The Nucleotide Sequence of Coxsackievirus A9; Implications for Receptor Binding and Enterovirus Classification

By KI HA CHANG,1 PETRI AUVINEN,2 TIMO HYYPIÄ2 AND GLYN STANWAY1

1Department of Biology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K. and 2Department of Virology, University of Turku, SF-20520 Turku, Finland

(Accepted 25 August 1989)

SUMMARY

The complete nucleotide sequence of the genome of coxsackievirus A9 (CAV-9) has been determined from cDNA cloned in Escherichia coli. Excluding the 3' poly(A) stretch, the RNA genome is 7452 nucleotides long and encodes a single polyprotein of 2201 amino acids. Comparison of the nucleotide and predicted amino acid sequences with those of the coxsackieviruses B1, B3 and B4 reveals a surprising degree of homology, with overall amino acid homologies of 86.9%, 86.2% and 87.0%, respectively. In contrast, there is much less homology to another coxsackie A virus, CAV-21, 60.4% overall amino acid homology. This demonstrates the high degree of diversity within the CAV group and indicates that the current classification does not directly correlate with molecular genetic properties. One major feature of CAV-9 is an insertion, relative to all other enteroviruses sequenced to date, which is located at the C terminus of VP1, and includes an arginine-glycine-aspartic acid tripeptide. Such sequences in a number of other proteins are known to have activity in promoting attachment to cell receptors and the implications for CAV-9 receptor binding are discussed.

INTRODUCTION

Coxsackieviruses, which include 23 group A serotypes (CAV-1 to CAV-22, and CAV-24) and six group B serotypes (CBV-1 to CBV-6), are common infectious agents that cause a wide spectrum of diseases, ranging from cold-like symptoms to severe infections of the central nervous system (Moore, 1982). They belong to the enterovirus genus of the family Picornaviridae and are small (28 nm), icosahedral particles with a coat made up of 60 copies of each of four proteins (VP1 to VP4), enclosing a single-stranded, positive-sense RNA genome which, in the case of enteroviruses, contains approximately 7500 nucleotides. The RNA has a 3' poly(A) tail and a small protein, VPg, covalently attached to the 5' terminus. The RNA is translated directly into a single polyprotein, which is cleaved by virus-encoded proteases to give the mature virus proteins (Rueckert, 1985).

The sequences of the genomes of numerous picornaviruses are now known (e.g. Kitamura et al., 1981; Stanway et al., 1984a; Carroll et al., 1984; Palmenberg et al., 1984; Pevear et al., 1987). Sequence comparisons within and between groups of picornaviruses have led to an understanding of their genetic relationships (Palmenberg et al., 1989), molecular evolution (Minor, 1988) and pathogenicities (Stanway et al., 1984b; La Monica et al., 1986). The availability of infectious cDNA copies of the viruses (Racaniello & Baltimore, 1981; Kandolf & Hofschneider, 1985) has provided a means of elucidating the role of specific viral sequences in biological properties, such as the role of 5' non-coding sequences in viral replication and the
attenuation of poliovirus (PV) neurovirulence (Semler et al., 1986; Kuge & Nomoto, 1987; Westrop et al., 1989). In addition, three-dimensional structures for PV1 (Hogle et al., 1985), rhinovirus 14 (Rossman et al., 1985), mengo virus (Luo et al., 1987) and foot-and-mouth disease virus (FMDV) (Acharya et al., 1989) have led to a further understanding of picornavirus function, including the identification of putative receptor-binding sites. Direct evidence for the location and nature of these structures has recently been presented (Colonno et al., 1988; Fox et al., 1989).

Recently, certain CBVs have also been studied in molecular detail. The complete nucleotide sequences of three serotypes [CBV-1 (Iizuka et al., 1987), CBV-3 (Lindberg et al., 1987) and CBV-4 (Jenkins et al., 1987)] have been determined, providing information on the genome organization of these viruses. Sequence comparisons have shown that all the CBVs form a relatively closely related group and in addition, swine vesicular disease virus (SVDV) (Inoue et al., 1989) has been shown to be a close relative. CAVs, despite being significant and diverse pathogens and a larger group, have not been studied as much and only one sequence has been determined (CAV-21; Hughes et al., 1989). This serotype is more closely related to the three PV serotypes and may be a recombinant virus, since the 3' homology to PVs is uniquely high (>95% at the amino acid level; Hughes et al., 1987).

