The nucleotide sequences of wild-type coxsackievirus A9 strains imply that an RGD motif in VP1 is functionally significant

Ki Ha Chang,1 Carol Day,2 Jackie Walker,2 Timo Hyypiä3 and Glyn Stanway1

1Department of Biology, University of Essex, Colchester CO4 3SQ, 2National Enterovirus Reference Laboratory, Epsom, Surrey KT19 8PB, U.K. and 3Department of Virology, University of Turku, SF-20520 Turku, Finland

We have shown previously that, compared to other enteroviruses, the coxsackievirus A9 (CAV-9) prototype strain, Griggs, contains a C-terminal extension to the capsid protein VP1 and that within this extension there is an RGD (arginine-glycine-aspartic acid) motif. To determine whether these features are found in other CAV-9 strains and therefore analyse whether they are likely to be functionally important, we have determined the nucleotide sequence of the appropriate region from five strains, isolated over a 25 year period. The results indicate that there is considerable diversity between the strains and there is little correlation between nucleotide sequence identity and date of isolation. All isolates exhibit the VP1 extension and although its amino acid sequence is otherwise variable, the RGD motif is common to all. This conservation of sequence, within a region which can otherwise vary, implies that the RGD sequence must be functionally significant. The VP1 extension shows similarity to sequences found in foot-and-mouth-disease virus strains and to part of the precursor of the cellular protein, human transforming growth factor β, and the possible significance of these observations is discussed.

Introduction

Coxsackie A viruses (CAVs), members of the enterovirus genus of the Picornaviridae family, exhibit a wide range of clinical manifestations, including cold-like symptoms, skin diseases, eye infections and central nervous system involvement (Grist & Reid, 1988). The viruses are composed of a ssRNA genome enclosed within a coat made up of 60 copies of four capsid proteins, VP1, -2, -3 and -4. Although poorly characterized at the molecular level compared to enteroviruses such as poliovirus and coxsackie B (CBV) viruses, our molecular understanding of CAVs has improved recently with the complete nucleotide sequence determination of two prototype strains, CAV-21 Coe and CAV-9 Griggs (Hughes et al., 1989; Chang et al., 1989). CAV-21 is remarkably poliovirus-like, particularly toward the 3' end of the genome, whereas CAV-9 shares considerable identity with CBVs. The most striking feature of the CAV-9 strain is an insertion in the VP1-encoding region compared to other sequenced enteroviruses. This encodes an apparent 17 amino acid extension to the C terminus of the VP1 protein and within this extension there is an RGD (arginine-glycine-aspartic acid) sequence, a motif known to be involved in receptor interactions in a number of cellular systems and in attachment of foot-and-mouth-disease virus (FMDV) to its cell receptor (Ruoslahti & Pierschbacher, 1987; Surovoy et al., 1988; Fox et al., 1989).

There is evidence that in rhinoviruses, which are highly homologous to enteroviruses in structural and functional terms, receptor attachment is mediated by interactions within narrow pits or canyons which encircle the fivefold axes of the icosahedral particle (Rossmann et al., 1985; Colonno et al., 1988). The 'canyon hypothesis' states that the dimensions are such that the relatively narrow cell receptor can penetrate the canyon but antibodies, which are larger, are excluded (Rossmann et al., 1985). In this way the integrity of virus–cell interactions can be maintained long-term without the pressure to change to avoid immune intervention. Many CAVs, including serotypes 13, 18 and 21, attach to the same receptor [intercellular adhesion molecule 1 (ICAM-1)] as the majority of rhinoviruses, presumably in an analogous manner (Colonno, 1987; Greve et al., 1989; Staunton et al., 1989). The CAV-9 receptor is distinct from ICAM-1, and from the receptor used by the closely related CBVs (Crowell et al., 1987), and the possession of the unique VP1 extension and RGD sequence raises the possibility that CAV-9 recognizes its receptor by a mechanism different from that of the other CAVs and CBVs that have been studied. It is possible that it has some functional similarity to the distantly related picornavirus, FMDV,
which attaches to its receptor by an RGD-mediated mechanism (Surovoy et al., 1988; Fox et al., 1989). To investigate further the significance of the VP1 extension and RGD motif, partial sequence data have been obtained from five wild-type CAV-9 strains. The data represent the first molecular analysis of clinical isolates of CAVs and enable preliminary analysis of the degree of variation within the CAV-9 serotype. More importantly, they strongly imply that the RGD motif has functional significance and give some indication of the possible origin of the VP1 insertion.

