The Complete Nucleotide Sequence of a Bovine Enterovirus

By J. A. P. EARLE, R. A. SKUCE, C. S. FLEMING, E. M. HOEY* AND S. J. MARTIN

Department of Biochemistry, The Queen's University of Belfast, 97 Lisburn Road, Belfast BT9 7BL, U.K.

(Accepted 19 October 1987)

SUMMARY

The complete nucleotide sequence of the genome of a bovine enterovirus (strain VG-5-27) has been determined using molecular cloning and DNA sequencing techniques. Excluding the poly(A) tract the genome was 7414 nucleotides long and contained a 5' non-coding region which at 818 nucleotides was longer than that of most picornaviruses. The single large open reading frame encoded a potential polyprotein of 2175 amino acids which showed considerable homology with other enteroviruses. This homology has allowed us to predict the possible cleavage sites of the polyprotein and to identify other features of structural and functional significance which might help us to understand the constraints involved in the evolutionary divergence of these viruses.

INTRODUCTION

Bovine enteroviruses are members of the picornaviridae which are endemic in cattle populations without appearing to cause any disease. Many strains have been isolated and these have been shown to fall into two serotypes (Knowles & Barnett, 1985). We have extensively studied the biochemical characteristics of one bovine enterovirus (BEV; strain VG-5-27), a Northern Ireland isolate which belongs to serotype 1. Like other picornaviruses, purified virions contain a single-stranded positive-sense RNA molecule (Todd & Martin, 1979) linked to a small polypeptide, VPg. This genome is surrounded by a capsid coat consisting of 60 copies of four polypeptides, VP1, VP2, VP3 and VP4 (Hoey & Martin, 1974). One of these, VP4, is myristylated at its N-terminal residue (Chow et al., 1987). The RNA is translated into a large polyprotein which is subsequently cleaved to give the various virus-induced polypeptides found in infected cells. Pactamycin mapping experiments have established that the gene order is similar to that of other enteroviruses such as the polioviruses (unpublished results).

The advent of molecular cloning techniques has allowed the genomes of various picornaviruses, representative of the four genera, to be cloned and sequenced. Surprisingly the enteroviruses and rhinoviruses have been found to be much more closely related than previously thought (Stanway et al., 1984). Moreover, X-ray crystallographic studies have shown that the structures of poliovirus type 1 (PV1) and rhinovirus 14 (HRV14) are very similar (Hogle et al., 1985; Rossmann et al., 1985), implying that they have arisen from a common ancestor. All the viruses so far sequenced from these two genera have been of human origin. In this study we have determined the sequence of BEV and have carried out a comparison of the predicted amino acid sequence of each encoded protein with those of two human enteroviruses, PV1 (Kitamura et al., 1981) and coxsackievirus B1 (CB1; Iizuka et al., 1987) and with a rhinovirus, HRV14 (Stanway et al., 1984). The extensive homology that exists has enabled us to predict the probable cleavage sites of the BEV polyprotein and to demonstrate a close structural relationship between the animal and human picornaviruses.

METHODS

Virus. Bovine enterovirus (strain VG-5-27), a Northern Ireland isolate, was originally obtained from Professor J. B. McFerran, Veterinary Research Laboratories, Stormont, Belfast, U.K. It was plaque-purified twice and propagated in BHK cells as described by Martin et al. (1970).

0000-8006 © 1988 SGM
**Virus purification.** Virus was purified by differential and sucrose gradient sedimentation using the method of Martin et al. (1970).

**RNA extraction and analysis.** RNA was obtained from virions by the method of Todd & Martin (1979), purified by Sephadex G-50 chromatography and recovered by precipitation with ethanol. It was subsequently analysed by electrophoresis in 1% agarose gels using the methylmercury denaturing conditions described by Bailey & Davidson (1976).