To gain further information on the degree of diversity within the coxsackievirus group and to study the molecular basis of the wide range of biological properties, particularly pathogenicity and receptor binding, we have determined the complete nucleotide sequence of CAV-9. This virus was chosen for study as it is one of the most pathogenic CAVs in terms of disease severity and occurrence and, unlike many of this virus group, can be propagated readily in tissue culture (Moore, 1982). Comparisons of the sequence with those of other enteroviruses show that coxsackieviruses can be divided into at least two groups in molecular terms, but that these do not correspond to the classical A and B divisions.

METHODS

Virus propagation and RNA preparation. Coxsackievirus A9 (CAV-9, Griggs strain), obtained from the American Type Culture Collection, was plaque-purified three times in LLC-MK2 cells. The virus stock showed a typical neutralization pattern with the LBM pools and with CAV-9-specific antiserum (WHO), as did the purified virus.

For RNA preparation [3H]uridine-labelled virus was grown in Roux bottles. The cells were harvested when the cytopathic effect was complete and after three cycles of freezing and thawing the cell debris was removed by centrifugation (1000 g, 10 min, 4°C). SDS was added to the supernatant (final concentration 1%), which was layered onto 30 to 60% (w/v) sucrose step gradient and centrifuged at 24000 r.p.m. for 4 h at 20°C in a Beckman SW28 rotor. The pellet was suspended in 32.5% CsCl and further purified by equilibration centrifugation (SW50.1, 50000 r.p.m., 18 h, 4°C). The gradient was unloaded by collecting 200±l fractions and the radioactivity determined in 5±l samples. Virus was collected by pelleting and stored at -70°C until used. RNA was purified by proteinase K (100 µg/ml; Merck) digestion in the presence of 0.5% SDS, followed by phenol and chloroform extractions and ethanol precipitation. Northern blotting analysis of purified RNA, using random primed CAV-9 cDNA as probe, revealed a single band equal in size to CAV-3 RNA.

cDNA cloning. Two µg of CAV-9 RNA was reverse-transcribed and cloned by the modified cDNA-RNA method, using pBR322 cut and dG-tailed at the PstI site (Stanway et al., 1984c; Hughes et al., 1988). Over 800 recombinants were obtained and overlapping cDNA clones, which together spanned the whole genome, were identified by hybridization.

DNA sequencing and analysis. CAV-9 cDNA also contains three internal PstI fragments (positions 531, 2542 and 6451) in addition to those at the termini introduced during the cloning procedure. Following excision with PstI, purified cDNA fragments were circularized with T4 DNA ligase, sonicated to give fragments in the range 300 to 600 bp, end-repaired and inserted into M13 mp19 cleaved with HincII (Deininger, 1983). M13 recombinants containing the randomly generated CAV-9 inserts were then sequenced by the chain termination method (Sanger et al., 1977). The whole of the sequence was determined at least twice, approximately 70% was obtained in both orientations and the close similarity to other enterovirus sequences further reduces the possibility of error. The sequence was assembled and analysed using the Staden Plus DNA software (Amersham); the programs FASTP and FASTN running on a VAX 8600 were used to scan the NBRF (release 20), SWISSPROT (release 9) and EMBL (release 17) databases (Lipman & Pearson, 1985).
RESULTS AND DISCUSSION

Nucleotide sequence

The sequence derived for CAV-9 cDNA is shown in Fig. 1. At 7452 nucleotides, excluding the poly(A) tract, CAV-9 has the longest genome of all the human enteroviruses sequenced to date. The polyprotein coding region consists of 2201 codons, is preceded by a 5' non-coding region of 743 nucleotides and is followed by a 106 nucleotide 3' non-coding region. The nucleotide composition of the genome without the 3' poly(A) tract is 28.6% A, 24.5% G, 22.9% C and 24.0% U and so shows the slight A excess always seen in enteroviruses. An analysis of dinucleotide frequencies indicates the rarity of CG, typical of picornaviruses and the codon usage (Table 1) is as expected for an enterovirus, with a roughly equal distribution of third position nucleotides. The slight imbalance of --C/--G observed is characteristic of a CBV-like virus, while CAV-21 and PVs have a very slight --A/--U imbalance (Hughes et al., 1989).