Methods

**Viruses.** Five CAV-9 strains, isolated from clinical cases in England over a 25 year period, were cultured in HEL cells prior to analysis. They were designated CO87, CO85, CO79, CO75 and CO62 respectively, according to the year when they were isolated.

**Oligonucleotide primers.** Oligonucleotides corresponding to positions 3663 to 3689 (OL94; 5' GGTCCCCTGACACCACGTT-GATTCC 5') and 3180 to 3199 (OL95; 5' ACCAA-ACCTTGGTTGTCGCCTAG 3') of the sequenced CAV-9 strain (Griggs) were used as primers. Polymerase chain reaction (PCR) amplification of virus nucleic acid should give a 510 bp product covering the region encoding the C terminus of VP1.

**RNA purification and reverse transcription.** To a 100 µl aliquot of tissue culture supernatant, tRNA (0.5 µg) was added as a carrier and vanadyl ribonucleoside complexes (10 mM) as an RNase inhibitor. The sample was extracted three times with an equal volume of phenol:amyl alcohol (50:1). RNA was precipitated with absolute ethanol and washed with 70% ethanol. Reverse transcription was carried out in a solution (40 µl) containing 50 mM-Tris-HCl pH 8.3, 75 mM-KCl, 10 mM-DTT, 3 mM-MgCl₂, 0.5 mM-dNTPs, RNasin (40 units), primer OL94 (0.1 nmol) and Moloney murine leukaemia virus reverse transcriptase (200 units) at 37 °C for 1 h (Gama et al., 1988, 1989).

**Amplification of viral cDNA using PCR.** Thirty cycles of PCR were performed on a 15 µl aliquot of the reverse-transcribed product in a solution (100 µl) containing 10 mM-Tris-HCl pH 8.3, 50 mM-KCl, 2.5 mM-MgCl₂, 0.2 mM-dNTPs, OL94 and OL95 primers (0.1 nmol) and 2.5 units of Taq DNA polymerase (Cetus). Each cycle included a 1 min 95°C denaturation step, a 2 min 55°C annealing step and a 4 min 72°C extension.

**DNA sequencing and analysis.** The PCR-generated DNA fragments were purified by agarose gel electrophoresis, electroelution and gel filtration. They were treated with the Klenow fragment of DNA polymerase I in the presence of nucleoside triphosphates and were ligated into HindII-cut M13mp18 to generate templates for sequencing. Sequencing was performed using the chain termination method and 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. A phylogenetic tree was calculated using the FITCH and PHYLIP suites of programs (Felsenstein, 1989).

**Results**

PCR amplification using the primers OL94 and OL95 yielded a DNA fragment from each of the virus isolates. Those from CO75, CO79 and CO85 were 510 bp in length, the size predicted from the sequence of the previously studied Griggs strain (Chang et al., 1989), whereas the CO62 and CO87 products were shorter than predicted (400 bp). Further analysis indicated that the latter products did not contain all the expected 3' sequences, presumably due to internal binding of the 3' PCR primer (OL94). Since this region was not of primary importance it was not investigated further in these two isolates. The nucleotide sequences of the PCR fragments, derived after cloning into M13, are shown in Fig. 1 (a). When compared to one another and to the prototype Griggs strain, nucleotide sequence differences in the range 17.2 to 24.4% were observed (Table 1). All differences were substitutions except the loss of one triplet within the VP1 insertion in CO79 and CO87. Considerable divergence (6.8 to 17.0%) was also seen at the amino acid level (Fig 1 b, Table 1).

