picornavirus non-structural proteins, with the exception of the L protein of aphtho- and cardioviruses which is encoded at the 5’ extremity of the ORF, are encoded towards the 3’ end of the genome. In addition to L, there are basically seven non-structural proteins, 2A, 2B, 2C, 3A, 3B, 3C and 3D, although it seems that in some cases precursors are required for activity. An example is the immediate precursor of 3C and 3D (3CD) which is involved in the processing of capsid protein precursors in poliovirus (Ypma-Wong et al., 1988a). Comparisons of non-structural proteins from different picornaviruses generally show that they are more conserved than the capsid proteins, presumably reflecting the lack of immune pressure and the need to maintain the active sites of enzymes.
The most conserved protein is 3D, the virus RNA-dependent RNA polymerase, which is over 95% identical between some enterovirus serotypes and shows extensive amino acid identity even between groups [e.g. HRV-14/FMDV, 33%; HRV-14/EMCV, 33%; HRV-14/poioivirus type 3, 65% (Stanway et al., 1984a)]. The functional domain of the polymerase is thought to include a GDD (Gly-Asp-Asp) motif conserved in all picornaviruses and similar to the YXDD (Tyr-Xxx-Asp-Asp) found in a wide range of other RNA virus polymerases including retrovirus reverse transcriptase (Kamer & Argos, 1984).

Sequence alignments have indicated that another highly conserved protein (2C) is also probably involved in RNA synthesis (Dever et al., 1987; Hodgman, 1988). It was found that 2C in FMDV shows a typical NTP-binding domain made up of three consensus elements GXXXXGK, DXXG and NKXD, allowing the prediction that it will bind a nucleotide as part of its biological activity. The first of these elements is found in all picornavirus 2Cs although the other two are often absent. Further analysis suggests that 2C, in common with the corresponding protein from other RNA viruses, exhibits some of the motifs found in a new superfamily of helicase proteins (Gorbulew et al., 1989a, b). Helicases promote DNA, RNA and RNA-DNA duplex unwinding and such an activity is quite likely to be important during RNA replication in picornaviruses. There is direct evidence that poliovirus 2C is involved in RNA synthesis, as mutations can lead to defects in this function, but it has not yet been shown that it has an RNA helicase activity (Li & Baltimore, 1988). A revertant of a poliovirus RNA synthesis mutant has a defect at the uncoating step, suggesting that 2C may also have a function in determining virion structure (Li & Baltimore, 1990).

The strategy of translation of the virus RNA into one long polyprotein necessitates cleavage to give the mature proteins. This is brought about by at least three proteolytic activities, all of which are thought to be encoded by the virus (reviewed by Kräusslich & Wimmer, 1988). 3C is known to carry out the majority of the primary cleavage of the polyprotein. In enteroviruses, 2A cleaves at the P1–P2 junction (i.e. at its own N terminus), separating the capsid precursors from those of the non-structural proteins (Toyoda et al., 1986). It is probably also involved in cleaving the cellular protein, p220, which is necessary for cap-dependent translation, and in this way may mediate host cell protein synthesis shut-off (Kräusslich et al., 1987). 2A in cardioviruses cleaves at its C terminus and does not have a role in host cell shut-off. Remarkably, the FMDV 2A is only 16 amino acids in length and yet is fully active in cis cleavage of the 2A–2B junction (Clarke & Sangar, 1988; M. D. Ryan, personal communication).

It is believed that the enterovirus 2A is also a member of the trypsin-like protease family, although it seems to belong to a different subclass from 3C. Again, cysteine replaces serine in a catalytic triad also including conserved histidine and aspartic acid residues (Bazan & Fletterick, 1988). These features are not shared by the cardio- and aphthoviruses.

In aphthoviruses a further protease has been demon-
strated: the L protein, encoded at the N terminus of the polyprotein, has been shown to cleave itself from the capsid precursors and to be involved in cleavage of p220 and host cell protein shut-off (Strebel & Beck, 1986; Devaney et al., 1988). Although cardioviruses also encode an L protein, this does not seem to be a protease and its removal from the capsid precursors is brought about by 3C (Parks et al., 1986).