CAV-9 was compared with previously sequenced enteroviruses (Toyoda et al., 1984; Lindberg et al., 1987; Iizuka et al., 1987; Jenkins et al., 1987; Inoue et al., 1989; Earle et al., 1988; Hughes et al., 1989) in the non-coding and coding regions and the results of these comparisons, which are shown in Tables 2 and 3, indicate that throughout the genome CAV-9 is highly homologous to the three sequenced CBVs and to SVDV. Much less homology is seen between CAV-9 and the PVs and, interestingly, between CAV-9 and the other CAV sequenced (CAV-21; Hughes et al., 1989). A bovine enterovirus (BEV) recently sequenced (Earle et al., 1988) is comparatively distantly related to CAV-9.

5' Non-coding region

The degree of homology of CAV-9 to the CBVs is high (83.6% to 86.4%) and is similar to that seen between the viruses (approximately 85%; Jenkins et al., 1987). The sequence bears about

<table>
<thead>
<tr>
<th>Table 1. Nucleotide frequencies (%) in CAV-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position in codon</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>U</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Nucleotide homology (%) between CAV-9 and other enteroviruses in the non-coding regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBV-1</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>5' Non-coding</td>
</tr>
<tr>
<td>3' Non-coding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Amino acid homology (%) between the proteins of CAV-9 and other enteroviruses</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBV-1</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>VP4</td>
</tr>
<tr>
<td>VP2</td>
</tr>
<tr>
<td>VP3</td>
</tr>
<tr>
<td>VP1</td>
</tr>
<tr>
<td>2A</td>
</tr>
<tr>
<td>2B</td>
</tr>
<tr>
<td>2C</td>
</tr>
<tr>
<td>3A</td>
</tr>
<tr>
<td>VPg (3B)</td>
</tr>
<tr>
<td>Protease (3C)</td>
</tr>
<tr>
<td>Polymerase (3D)</td>
</tr>
</tbody>
</table>
The nucleotide and predicted amino acid sequences of cDNA representing the entire genome of CAV-9. Protein boundaries were located by alignment with other enteroviruses. Underlined nucleotides are those participating in the 5' stem-loop and the conserved structure possibly involved in translation.

Fig. 1. The nucleotide and predicted amino acid sequences of cDNA representing the entire genome of CAV-9. Protein boundaries were located by alignment with other enteroviruses. Underlined nucleotides are those participating in the 5' stem-loop and the conserved structure possibly involved in translation.

the same degree of similarity to the PVs as to CAV-21 and this is substantially less than the relationship to the CBVs.

CAV-9 RNA shows several features found in other enteroviruses and a typical stem-loop structure, with 10 nucleotides in the stem (underlined in Fig. 1), is found near the 5' terminus (Larsen et al., 1981). Several sequence blocks highly conserved relative to the other enteroviruses are observed throughout the region. These are at positions 60 to 90, 450 to 480 and 540 to 570 and correspond to those previously described as highly conserved between rhinovirus 14 and PV-3 (Stanway et al., 1984a). In addition, nucleotides 1 to 35 and 513 to 645 are nearly identical to the corresponding regions of CBVs and SVDV. The notable exception is at the 5' terminus where one extra U residue, relative to all previously sequenced enteroviruses and rhinoviruses, is
observed in the cDNA. A conserved, predicted stem–loop structure, which we have recently described for rhinoviruses, is found in CAV-9 (positions 591 to 620), as in the other enteroviruses (Fig. 2; AlSaadi et al., 1989). Few nucleotide differences are seen in this feature when enteroviruses are compared and these are largely found in the loop, maintaining the overall structure. With the exception of BEV, the structure has a high predicted stability in each case. These observations suggest a functional role and it is interesting that in PVs and rhinoviruses there is evidence that a segment of RNA containing this region is necessary for efficient translation of the virus proteins (Bienkowska-Szewczyk & Ehrenfeld, 1988; AlSaadi et al., 1989). Preceding the loop in all the viruses, including BEV, there is a characteristic pyrimidine-rich stretch of approximately 17 nucleotides, which may also play some role in translation. The 100 to 120 nucleotides following the loop and preceding the AUG that initiates the open reading frame are the most variable region of the genome between different enteroviruses (Toyoda et al., 1984) and CAV-9 conforms to this pattern. The number of bases in this region that differ between CAV-9 and the three sequenced CBVs (CBV-1, CBV-3 and CBV-4) are 28/98, 29/98 and 32/98, respectively. Interestingly, one base difference (C-U) at position 674 was observed from the sequencing of two independent cDNA clones and this may suggest that the variable region is subject to considerable genetic drift even within an enterovirus serotype.