**Cloning procedure.** Purified RNA (10 μg) was suspended in 25 μl 10 mM-methylmercuric hydroxide and 1 μl RNAsin (4 units/μl, Promega Biotec) and allowed to stand at room temperature for 15 min. Ten μl 700 mM-2-mercaptoethanol was added and after a further 5 min the RNA was reverse-transcribed in a 150 μl incubation mixture containing 50 mM-Tris-HCl pH 8.3, 100 mM-NaCl, 10 mM-dithiothreitol, 1 mM each of dATP, dCTP, dTTP, 0.2 mM-[3H]dGTP (sp. act. 2 mCi/μmol), 100 μg/ml actinomycin D, 20 μg/ml oligo(dT)12-18 and 500 units/ml avian myeloblastosis virus reverse transcriptase (Life Sciences, St Petersburg, Fla., U.S.A.) by incubation at 42 °C for 2 h. The reaction was terminated by the addition of 10 μl 100 mM-EDTA and the mixture was extracted with phenol and passed over a Sephadex G-50 column. The RNA/cDNA hybrids eluted in the void volume were precipitated with ethanol and a portion was analysed on methylmercury denaturing gels as described above. They were subsequently tailed with dCMP residues using terminal transferase as described by Cann et al. (1983). Similarly the PstI-cleaved vector pBR322 was tailed with dGMP residues. Equimolar amounts of tailed hybrid and plasmid were co-precipitated with ethanol, suspended in 10 mM-Tris-HCl pH 7.8, 150 mM-NaCl, 1 mM-EDTA and incubated at 68 °C for 15 min followed by slow cooling to room temperature. The annealed DNA was used to transform *Escherichia coli* ED8767 and tetracycline-resistant, ampicillin-sensitive colonies were screened further by colony hybridization (Grunstein & Hogness, 1975) with 32P-labelled cDNA probes prepared by reverse transcription of BEV RNA using random pentanucleotide primers.

**DNA sequencing.** Inserts from recombinant plasmids were digested with various restriction enzymes and inserted by ligation into the replicative form (RF) of M13mp18 or mp19 bacteriophage (Messing & Vieira, 1982; Yanisch-Perron et al., 1985), cleaved at compatible sites. Alternatively plasmid DNAs were cleaved randomly with DNase I (Anderson, 1981) and fragments of 300 to 1000 bp eluted from agarose gels. They were repaired with T4 DNA polymerase in the presence of the four deoxynucleoside triphosphates prior to ligation to SmaI-digested M13mp8 RF. Each fragment was sequenced by the dideoxy chain termination method of Sanger et al. (1977). In certain instances the nucleotide sequences of cDNAs were determined directly from plasmids by the method of Hattori & Sakaki (1986) using synthesized BEV-specific primers.

**Sequence analysis.** DNA sequences from individual M13 or plasmid clones were compiled and analysed using an IBM AT personal computer and the Microgenie suite of programs developed by Queen & Korn (1984).

### RESULTS AND DISCUSSION

**Synthesis of cDNA**

The RNA extracted from BEV was analysed on methylmercuric hydroxide agarose gels and visualized by staining with ethidium bromide. All preparations used in this study showed a single discrete band with no evidence of degradation. Following denaturation with methylmercuric hydroxide, this RNA was reverse-transcribed using oligo(dT)12-18 as primer. Yields of cDNA, calculated from the incorporation of [3H]dGTP, varied from 30 to 50%. When electrophoresed in parallel to the RNA on denaturing gels the majority of the cDNA migrated to a similar position as a sharp band, indicating that the principal product was full-length (Fig. 1). If prior denaturation of the template was not carried out then a number of subgenomic sized bands were observed perhaps due to the enzyme stopping at a number of places because of secondary structure.