The final cleavage to be performed is the generation of VP4 and VP2 from VP0, the last step of maturation of the virus particle. This presumably occurs inside the capsid and autocatalysis has been suggested on the basis of structural studies (Arnold et al., 1987). It is proposed that a serine residue (found at position 10 of VP2 in a number of picornaviruses) and the virion RNA itself are involved. However, there is no serine in a functionally similar position in FMDV and therefore this theory must remain tentative (Acharya et al., 1989).

VPg, a small protein of about 23 residues, is found linked covalently to the 5' terminus of all newly synthesized picornavirus RNAs, both positive- and negative-sense. Comparisons between different picornaviruses indicate little of importance other than the presence of a conserved tyrosine residue (at position 3) involved in the linkage to the RNA. It is believed that VPg is involved in RNA replication and it has been shown that VPg can become covalently attached to the RNA in the absence of other factors (Tobin et al., 1989). This suggests that either VPg, the RNA or a combination of the two represents the catalytic centre for this reaction.

The other non-structural proteins, 2B and 3A, remain ill understood although the former may be a determinant of host range in rhinoviruses (Lomax & Yin, 1989).

3' Untranslated region

The 3' UTR differs in length among the picornaviruses, ranging from about 40 nucleotides in rhinoviruses to 126 in EMCV. The region in rhinoviruses is AU-rich and contains few C residues; thus there is little stable secondary structure predicted. However, in other picornaviruses such structures can be predicted, and conservation and covariance suggest that these may have some functional importance (Inoue et al., 1989; Auvinen et al., 1989). Enteroviruses can be divided into two groups on the basis of 3' UTR sequences, a poliovirus group (which also contains coxsackievirus A21 and bovine enterovirus) with approximately 72 3' UTR residues and a coxsackie B virus group (which also includes swine vesicular disease virus, coxsackievirus A9 and Echovirus types 6 and 11) with about 100 residues. The difference corresponds to a complete stem-loop structure predicted in coxsackie B viruses but absent in polioviruses (Auvinen et al., 1989). Otherwise, little is known about the 3' UTR, although it may be expected to contain signals for polymerase binding.

The 3' UTR does not contain a polyadenylation signal since the poly(A) tract found at the extreme 3' end of all picornavirus genomes [and the poly(U) of the negative-sense RNA] is encoded genetically (Rueckert, 1985). There may, however, be some post-transcriptional addition. The poly(A) tracts are heterogeneous in length but the average size varies between 35 residues in cardioviruses and 100 in aphthoviruses.

**Codon usage**

As sequence data have accumulated, it has become clear that different picornavirus groups exhibit distinct nucleotide compositions even when the viruses show otherwise close homology (Palmenberg, 1987; Hughes et al., 1988). Closer analysis reveals that this is due largely to the viruses having characteristic codon preferences, with considerable variation in the ratio of each nucleotide found in the third position of triplets i.e. the position often free to change without altering the amino acid encoded. An informative parameter to use in the analyses is percentage C + G in this position and further discrimination can be seen on the basis of the relative numbers of --A to --A + --U codons. The coding region of 21 completely sequenced serotypes, together with two partial sequences (FMDV strains C and SAT 3) are analysed in Fig. 2. Using this approach, it can be seen that the picornaviruses cluster into three main groups. Aphthoviruses are at one extreme, having a distinct preference for C and G in the third position of their codons. Rhinoviruses are at the other extreme, with a marked preference for A + U, while enteroviruses and cardioviruses are intermediate. HAV is more similar to rhinoviruses than to any of the other viruses on the basis of --C/--G codons but can be distinguished by its preference for --U over --A + --U. TMEV differs from EMCV in the same fashion, despite being very close to this virus in terms of amino acid homology, and bovine enterovirus is also separated from the main group of enteroviruses by a preference for --U. A closer analysis of the enteroviruses reveals further subgrouping with coxsackie B viruses, swine vesicular disease virus and coxsackievirus A9 having a slight tendency towards --G and --C, the polioviruses being neutral and coxsackievirus A21 tending towards --A and --U. Similarly HRV-14 is distinct from the other rhinoviruses, having a slightly more enterovirus-like pattern.