**Capsid proteins**

The capsid proteins and in particular VP1 are much more diverse than the non-structural proteins when compared with other enteroviruses (Table 2), but there is considerably more similarity to the CBVs than to CAV-21 and the PVs. The CAV-9 capsid protein sequences are aligned with those of CBV-4 (as a typical CBV) and CAV-21 in Fig. 3. It can be seen that many differences are concentrated into certain areas, which correspond largely with regions of known antigenic importance in other viruses (PVs have been studied particularly well; Minor et al., 1986), and are often in looped-out regions of the three-dimensional structure (Rossmann et al., 1985; Hogle et al., 1985). These alignments therefore allow the prediction of the CAV-9 antigenic domains and these are indicated in Fig. 3.

The most interesting feature of the capsid proteins is a relative insertion of approximately 17 amino acids at the C terminus of VP1. The sequence of this region, which was confirmed from two independent cDNA clones, is compared with the other enteroviruses in Fig. 4. There is a high degree of homology at the C-terminal side of the insertion with all the sequences except CBV-3 (Lindberg et al., 1987). The CBV-3 sequence can be made to conform with the other sequences if a one base deletion, followed 11 codons later by a one base insertion, is introduced. The close alignment with other enteroviruses, together with the absence of any other likely cleavage sites, strongly suggest that the insertion is at the C terminus of VP1, rather than the N
terms of 2A. Although there is some length heterogeneity at this point between the enteroviruses, the insertion seen in CAV-9 of between 15 (relative to CBV-3) to 19 (PV-2 and 3) amino acids is very large and presumably has important implications for the folding of VP1 and the overall three dimensional structure. Since the extra region is highly hydrophilic it is possible that it is exposed at the surface of the virus. If this is so, it is very interesting that an RGD (arginine-glycine-aspartic acid) tripeptide is found within the predicted sequence. Such a motif is known to play a role in cell adhesion in many proteins present in extracellular matrices and in the blood (Ruoslahti & Pierschbacher, 1987) and one located within the VP1 protein of FMDV has recently been shown to be involved in the binding of the virus to its cellular receptor. Furthermore, the C-terminal region of VP1 has also been identified as having receptor-binding
VP1/P2-A

CAV-9 PRLCQYKAFWGTPTIPDVPTQKDWQNLTVTQTPQSWLRGMLSTLTNSAFGQGQSAVYGNSRYVRVNRHL
CBV-1 PRLCQYKKNFRETQGVTTQRSN--TTT------------TQGAFGQSGAYVFGNTRVVRNHRL
CBV-3 PRLCQYKAKNWNPQGVTQTVTSQ---TTT-----------------------MNQGATWTTQGQVGCDYVRVNRHL
*CBV-3 PRLCQYKAKNWNPQGVTQTVTSQ---TTT-----------------------MNQGATWTTQGQVGCDYVRVNRHL
SVDV PRLCQYKAKWNPQGVTQTVTSQ---TTT-----------------------MNQGATWTTQGQVGCDYVRVNRHL
PV-1 PRAVLYDGGE-VYDGKLISPSLPSLKK--LTT------------------YTGFYHQKAVYTAGYICNYHL
PV-2 PRAVLYDGGE-VYDGKLISPSLPSLKK--LTT------------------YTGFYHQKAVYTAGYICNYHL
PV-3 PRAVLYDGGE-VYDGKLISPSLPSLKK--LTT------------------YTGFYHQKAVYTAGYICNYHL
BEV PRQAPYKKNVLVFSGSQDSRCSNRAS--LTT------------------YTGFYHQKAVYTAGYICNYHL