The RNA/cDNA hybrids were tailed directly with dCMP residues and annealed to dG-tailed PstI-restricted pBR322. The recombinant mixture was used to transform competent *E. coli* ED8767 cells. Colonies with the correct antibiotic sensitivity phenotype were subjected to hybridization with BEV-specific probes. Over 100 colonies showing a strong signal were selected and small scale plasmid preparations made from 1 ml cultures. Inserts excised with PstI ranged in size from 0.2 to 5.5 kb on agarose gels. The largest clones were restriction enzyme-mapped and aligned using common restriction sites, cross-hybridization and sequence data. The restriction enzyme map is shown in Fig. 2.
Fig. 1. Analysis of BEV RNA and cDNA on methylmercuric hydroxide gels. Lane 1, ribosomal RNA markers; lane 2, \[^{3}H\]uridine-labelled RNA from purified virions; lane 3, cDNA produced by reverse transcription of virion RNA in the presence of \[^{3}H\]dGTP.

![Image of gel analysis](image)

Fig. 2. Restriction endonuclease cleavage map of the BEV genome. The clones shown were used for the determination of the genome sequence. P, PstI; B, BamHII; E, EcoRI; Xb, XhoI; Ek, EcoK; Xh, XhoI; H, HindIII; Sa, SacI; Sc, ScaI.

![Image of restriction map](image)
Fragments generated from BEV cDNA inserts were subcloned into M13 bacteriophage vectors and sequenced using the dideoxy chain termination method of Sanger et al. (1977). One stretch of the genome of approximately 100 nucleotides proved consistently difficult to subclone, and was therefore sequenced directly from two plasmids, pBEV 62 and pBEV 70, using a synthesized BEV-specific primer. This sequence was shown later to contain an EcoK site. Two plasmids, pBEV 11 and pBEV 73, were found to contain a poly(A) tract and were assumed to extend to the 3' end of the genome. Two others, pBEV 62 and pBEV 83, shared the same sequence at one end of the insert. The extensive homology of this region to the 5' end of the RNAs of polioviruses and HRV14 indicated that it represented the 5' end of the BEV genome. Thus we believe that we have the complete genome cloned. The sequence of this and of the predicted translational product are shown in Fig. 3. All areas of the genome were sequenced at least three times and from at least two plasmids. The majority was determined from both strands. The complete genome is 7414 nucleotides long and contains a large open reading frame coding for a polyprotein of 2175 amino acids starting from an AUG codon at position 819 and stopping at a termination site 71 nucleotides from the poly(A) tract.

5' Non-coding region

All picornavirus RNAs sequenced so far contain a long 5' untranslated region, the exact function of which is not yet known. That of BEV (818 bases) is longer than that of other enteroviruses. The corresponding regions in PV1 and CB1 are approximately 740 nucleotides whereas that of HRV14 is 610 nucleotides. However there is substantial sequence homology between the extreme 5' end of these virus RNAs, with 75 to 80% homology occurring over the first 94 nucleotides. At this point BEV seems to have an insertion of approximately 100 nucleotides before the sequences align again. Lindberg et al. (1987) have also observed sequence variation in this region of the genome. They noted that poliovirus had a 5 base deletion and HRV14 a 17 nucleotide insertion relative to coxsackievirus B3 (CB3) and speculated that such variation might determine the special character of each virus. The insertion in BEV at this point is unusually long, being of the order of 100 nucleotides, but the fact that exactly the same sequence was obtained from two plasmids indicates that it was not an artefact of the cloning procedure. Moreover a plasmid containing the whole genome derived from two of the clones sequenced is infectious when transfected into tissue culture cells. Therefore the insertion is likely to be of viral origin. Allowing for small length variations, nucleotides 204 to 685 of BEV align with nucleotides 93 to 603 of PV1 and CB1 with approximately 60% homology and with nucleotides 102 to 685 of HRV14 with 50% homology. The region from nucleotides 590 to 685 (in BEV) is particularly highly conserved, implying an important functional role for this region of the genome, perhaps for packaging or for ribosome binding. Thereafter there is considerable divergence of sequence between BEV, PV1 and CB1 although there is approximately the same number of bases (130) before the translation initiation site. This 130 base variable region appears to be absent from HRV14 and HRV2 (Skern et al., 1985), and might therefore be a common distinguishing feature of enteroviruses.