It has been known for some time that organisms exhibit distinct codon preferences which may, in some
cases at least, be correlated with the abundance of different tRNAs and the level of expression of individual genes (Grantham et al., 1980; Chen & Inouye, 1990). Mammalian genes tend to be high in G + C in the third position and it is surprising that only FMDVs conform to this pattern, with most viruses showing little preference or, in the case of rhinoviruses and HAV, a marked bias towards A + U. It is not at all clear why the viruses differ in codon usage but one possible correlation is with temperature. Rhinoviruses infect the nose, which has a below average temperature and this is probably why their temperature optimum for growth in vitro is around 33 °C. RNA secondary structure, which may interfere with RNA replication or translation, will be more stable at 33 °C than 37 °C but is presumably lessened by a drift to higher A + U content, as A-U interactions are weaker than G-C. Therefore rhinovirus RNA may have accumulated A + U where these would not change the corresponding amino acid sequence and adversely affect protein function (i.e. mainly in the third position of codons). This supposition is supported by the slight -A + -U preference in coxsackievirus A21 which also infects the nose (Hughes et al., 1989). Stronger support comes from an analysis of enterovirus 70, a virus with a low temperature optimum which infects the eye; this has a clear preference for -A + -U (Ryan et al., 1990). This hypothesis does not explain the strong bias against -G + -C in HAV which infects the liver; another possibility is that different tissue-specific effects are being observed. It is interesting that genes encoding mammalian isozymes expressed in the liver show a lower -G + -C content (59%) than those expressed in muscle (80%) (Newguard et al., 1986). Finally, the observations may stem from non-adaptive phenomena. If the RNA polymerase were more prone to making errors during the addition of a particular base, then this base could become gradually less abundant, being retained only where a mutation would be lethal or deleterious to virus function. Codon preference may well be the result of a number of these and other pressures; this enigmatic phenomenon merits further investigation.

**Molecular relationships**

The availability of many picornavirus sequences has enabled detailed comparisons to be made which have led to an understanding of molecular relationships within the family (Palmenberg, 1989). Surprising results have emerged from these comparisons although there is considerable agreement with the original divisions made on the basis of physicochemical and pathogenic properties. The major surprise was that HAV, classified as an enterovirus, is only distantly related to the other enteroviruses and, indeed, to any other picornavirus, and clearly should be considered as a member of a separate genus (Najarian et al., 1985; Cohen et al., 1986; Palmenberg, 1989). Of the four current genera, the cardioviruses and the aphthoviruses form distinct groups whereas the rhinoviruses and enteroviruses overlap considerably. Although there may be a case for considering these as one large entero/rhinovirus genus, the current classification corresponds well to the specific molecular properties discussed above, the length of the 5' UTR and codon usage.

A summary of the relationship between available P1 encoding sequences (using only one member of each serotype) is shown in Fig. 3. The figure is based on the data of Palmenberg (1989) and several additional serotypes have been included. It can be seen that there is a considerable degree of nucleotide sequence identity between viruses within conventional genera (with the exception of HAV), the minimum identity level being nearly 50% (bovine enterovirus to the other enteroviruses). Within the aphthoviruses there is a minimum of about 55% nucleotide identity and so some aphthoviruses are less related than are EMCV and TMEV (60%) despite the latter viruses having previously been assigned to different genera. Molecular relationships among viruses within each genus are complex, particularly within the enteroviruses and rhinoviruses which comprise the great majority of picornavirus serotypes. Conventionally, enteroviruses are subdivided on the basis of pathogenic properties in man and experimental animals into the poliovirus, coxsackievirus A, coxsackievirus B and Echovirus groups. This classification holds well at the molecular level for the three poliovirus serotypes and for the coxsackie B viruses, within both...
groups there being a minimum of 70% P1-encoding nucleotide sequence identity. Between these subgroups there is much less similarity (55%). The relationship between molecular genetical properties and previous classification breaks down to some extent for the other subgroups. Coxsackie A viruses fall into at least three homology groups. Coxsackievirus A21 is very similar to the polioviruses (particularly in the 3' end of the genome) whereas coxsackievirus A9 is effectively a member of the coxsackievirus B group in molecular terms (Hughes et al., 1989; Chang et al., 1989). Coxsackievirus A2 is less related to other enterovirus subgroups (C. Horsnell et al., unpublished results). Partial sequence data suggest that serotypes 18 and 24 are similar to coxsackievirus A21 whereas A10 is more similar to A9 (P. R. Horsnell et al., unpublished results). Further work is necessary to understand fully this complex subgroup. The other enterovirus subgroup, Echoviruses, have been studied little. Sequence data indicate that at least two serotypes, 6 and 11, are closely related to coxsackie B viruses and hybridization results show that this observation can be extended to cover the majority of the serotypes (Auvinen et al., 1989; Auvinen & Hyypiä, 1990; P. Auvinen, personal communication). Echovirus type 22, on the other hand, has only a very low degree of identity to any other enterovirus and may eventually be considered the prototype of a new picornavirus genus (Auvinen et al., 1989; Auvinen & Hyypiä, 1990; T. Hyypiä, personal communication). Recently isolated enteroviruses are not subgrouped and are given the generic name and a number. Enterovirus 70 has been sequenced (Ryan et al., 1990) and the results suggest that it does not fall into any of the nucleotide identity groups defined above.