The relationship between the strains at the nucleotide sequence level is shown in Fig. 2 and it can be seen that the most similar strains are CO62/Griggs, CO75/CO85 and CO79/CO87. There is no correlation between date of isolation and similarity, CO85 and CO87 being among the most divergent of the isolates. The region sequenced corresponds roughly to that analysed in extensive molecular studies on poliovirus epidemiology and the upper level of variation seen in CAV-9 strains (up to 24.4% at the nucleotide level) is similar to the maximum degree of variation seen within a poliovirus serotype (Rico-Hesse et al., 1987). This suggests that the region would be useful for future, more extensive studies on CAV epidemiology. The finding that there is not necessarily a correlation between date of isolation and nucleotide sequence identity parallels the observation already made for poliovirus. It suggests that there are

<table>
<thead>
<tr>
<th>Strain</th>
<th>CO87</th>
<th>CO85</th>
<th>CO79</th>
<th>CO75</th>
<th>CO62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide sequence differences (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Griggs</td>
<td>23.2</td>
<td>20.7</td>
<td>22.5</td>
<td>21.1</td>
<td>18.2</td>
</tr>
<tr>
<td>CO87</td>
<td>12.0</td>
<td>24.2</td>
<td>21.5</td>
<td>22.9</td>
<td>24.4</td>
</tr>
<tr>
<td>CO85</td>
<td>11.7</td>
<td>15.1</td>
<td>23.7</td>
<td>17.2</td>
<td>22.8</td>
</tr>
<tr>
<td>CO79</td>
<td>11.0</td>
<td>10.7</td>
<td>14.5</td>
<td>24.4</td>
<td>24.2</td>
</tr>
<tr>
<td>CO75</td>
<td>10.0</td>
<td>17.0</td>
<td>13.8</td>
<td>12.4</td>
<td>23.1</td>
</tr>
<tr>
<td>CO62</td>
<td>6.8</td>
<td>13.3</td>
<td>12.0</td>
<td>12.0</td>
<td>16.2</td>
</tr>
</tbody>
</table>

| Amino acid sequence differences (%) |

**Table 1. Nucleotide and amino acid sequence differences between the CAV-9 isolates**
multiple genetic lineages of CAV-9 circulating in the geographic area from which the isolates were obtained (England).

The genomic region studied in the clinical isolates is of particular interest because it includes the RGD motif and VP1 extension-encoding sequences initially observed in the Griggs strain. Sequence analysis of the isolates revealed that they all contained both these features (Fig. 1) and that although the RGD tripeptide sequence is absolutely conserved within the strains, the surrounding regions are highly variable. Indeed, virtually all differences seen between the strains in VP1 are located within the 14 amino acids flanking the RGD motif. In contrast, the part of the 2A protein sequenced exhibits a much more even distribution of differences. The most frequent substitutions (or deletions) in the region (RGD is positions 1 to 3) occurred at amino acid positions $-6$, $-5$, $-2$ and $+6$ (Fig. 1b). Three differences between the six strains analysed (Griggs and five clinical isolates) are seen at position $-6$, four at position $-5$, three at position $-4$ and five at position $+6$. The alanine residue missing at position $-5$ in both CO87 and CO79 confirms the conformational flexibility of the region. In addition to the RGD motif, further invariant amino acids are seen at positions $-4$ and $-3$ (QS) and $+7$ (L).