The coding region

Cleavage of the polyprotein

The methionine codon which initiates translation occurs at position 819 and gives rise to an open reading frame coding for 2175 amino acids, sufficient to give rise to all the virus-induced polypeptides. The gene order is the same as that of other enteroviruses and considerable homology exists at the amino acid level between BEV, PV1, CB1 and HRV14. This homology has allowed us to predict probable cleavage sites between the majority of the virus proteins (Table 1). We have been unable to predict precisely the boundary between VP3 and VP1, but have indicated possible sites. The majority of other sites cleaved are Q/G bonds, especially in the P3 region of the polyprotein, and are probably mediated by the viral protease 3C. However the cleavage site between P1 and P2 is predicted to be Y/G as in poliovirus where it is thought to be
Table 1. Comparison of the predicted cleavage sites of the BEV polyprotein with those of PV1, CB1 and HRV14

<table>
<thead>
<tr>
<th>Proteins</th>
<th>BEV</th>
<th>PV1</th>
<th>CB1</th>
<th>HRV14</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP4/VP2</td>
<td>K/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
</tr>
<tr>
<td>VP2/VP3</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/T?</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/T?</td>
</tr>
<tr>
<td></td>
<td>Q/N?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP1/2A</td>
<td>Y/G</td>
<td>Y/G</td>
<td>T/G</td>
<td>Y/G</td>
</tr>
<tr>
<td>2A/2B</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/S</td>
<td>Q/G</td>
<td>Q/N</td>
<td>Q/A</td>
</tr>
<tr>
<td>2C/3A</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>3A/VPg (3B)</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>VPg/protease (3C)</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
</tbody>
</table>

Table 2. Amino acid sequence homologies (%) of BEV proteins with those of PV1, CB1 and HRV14

<table>
<thead>
<tr>
<th>Protein</th>
<th>BEV/PV1</th>
<th>BEV/CB1</th>
<th>BEV/HRV14</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP4</td>
<td>60</td>
<td>58</td>
<td>42</td>
</tr>
<tr>
<td>VP2</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>VP3</td>
<td>41</td>
<td>47</td>
<td>44</td>
</tr>
<tr>
<td>VP1</td>
<td>33</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>2A</td>
<td>56</td>
<td>62</td>
<td>43</td>
</tr>
<tr>
<td>2B</td>
<td>43</td>
<td>63</td>
<td>42</td>
</tr>
<tr>
<td>2C</td>
<td>58</td>
<td>63</td>
<td>54</td>
</tr>
<tr>
<td>3A</td>
<td>50</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td>VPg</td>
<td>26</td>
<td>30</td>
<td>43</td>
</tr>
<tr>
<td>Protease (3C)</td>
<td>57</td>
<td>59</td>
<td>47</td>
</tr>
<tr>
<td>Polymerase (3D)</td>
<td>69</td>
<td>70</td>
<td>61</td>
</tr>
</tbody>
</table>

mediated by protease 2A intramolecularly immediately after this protein is synthesized (Toyoda et al. 1986). The close homology that exists and the conservation of the sequence PGDCGGLLRCFHG at the suggested catalytic site would imply that 2A serves a similar function in BEV polyprotein processing. The last bond to be cleaved occurs on virus maturation when VP0 is processed to VP4 and VP2 by breakage of a K/S bond in BEV compared to a N/S bond in PV1 and HRV14. This cleavage is thought to be autoproteolytic, occurring when the RNA genome enters the capsid shell (Rossmann et al., 1985; Hogle et al., 1985). A serine residue at position 10 in the VP2 of HRV14 has been implicated in the catalytic process (Rossmann et al., 1985) and is conserved in all of the viruses compared, including BEV.

*Comparison with other picornaviruses*

The degree of homology shared by each BEV protein with its PV1, CB1 and HRV14 counterpart is shown in Table 2 and the high levels of conservation observed would indicate that these viruses have probably stemmed from a common ancestor. The homology levels vary for different proteins giving some indication of which have been able to tolerate most changes during the evolutionary process.