Fig. 4. Alignment of the VP1/2A boundary from all sequenced enteroviruses, illustrating the unique CAV-9 VP1 C-terminal extension. Predicted cleavage sites are underlined and hyphens denote spaces introduced into the sequences to optimize alignment with CAV-9. *Derived amino acid sequence which can be obtained by making a one base deletion in the published sequence followed 11 codons later by an insertion to maintain the reading frame. †Cleavage sites in PV-1 were determined by amino acid sequencing (Larsen et al., 1982; Pallansch et al., 1984).

properties in FMDV (Acharaya et al., 1989; Fox et al., 1989). In FMDV the C-terminal 17 amino acids of VP1 form a long arm, which is located on the surface of the capsid and is unique among the picornaviruses studied to date, but the C-terminal extension of CAV-9 described above may be related and the two viruses may share some functional homology in this region. If there is some involvement of the RGD sequence in CAV-9 it would suggest that this virus binds to cells by a mechanism different to that of the CBVs, even though there is otherwise close molecular homology, because there is no similar feature in the capsid proteins of these viruses. Our work may therefore have important implications for an understanding of receptor binding, tissue tropism and evolution of enteroviruses.

Non-structural proteins

The degree of homology between the P2 and P3 proteins of CAV-9 and those of the CBVs is very high (> 95% overall) and again approaches that seen between any two CBVs. The order of similarity seems to be CBV-1 > CBV-4 > CBV-3 > SVDV, but all the figures are very close and there is no clear evolutionary path.

In the P2 region, 2A, which has been assigned a definite function as a second protease in poliovirus (Toyoda et al., 1986), is slightly less conserved than the other two protein regions, 2B and 2C, the function of which remains obscure. The involvement in the 2A active site of the amino acid sequence PGDCGGILRC has been suggested. This sequence is conserved between other enteroviruses and is also present in the CAV-9 2A. The P3 region is processed to give 3A, VPg (3B), the 3C protease and RNA-dependent RNA polymerase (3D). The 3C protease shows the highly conserved sequences (GQCGGV and HVGGNG) noted by others (Inoue et al., 1989), including the cysteine (residue 147) and histidine (residue 161) believed to be part of the active site. The VPg sequence of CAV-9 differs from those of CBVs and SVDV by only one or two amino acids.

3′ Non-coding region

CBVs, together with echoviruses 6 and 11, differ from PVs and CAV-21 in the length of the 3′ non-coding region, since the former group have an extra 28 nucleotides after the stop codon (Auvinen et al., 1989). As expected from the other analyses, CAV-9 conforms to the CBV pattern rather than to CAV-21 or PVs and there is a considerable homology. As is the case at the 5′ terminus, which has an extra U residue, there are also extra residues (AGG) at the 3′ terminus when compared with CBVs.

Protein processing

The polyprotein cleavage sites have been established for poliovirus (Larsen et al., 1982; Pallansch et al., 1984) and the reasonable degree of homology fortunately enables the CAV-9 cleavage sites to be identified with some confidence (Table 4). The majority of the cleavages performed by the 3C protease are likely to be between QG amino acid pairs, the exception being...
Table 4. Proposed proteolytic cleavage sites in the polyproteins of enteroviruses

<table>
<thead>
<tr>
<th>Boundary</th>
<th>CAV-9</th>
<th>CBV-1</th>
<th>CBV-3</th>
<th>SVDV</th>
<th>CAV-21</th>
<th>PV-3</th>
<th>BEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP4/VP2</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>K/S</td>
</tr>
<tr>
<td>VP2/VP3</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>VP3/VP1</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>2A/2B</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>2B/2C</td>
<td>Q/N</td>
<td>Q/N</td>
<td>Y/G?</td>
<td>Q/N</td>
<td>Y/R?</td>
<td>Q/N?</td>
<td>Q/N?</td>
</tr>
<tr>
<td>2C/3A</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>VPg/Protease</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>Protease/Polymerase</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
</tbody>
</table>

the 2B/2C cleavage, which may well occur at a QN conserved between CAV-9, the CBVs and SVDV. The VP0 protein cleavage, possibly performed autocatalytically, almost certainly occurs at an NS (Arnold et al., 1987). However, as is often the case, the site of the P1/P2 cleavage performed by 2A is difficult to assign by alignment. It is likely that an FG or an HG are utilized and as they are almost adjacent, which one is used does not affect the above discussion on the VP1 insertion.