The presence of a conserved RGD motif within a region which can otherwise show a high degree of variability suggests that it must play a vital role in viral replication, most likely at the level of receptor binding. Since CAV-9 is very similar to the CBVs in molecular terms, the VP1 extension, which is unique among picornaviruses has been demonstrated, although the viruses involved are usually very similar (Hirst, 1962; King et al., 1985). An example occurring in humans is poliovirus, observed in vaccinees (Kew & Nottay, 1985).
acquired by heterologous recombination with a different virus. Indeed, we have already reported some similarity of the CAV-9 Griggs extension to a region of VP1 in FMDVs known to play a role in receptor binding. This similarity is also reflected in the other CAV-9 isolates. In addition to the RGD sequence itself, notable features include a leucine residue at position +7 (RGD-L), found in nearly all FMDV strains and common to the CAV-9 isolates. In addition, a leucine or methionine (L/M) frequently follows the RGD sequence in FMDV strains as in the CAV-9 isolates (N. Knowles, personal communication). There is a particular resemblance between CAV-9 (Griggs) and FMDV A24 (Makoff et al., 1982) which, in the region in question, has the sequence VGGSGRRGDMGSLAAR (underlined amino acids are identical to those seen in the CAV-9 Griggs VP1 extension). A large block of identity, QRATSRGGD-LAVLAQR, is also seen between CO85 and FMDV SAT3 (Brown et al., 1989).

Recently, there have been two reports of the identification of RNA viruses which contain extra nucleic acid sequences similar to cellular RNAs. The first was an influenza virus variant possessing a 28S ribosomal RNA insert in the haemagglutinin gene (Khatchikian et al., 1989). Subsequently, a togavirus with an inserted sequence encoding animal ubiquitin was characterized (Meyers et al., 1989). In both cases, an altered pathogenicity was observed. Heterologous recombination between picornaviruses and cellular RNAs is clearly a possibility and to investigate this further, the sequence of the Griggs strain VP1 extension was compared against DNA and protein sequence databases. Both orientations of the nucleotide sequence were used in a search of the EMBL database but no significant matches were identified. Initially, a search of the protein databases also yielded no significant identities. However, it was reasoned that if the extension had been captured as a functionally active unit, the original cellular or other protein would contain the extension had been captured as a functionally active unit, the original cellular or other protein would contain

allowing these insertions gives a match of 12/24 amino acids for the Griggs strain. The corresponding matches for the other CAV-9 isolates are: CO87, 12/23; CO85, 12/24; CO79, 12/23; CO75, 13/24; CO62, 12/24. In addition to this overall high degree of identity, it may be significant that the residue following the RGD is methionine (M) in the TGF-β1 precursor proteins (as in some of the CAV-9 isolates and FMDV strains) and that the +7 residue is isoleucine (I), a close structural analogue of the highly conserved leucine (L) found at this position in the CAV-9 isolates and FMDV viruses.

Discussion

The main point of the work described here was to determine whether the VP1 C-terminal extension seen in the Griggs strain (Chang et al., 1989) is present in other CAV-9 strains and whether the RGD motif is coincidental or is likely to play some role in the biology of the virus. The results show that in all the strains analysed, isolated over a 25 year period, both the extension and the RGD motif are retained. The perfect conservation of the RGD sequence is in contrast to many of the flanking amino acids, which showed extensive variation. Indeed, virtually all the amino acid differences seen in VP1 were concentrated into this region. This shows clearly that the VP1 extension has a high degree of conformational flexibility and since the RGD motif is perfectly conserved it must confer some sequence-specific selective advantage. This is most likely at the level of receptor binding, in view of the known involvement of the RGD motif in several specific cell–cell interactions and in the attachment of FMDV to its cell receptor (Pierschbacher & Ruoslahti, 1984; Surovoy et al., 1988; Fox et al., 1989).

It is interesting that the CAV-9 extension of all the isolates bears some similarity to the sequence of several FMDV strains from which the ‘consensus sequence’ RGD(L/M)–L can be derived (Makoff et al., 1982; Carroll et al., 1984; Forss et al., 1984; N. Knowles,
personal communication). We have recently completed the nucleotide sequence of echovirus 22 which differs from other picornaviruses and can be considered to represent a distinct genus (T. Hyypiä et al., unpublished results). The VP1 protein of this virus contains an RGD motif near its C terminus and this is in an identical context (RGDM-L), suggesting that all these viruses may share some functional homology at the receptor level. Thus, integrins, which interact with RGD-containing molecules, may be used quite frequently by picornaviruses to bring about cell attachment. These comparisons with other picornaviruses indicate an important role for the +4 and +7 residues [RGD(L/M)-L] in the function of the RGD motif and perhaps explain why they are well conserved, but the invariant nature of the −4/−3 QS remains enigmatic. This dipeptide is not seen in any of the FMDV strains or cellular, RGD-containing proteins which were analysed and a further assessment of its significance must await more extensive sequence analysis of CAV-9 isolates, or the construction of specific CAV-9 mutants using infectious cDNA.