*The P1 region*

The four structural proteins of the virion are derived from the P1 region. X-ray crystallographic studies have indicated that the three-dimensional structure of the polioviruses
Fig. 3. The nucleotide sequence of the BEV genome and the predicted amino acid sequence of the polyprotein. The proposed cleavage sites, determined by alignments with PV1, CBl and HRV14, are shown by arrows.
Fig. 4. Diagrammatic comparison of the structural proteins of BEV, PV1, CBI and HRV14. (a) VP1, (b) VP2, (c) VP3 and (d) VP4. Black areas show completely conserved amino acid sequences. The position of $\alpha$ helices and $\beta$ sheets and the relative sizes of loop structures are marked. These were determined from sequence alignments and from the X-ray crystallographic data obtained for PV1 (Hogle et al., 1985) and HRV14 (Rossmann et al., 1985).

and HRV14 are very similar (Hogle et al., 1985; Rossmann et al., 1985) and this is probably the case for BEV as well. The close homology between the three viruses has allowed us to align amino acid sequences and to map those of BEV onto the schematic representations of the four structural proteins of PV1 and HRV14. In general those sequences which form the core of the proteins are more conserved than those which occur at the surface of the virion where presumably changes can be accommodated more easily. VP4 and VP2 are the two most conserved polypeptides. VP4 has been shown to be located internally in the picornavirus capsid structure (Rossmann et al., 1985; Hogle et al., 1985) and therefore probably receives little immune pressure to mutate. VP2 has been shown to contain antigenic sites in both PV1 and HRV14 (Minor et al., 1986; Sherry et al., 1986). These sites occur on similarly located puffs which protrude from the surface of the virion. Interestingly this puff is much smaller in BEV.

The first 134 N-terminal amino acids align for all three viruses as do the last C-terminal 90. There are only 25 amino acids in the puff between these two regions in BEV compared to 48 in PV1, 42 in CBI and 39 in HRV14 (Fig. 4b). Whether these amino acids are antigenically important in BEV remains to be determined.

The other two proteins, VP1 and VP3, are less well conserved although in each case there are considerable blocks of homology shared between the viruses. These probably have structural importance. In VP3 there are two markedly conserved regions. One, an eight amino acid stretch AYTPPGGD (starting at residue 132 in BEV), has been shown to be situated near the receptor canyon floor of HRV14 and may be necessary for the topology of this structure (Rossmann et al., 1985). The other conserved region is a 17 amino acid stretch (residues 146 to 162 in BEV). This
may play an essential role in the core structure of VP3 or be important in protein interactions that occur when the virion is assembled.

VP1 is the least conserved of the viral proteins and BEV VP1 shares only 35% homology with those of the other viruses. This protein has been shown to contain important antigenic sites in both PV1 and HRV14 (Minor et al., 1986; Sherry et al., 1986), again supported on loop structures protruding from the surface. As expected the amino acids located at these and corresponding sites were found to be different and the loops varied in length when the four viruses were compared (Fig. 4a). The small protrusion which contains site 1 in poliovirus and NIm1A in HRV14 is absent in BEV. However, the major protrusion from the surface between β-sheets G and H is the same size as that in CB1 and HRV14 but smaller than that of PV1. In addition part of the C-terminal region of VP1 is thought to line the receptor wall and therefore would be expected to vary in viruses with different host ranges. The major conserved region is a 10-amino acid stretch MYVPPGAPVP (residues 149 to 158 in BEV) which again has been assigned to the receptor floor in HRV14 (Rossmann et al., 1985).

The P2 region

Of the three proteins derived from the P2 region only 2A has been assigned a definite function as a second protease in poliovirus (Toyoda et al., 1986), and is more conserved between BEV and the enteroviruses than between BEV and HRV14. This is also the case with 2C which has been implicated in replication and with guanidine sensitivity (Pincus et al., 1986). The function of 2B remains obscure and this polypeptide shows the least homology in the P2 region when that of BEV is compared with PV1 and HRV14. There is however a marked higher level of homology with the 2B of CB1 which, together with the overall higher homology, might indicate that BEV is most closely related to this virus in evolutionary terms.