The determination of the nucleotide sequence of two animal enteroviruses has shed further light on the genus (Inoue et al., 1989; Earle et al., 1988). Swine vesicular disease virus is of particular interest since it is very highly homologous to the coxsackievirus B group (virtually as closely related as they are to one another). This is probably the basis of its known immunological relationship to coxsackievirus B5 (not yet sequenced). Bovine enterovirus is homologous to the other enteroviruses but is not particularly similar to any of the other subgroups, although in some regions of the genome it is most closely related to enterovirus 70 (M. D. Ryan et al., unpublished results).

Most rhinoviruses share at least 70% nucleotide identity, although HRV-14 is more divergent (55%). In view of the relative dissimilarity of HRV-14 to other rhinoviruses sequenced to date, it is unfortunate that this serotype has been studied in great detail as a typical member of the genus. Until recently, the rhinovirus sequence data suggested that HRV-14 is unique but it is clear now, from partial data generated by PCR, that HRV-3 is a close relative (R. E. Gama et al., unpublished results). This accords well with the results of a novel approach which classifies rhinoviruses into two groups on the basis of their susceptibility to antiviral compounds (Andries et al., 1990). These drugs bind within a hydrophobic pocket located underneath the surface canyon to prevent uncoating, and it has been suggested that the shape of this pocket varies between members of the two groups. Sequence data from other members of the HRV-14/-3 group are required to establish the validity of this model, which may have important implications for virus function and evolution.

At least partial data are available for the seven FMDV serotypes and this indicates that the A, C, O and Asia serotypes can be considered as one subgroup and the
SAT serotypes as another (Palmenberg, 1989). Members of the cardioviruses (e.g. EMCV and mengovirus) are closely related and form a subgroup distinct from TMEV.

Based on some of the above data it has been suggested that Picornaviridae can be divided into 11 subgroups or subgenera, each member of which shares >66% nucleotide homology with other members (Palmenberg, 1989). These subgenera are HRV-14-like, other rhinoviruses, poliovirus, coxsackievirus type A, coxsackievirus type B, bovine enterovirus, FMDV SAT-like, FMDV A/O/C/Asia-like, TMEV, mengovirus-like, and HAV. The number of subgenera may continue to grow as more data accumulate; already to this list should be added echovirus 22 and enterovirus 70. In addition the coxsackievirus subgenera need to be refined as coxsackievirus B1/B3/B4/A9/swine vesicular disease virus, coxsackievirus A21, and coxsackievirus A2.

There can be considerable variation within a serotype and between subtypes of a serotype, in addition to the variation between serotypes, subgenera and genera (data not shown). Strains of the same serotype of TMEV, poliovirus or FMDV can show as little as 90% amino acid identity in VP1 (Palmenberg, 1989; Hughes et al., 1986). In contrast, HAV strains are usually almost identical at the amino acid level and it has been shown that a simian strain shares about 83% nucleotide identity with human isolates in the P1 region (Brown, E. A., et al., 1989).

The majority of differences between closely related picornaviruses are concentrated in particular areas of the genome, mainly those encoding the surface residues of the capsid proteins, and particularly those outside the β-barrel core structure (Rossmann et al., 1985). These are often involved in determining the antigenic properties of the viruses and variation here leads to the diversity of virus serotypes. The ability of the capsid to accommodate such changes may be a factor in the number of serotypes seen for each virus. Thus, there are 100 characterized rhinoviruses, most of which are closely related in overall molecular terms but are immunologically distinct. On the other hand there are only three poliovirus serotypes despite the highly mutable nature of the RNA genome and the wide circulation of the viruses. The observation already discussed that the antigenic loop of a poliovirus type 2 strain is a determinant of pathogenicity may be of relevance here since this would imply that less variation than expected is possible (Murray et al., 1988a, b). A further type of limitation may be the observed retention of a trypsin-sensitive site in the antigenic loop of poliovirus type 3 which may be involved in escape from the immune system and so be an obligatory or highly advantageous feature (Minor et al., 1987).