Implications for classification and receptor binding

Previous work has shown that coxsackieviruses share the typical features of enteroviruses, but can be distinguished from other members of this group by their pathogenicity for suckling mice. They are divided into CAVs and CBVs on the basis of the lesions produced in these mice. CAVs produce a diffuse myositis with acute inflammation and necrosis of fibres of voluntary muscles, whereas CBVs give focal areas of degeneration in the brain, focal necrosis of skeletal muscle and inflammatory changes in the dorsal fat pads (Grist et al., 1978). The basis of the characteristic pathogenicities for suckling mice is not clear, but one possible correlate is receptor specificity as it seems probable that all CBVs attach to the same cell receptor and there is also a good correlation between classification and receptor specificity in the CAVs. However, few CAVs have been tested and of these CAV-21 has some anomalous properties (Colonno, 1987; Crowell et al., 1987). There is no systematic information available on CAV-9, except that its attachment cannot be blocked by a CBV receptor-specific antibody (Crowell et al., 1986). The sequencing work described in the present paper, which shows that CAV-9 and CAV-21 are relatively dissimilar and that CAV-9 is closely related to CBVs, confirms and extends previous data based on hybridization (Hyytiä et al., 1984, 1987; Auvinen et al., 1989). Our results clearly show that the current criteria for classifying enteroviruses do not necessarily reflect the overall genetic relationships. The results, together with the CAV-21 sequence, indicate that CAVs fall into at least two groups, one PV-like (e.g. CAV-21) and one CBV-like (e.g. CAV-9). We have reached a similar conclusion from hybridization analysis of enteroviruses (Auvinen et al., 1989), which gave a crude division into these clusters, but there is still a need to obtain more sequence data from other members of CAVs to clarify the properties of this complex group.

One of the most interesting features of the sequence presented is the C-terminal extension to CAV-9 VP1. It has recently been shown that a strain of influenza virus of enhanced pathogenicity resulted from the insertion of a sequence derived from cellular rRNA into one segment of the virus genome (Khatchikian et al., 1989). Since the CAV-9 extension is unique among enteroviruses sequenced to date, it is probable that it has been acquired by heterologous recombination, either with a cellular mRNA or a different virus. However, when the inserted nucleotide sequence was used in both orientations to search the EMBL database no significant matches were observed. A search of the NBRF and SWISSPROT protein databases gave no significant amino acid matches with cellular proteins, but some homology was seen with the
VP1 loop of FMDVs which, as discussed above, is known to play a role in receptor binding. In addition to the RGD sequence, CAV-9 also shows a leucine residue located four amino acids downstream, which is seen in nearly all FMDVs (Palmenberg, 1989). There is particular homology with FMDV strain A24 (Makoff et al., 1982) which, in the region in question, has the sequence YGGSGRRGDMGSLAAR (underlined amino acids are homologous to those seen in the CAV-9 VP1 extension). Thus there may well be some functional homology between the sequences and they may have a common origin, the CAV-9 insertion being derived either from the same or a closely related cellular mRNA, or from an FMDV. The origin of any such inserted sequence may be difficult to discover since the implication is that the sequence may be on the surface and hence subject to immune pressures, which will tend to cause drift in residues not involved in key functions. Indeed, such hypervariability in residues surrounding the RGD in FMDV has been suggested as a mechanism by which this exposed sequence, obligatory for cell attachment, can be ‘camouflaged’ from the immune system by allowing the existence of several serotypes and subtypes (Acharya et al., 1989). Thus, unless the recombinant event is very recent, a great deal of divergence may have occurred from the original sequence. Although suggestive, our results do not definitely show the role of the RGD in receptor binding. Experiments are currently under way to mutate this sequence and also to see whether binding can be blocked by a synthetic RGD peptide (Fox et al., 1989). Studies are also being undertaken to elucidate the receptor grouping of CAV-9. The present work, together with these studies, is of great importance in gaining a better understanding of receptor specificity in enteroviruses and may eventually contribute to an understanding of the relationship between receptor binding and pathogenicity.

We are grateful to Christine North for assistance during the molecular cloning of the virus and thank Nick Knowles for pointing out the homology of the VP1 extension to FMDV A24. Part of the work was made possible by a grant from the Nuffield Foundation.

REFERENCES


Sequence of coxsackievirus A9


(Received 7 June 1989)