It is known that the functional RGD of FMDV is located in an exposed position on the virus surface (Acharya et al., 1989) and it is likely that the CAV-9 RGD must be similarly exposed if it is involved in receptor binding. Although this was not addressed directly in the present work, the variable nature of the VP1 extension suggests that it is under immune pressure. Furthermore, it has recently been shown that in intact CAV-9 (Griggs) particles the VP1 protein is trypsin-sensitive (Roivainen et al., 1991). Treatment with this protease or with intestinal fluid reduces the size of VP1 slightly, consistent with cleavage at an arginine residue, within or close to the RGD motif. This motif is thus most likely to be located at the surface of the virus in a trypsin-accessible site. Further direct evidence for the significance of the RGD motif in the Griggs strain is the observed blockage of infectivity by synthetic RGD-containing peptides (Roivainen et al., 1991). The data presented here add weight to this direct evidence for the involvement of the RGD motif in receptor binding and shows that the possession of the VP1 extension and RGD is widespread among CAV-9 strains rather than being unique to the Griggs strain.

An exposed location for the receptor-binding domain seems to indicate that CAV-9 cell attachment is not according to the canyon hypothesis since this supposes that key residues are concealed to avoid immune intervention (Rossmann et al., 1985). This allows the generation of antigenic diversity at the virus surface without the necessity of changing residues involved in cell attachment. However, it has been suggested that in FMDV, antigenic diversity is still possible while maintaining the critical RGD moiety, if the flanking regions are hypervariable (Acharya et al., 1989). The derivation of a consensus sequence of invariant CAV-9 residues (DINTV-QS-RGD-L-THG) shows that there is little conservation of RGD-flanking residues in the clinical isolates, which is in good agreement with this concept and implies that the same factors may be operative in CAV-9.

The observation of the VP1 C-terminal extension in CAV-9 may be important for our understanding of picornavirus evolution, particularly in the light of the apparent similarity to TGF-β1. Two RNA viruses have been identified which contain cellular RNA sequences and CAV-9 may represent the first observation of a picornavirus with such captured sequences (K hitchikian et al., 1989; Meyers et al., 1989). The region of similarity is found within the precursor of TGF-β1, but not in the mature growth factor and it is not clear whether the RGD motif has any significance in the biology of TGF-β1, a multifunctional peptide that controls proliferation and differentiation in several cell types (Sporn et al., 1986). Although not part of the biologically active, mature molecule, the cleaved N-terminal part of TGF-β is secreted along with TGF-β1 and another protein, the TGF-β1-binding protein. These proteins form a latent complex from which TGF-β1 has to be liberated before it becomes active. Recently, the TGF-β1-binding protein has been sequenced and found to contain many repeats, one element of which is a DIN or DID sequence, reminiscent of that seen in the TGF-β1 precursor itself and in the CAV-9 isolates (Kanzaki et al., 1990). The binding protein also contains an RGD motif. The discovery of RGD sequences in two proteins of the latent complex suggests strongly that they have a function, almost certainly in interacting with cell surfaces. The conservation of the sequence of the CAV-9 isolates, demonstrated here, adds weight to this conclusion and possibly implies that CAV-9 utilizes the putative receptor recognized by the TGF-β1 latent complex. The role of RGD sequences in the biology of TGF-β1 and the relationship to CAV-9 receptors both seem worthy of further study.

We would like to thank Nick Knowles for making available several unpublished FMDV sequences and for assistance with the PHYLIP programs.

References


(Received 29 July 1991; Accepted 14 November 1991)