The close relationship between viruses with similar properties and the functional homology between those more diverse indicate that the large number of picornaviruses have arisen by divergent evolution. More distant genetic lineages can also be identified taking into account genomic organization and three-dimensional structural motifs (Goldbach & Wellink, 1988). The clearest relationship is seen with a group of plant RNA viruses, the comoviruses (Fig. 1) which have a genome made up of two separately encapsidated segments of RNA, M RNA which encodes the structural proteins, and B RNA, the non-structural. Sequence comparisons indicate that the B RNA-encoded proteins are analogous to picornavirus non-structural proteins and the order of the genes is essentially identical. Thus, the extreme 3' gene encodes a polymerase and in each case this is located next to a protease gene (Argos et al., 1984; Kamer & Argos, 1984). VPg also has the same genomic location. In addition, there is a correspondence between the amino acid sequence of protein 2C of the picornaviruses and p58 of the comoviruses, both of which contain a consensus NTP-binding domain and some elements of a helicase (Dever et al., 1987; Gorbunova et al., 1989a, b; Habi & Symons, 1989).

On the basis of genome organization and translation strategy, positive-sense RNA viruses of plants have been divided into two groups, picornavirus-like and Sindbis-virus-like (Goldbach & Wellink, 1988). In addition to comoviruses, the picornavirus-like group contains the potyviruses whereas the Sindbis virus-like group contains several members, including the tobamoviruses and potexviruses. Detailed comparisons of helicase and polymerase motifs suggests that a third group, luteovirus-like, can be defined and that this forms an evolutionary link between the two groups identified originally (Habi & Symons, 1989). Amino acid sequence homology between the Sindbis virus core protein and FMDV VP3 has been noted (Fuller & Argos, 1987). Together with the common structure of these proteins, the eight-stranded β-barrel, and homology in the sequence and genomic location of the polymerase gene, it is likely that Sindbis virus and the picornaviruses have a common ancestor (Fuller & Argos, 1987; Rossmann & Johnson, 1989).

**Evolution of picornaviruses**

The generation of genetic and biological diversity within the picornaviruses is due to at least two mechanisms: firstly, the accumulation of mutations due to uncorrected errors in RNA replication; and secondly, RNA recombination. It is generally accepted that the lack of proofreading in viral RNA polymerases leads to a mutation rate orders of magnitude higher than that of the
host (reviewed by Smith & Inglis, 1987), direct experimental evidence coming from the measurement of the error rate of the RNA polymerase in vitro, from the incidence of reversion to or from a readily measurable phenotype such as temperature sensitivity or guanidine resistance, from T1 oligonucleotide mapping or from the frequency of neutralizing monoclonal antibody-escape mutants (Smith & Inglis, 1987; Ward et al., 1988; de la Torre et al., 1990). Error frequencies seem to depend on the method of measurement and vary from virus to virus but figures in the range $10^{-2}$ to $10^{-6}$/site/round of replication have been quoted. Assuming a value of $10^{-4}$, most picornavirus genomes will differ from the parental molecule and virus populations are likely to be extremely heterogeneous. The term ‘quasispecies’ has been used to describe this situation, essentially an equilibrium of RNA molecules each differing from the consensus at one or more locations (reviewed by Domingo et al., 1985).

Thus a pool of mutant genomes is generated from which variants can emerge rapidly if selective pressures are imposed. A good example is the live poliovirus type 3 vaccine strain which was attenuated by multiple passage of a wild-type virulent strain in monkey tissue and cell culture. The process introduced several mutations, one of which is at position 472 (C virulent, U attenuated) in the 5’ non-coding region (Stanway et al., 1984b). Upon vaccination, rapid reversion occurs, presumably due to low levels of pre-existent variants which have a growth advantage in the human gut (Evans et al., 1985; Minor & Dunn, 1988). Revertants are seen within a few hours and after 24 h all the viruses isolated have a C at position 472 and increased neurovirulence (Evans et al., 1985).