The P3 region

The polymerase protein 3D of BEV is the most conserved of all the viral proteins, sharing homologies of 60 to 70% with those of the other viruses. Hydrophobicity plots indicate that the overall structure of these proteins are very similar (data not shown), implying that during evolution very little change has been tolerated probably due to strict functional enzymic requirements. This may also be the case with 3C, the major viral protease, which is also well conserved. Two stretches of amino acids are common to the proteases of all four viruses. These are GQCGGV (residues 145 to 150 in BEV) and HVGGNG (residues 161 to 166) and both have been implicated in the active centre of poliovirus 3C, which is considered to be a cysteine protease (Argos et al., 1984). The 3A of BEV which does not have a known enzymic function is less well conserved than 3C or 3D, but with all three proteins BEV shows a closer relationship to the enteroviruses than to HRV14. The VPg protein however shares little homology with those of the other viruses although there are six common amino acids which can be aligned in all four (Fig. 5). Interestingly BEV VPg lacks an alanine residue at position 2, which is present in all the other enteroviruses but has a proline residue in common with HRV14. A proline residue also occurs at this position in the HRV2 VPg sequence (Skern et al., 1985).

The 3' non-coding region

The coding region of BEV ends with a single termination signal, UGA, 71 nucleotides short of the poly(A) tract. The length of the non-coding region is similar to that of PV1 (72 bases), but shorter than that of CB1 (102 bases) and longer than that of HRV14 (46 bases). The function of this region of the genome is unknown but it might be expected that if important signals were present, e.g. for polymerase binding or for packaging, then these might be conserved between the four viruses. Comparison of the data reveals that there are no nucleotide sequences of longer than 6 bases common to all. However the hexanucleotide UUAAUU occurs in the 3' end of all four virus RNAs although at different locations. Whether this sequence has an important function remains to be determined. Iizuka et al. (1987) have demonstrated that possible secondary structures occur in the 3' non-coding regions of CB1, CB3 and PV1 (Sabin). The nucleotides occurring in the stems of these structures are largely conserved in these three viruses
Fig. 5. Comparison of the predicted amino acid sequence of the VPg protein of BEV, PV1 (Kitamura et al., 1981), CB1 (Lindberg et al., 1987) and HRV14 (Stanway et al., 1984). Conserved amino acids are boxed.

Fig. 6. Predicted secondary stem-loop structure of the 3' non-coding region of the BEV genome. The free energy calculated according to the rules of Tinoco et al. (1973) was found to be 46 kJ/mol.

but not in BEV. However it is possible to predict that a similar structure might exist in this virus (Fig. 6). The function of such structures is unknown, but they may be important at some stage of replication since an 8 nucleotide insertion in the potential stem and loop structure of PV1 resulted in a mutant with a temperature-sensitive phenotype (Sarnow et al., 1986).

The close sequence homologies reported in this paper for BEV and common human viruses shows how mutational events can play a role in the adaptation of viruses to different hosts and target cells. Further information on other animal picornaviruses may help to unravel the evolutionary relationships of this group of viruses and provide genetic information about the sequences that determine host specificity and virulence. Also the information provided here on the variable regions of the capsid proteins points to novel strategies to be made for site-specific conversion of epitope sites that may allow the construction of hybrid viruses which could be useful as live vaccines. We have recently constructed an infectious cDNA clone of BEV (to be reported elsewhere) and are currently investigating these possibilities.

We would like to thank Miss Carol Lyons for assistance with tissue culture work and the Agriculture and Food Research Council for grant support during this study. C. S Fleming and R. A. Skuce were supported by D.E.N.1. scholarships.

REFERENCES


Sequence of a bovine enterovirus genome