Another study on the evolution of picornviruses in individuals observed sequence variations of up to 0.8% over a 3-week period (Kinnunen et al., 1990). Rapid evolution of FMDV is also seen and serotype C3 isolated during persistent infection of cattle showed variation of 0.9 to 7.4% per year (Gebauer et al., 1988).

The evolution of picornviruses in the field has been the subject of an extensive study (Rico-Hesse et al., 1987). Using RNA sequencing of several wild-type isolates across a 200 base region at the 2A–VP1 boundary, it was shown that isolates from the same geographical location clustered into groups of closely related strains, the place of isolation being more important than the year. The geographical origins of outbreaks could thus be ascertained, information vital for epidemiology and control of infections, showing the positive impact that molecular techniques are having in all aspects of picornavirology.

The second mechanism by which picornaviruses evolve is recombination between different strains. Recombination was first observed as a laboratory phenomenon several years ago and used to map genetic markers onto the poliovirus and FMDV genomes (Hirst, 1962; Cooper, 1977; King et al., 1982, 1985). It occurs, probably by a copy-choice mechanism, during dual infection of a cell. That recombination occurs in more natural systems is now clear, an example being the isolation of intertypic recombinants from individuals vaccinated with the trivalent oral polio virus vaccine (Kew & Nottay, 1985; Minor et al., 1986b; Cammack et al., 1988). Recombinants between vaccine and wild-type strains have also been reported (Rico-Hesse et al., 1987). Recombination between closely related viruses can thus play a role in virus evolution but possibly of more significance is the recent determination of the sequence of coxsackievirus A21, which seems to indicate that this virus is a recombinant in which the 3’ 2000 nucleotides have been donated by a poliovirus (Hughes et al., 1989). TMEV is a further possible recombinant since the majority of the genome is very similar to EMCV but part of the 5’ UTR is quite distinct (Pevear et al., 1987). Enterovirus 70 is another virus where recombination may have occurred in the 5’ UTR since this region is very similar to that of the polioviruses but the rest of the genome is highly divergent (Ryan et al., 1990). If recombination between more distant viruses can give rise to viable progeny, this mechanism could then be a major factor in the generation of viruses with altered host range and tissue tropisms, thus contributing significantly to the diversity and success of the picornaviruses.

Potentially, there is a third and even more radical evolutionary mechanism, the introduction of cellular genetic information into the virus genome. This has been observed in influenza virus where a variant possessing a 28S ribosomal RNA insert in the haemagglutinin gene has been isolated (Khatcikian et al., 1989). A togavirus with an inserted sequence encoding animal ubiquitin has also been characterized and, as in the influenza virus example, an altered pathogenicity was observed (Meyers et al., 1989). There are as yet no clear examples of picornaviruses containing captured cellular genetic information, although the insert at the C terminus of coxsackievirus A9 VP1 may have a cellular origin. Since this region is variable, probably in response to immune pressures as discussed earlier, a cellular source is difficult to prove. However there is a clear amino acid sequence similarity to part of the precursor of human transforming growth factor β1 and this may have been the origin of the insertion (K. H. Chang & G. Stanway, unpublished results). Picornaviruses probably arose from cellular nucleic acids and host genes may have accumulated during evolution. Evidence for this includes the suggested structural and functional homology of the coat proteins to concanavalin A and the similarity between euakaryotic trypsin-like proteases and the 2A and 3C picornavirus proteins (Rossmann & Johnson, 1989). If coxsackievirus A9 is an example of such an acquisition
on a more contemporary time-scale, this could have important implications for picornavirus evolution.

Conclusion

This review illustrates how molecular and structural analyses have contributed to a greater understanding of the rich biological diversity seen in the picornaviruses. In many cases the second phase of studies, based on increasingly powerful genetic engineering, biochemical and immunological techniques is now well under way and beginning to reveal the depth of the vast store of information which still lies encoded cryptically in the sequences and structures which have been determined. The advances made already suggest that the next few years will be even more fruitful than those since 1981 when the first picornavirus sequence was published.

I would like to thank Jane Stanway, Christine Horsnell and Pamela Hughes who contributed greatly to the preparation of this review.

References


