The Complete Nucleotide Sequence of Swine Vesicular Disease Virus

BY TORU INOUE,† TAIKO SUZUKI‡ AND KIICHI SEKIGUCHI‡

(Accepted 5 December 1988)

SUMMARY

The complete nucleotide sequence of the genome of the enterovirus swine vesicular disease virus (SVDV; H/3'76) isolated from a healthy pig has been determined using molecular cloning and DNA sequencing techniques. The RNA genome was 7400 nucleotides long, excluding the poly(A) tract, and appeared to encode a single polyprotein of 2185 amino acids. The predicted amino acid sequence of the polyprotein showed close homology (around 90%) to that of the previously sequenced coxsackieviruses B1, B3 and B4, and also showed homology (around 60%) to that of poliovirus. This homology allows us to predict the possible cleavage sites of the polyprotein and to identify other features of structural and functional significance, which seem to be important to the biological integrity of the virus. A detailed analysis of homology between SVDV and coxsackieviruses shows that non-structural proteins are highly conserved whereas the structural proteins are less well conserved. The 5' and 3' non-coding regions are also conserved, although there are several divergent nucleotide stretches. These stretches may differentiate SVDV from coxsackieviruses.

INTRODUCTION

Swine vesicular disease (SVD) was first observed in Italy in 1966 (Nardelli et al., 1968). Subsequent outbreaks occurred in Hong Kong (Mowat et al., 1972), the United Kingdom (Dawe et al., 1973), a number of other European countries (Burrows et al., 1974) and in Japan (Kodama, 1976). SVD is an infectious disease of pigs characterized by the appearance of vesicles on the tongue, in the mouth and on the feet and hocks, lesions which are indistinguishable from those caused by foot-and-mouth disease virus.

The causative agent, SVD virus (SVDV) belongs to the enterovirus genus, in the family Picornaviridae, and has a close serological relationship to coxsackievirus B5 (CB5) (Graves, 1973). However, SVDV can be distinguished from CB5 by cross-neutralization and immunodiffusion tests (Brown et al., 1973, 1976). Pigs inoculated with CB5 develop antibody to both SVDV and CB5 but do not develop the disease (Garland & Mann, 1974).

Recently the complete nucleotide sequences of three coxsackie B viruses (CBs), CB1, CB3 and CB4 have been determined (Iizuka et al., 1987; Lindberg et al., 1987; Jenkins et al., 1987). Like other picornaviruses, the genome of the CB is a single-stranded (ss) RNA with a positive polarity, and is composed of approximately 7400 nucleotides. A small protein, VPg, is attached at the 5' terminus, and a poly(A) tract is located at the 3' terminus. Little information is available on the nucleotide sequence of the SVDV genome except for 75 nucleotides at the 5' terminus (Porter et al., 1978).

In this study, we have cloned the genome of SVDV into a plasmid and determined its complete nucleotide sequence. We also carried out comparisons of its nucleotide and deduced amino acid sequences with the sequence of CBs and some other picornaviruses. The results indicate that SVDV is closely related to the CBs.
METHODS

Virus and cells. SVDV (H/3 '76) was isolated from the faeces of a clinically healthy pig in Hokkaido, Japan in 1976 by Dr M. Kodama (Kodama et al., 1980a,b). The virus was less pathogenic for pigs than isolates from diseased pigs, and vesicle formation was restricted to the sites of inoculation. The virus was plaque-purified twice and grown in IBRS-2 cell monolayers in roller bottles.

Virus purification. Infected cells were harvested 16 h post-infection when cells exhibited c.p.e. Virus was purified by a modification of the procedure described by Nomoto et al. (1979). Briefly, the clarified infected culture medium was made up to 2% Sarkosyl and 20 mM-disodium EDTA and centrifuged for 2 h at 27000 r.p.m. at 5 °C. The pellet was suspended in 0.05 M-Tris–HCl pH 7.4, 0.3 M-NaCl, 1 mM-disodium EDTA, 1% Sarkosyl, and the suspension was sedimented through 30% glycerol containing 0.3 M-NaCl, 0.01 M-Tris–HCl pH 7.4, 1 mM-disodium EDTA, 1% Sarkosyl for 2 h at 35000 r.p.m. at 5 °C. The pellet was resuspended in TNE (0.01 M-Tris–HCl pH 7.4, 0.1 M-NaCl, 1 mM-disodium EDTA) and banded by isopycnic centrifugation in CsCl (ρ = 1.34 g/ml) for 16 h at 40000 r.p.m. at 7 °C.

RNA extraction and characterization. RNA was obtained from purified virus by the method of Grubman et al. (1979) with several modifications. The purified virus was treated with protease K/SDS and extracted with phenol/chloroform. The extracted RNA was recovered by precipitation in ethanol, and the viral RNA was purified by sucrose density gradient centrifugation and oligo(dT)-cellulose chromatography. Viral RNA was subsequently analysed by 1% agarose gel electrophoresis. The fundamental integrity of the RNA was shown by an infectious RNA assay on monolayers of IBRS-2 cells (Grubman et al., 1979).

Preparation of SVDV cDNA clones. Double-stranded (ds) cDNA was synthesized from viral RNA using a cDNA synthesis kit (Amersham). The first strand RNA–cDNA hybrid was primed with oligo(dT) and synthesized using reverse transcriptase. The second strand was synthesized by replacing the RNA strand of the RNA–cDNA hybrid using RNase H and DNA polymerase I. T4 DNA polymerase was subsequently added to the reaction mixture to remove any small 3' overhang on the first strand cDNA. The ds cDNA was phenol-extracted and tailed with oligo(dC) using terminal transferase, as described by Deng & Wu (1981).

The dC-tailed ds cDNA was annealed to dG-tailed PstI-cleaved pBR322 cloning vector (Bethesda Research Laboratories; BRL). The annealed DNA was then used to transform competent cells of Escherichia coli strain HB101 (BRL).

Restriction fragments of the ds cDNA obtained by digestion with appropriate restriction enzymes were purified by 1% agarose gel electrophoresis and inserted into the corresponding site of the M13 cloning vector. The ligated DNA was used to transform competent E. coli JM109 cells.

The recombinant clones were screened by colony and Southern blot hybridizations (Maniatis et al., 1982) using oligo(dT)-primed [32P]-labelled single-stranded (ss) cDNA as a probe. Recombinant plasmids were isolated from strongly hybridizing colonies (Birnboim & Doly, 1979) and further characterized by restriction enzyme mapping. For the genome region close to the 5' terminus, a synthetic oligonucleotide (20-mer; nucleotides 665 to 684, Takara Syuzo) was used as a primer to synthesize ds cDNA.

Nucleotide sequence analysis. Inserts from recombinant plasmids were digested with various restriction enzymes, and each fragment was subcloned into M13mp18 or mp19 replicative form DNA cleaved at compatible sites. Subclones that harboured a large insert were treated with exonuclease III and S1 nuclease to construct stepwise deletion sets (Henikoff, 1984). Each fragment was sequenced by the dideoxynucleotide chain termination method of Sanger et al. (1977).

The primer extension method was employed for sequencing the extreme 5' terminal region of the genome, whose sequence was not contained in any of the cloned cDNAs obtained. A synthetic oligonucleotide that was complementary to a region of the viral RNA (nucleotides 91 to 110) was radiolabelled using polynucleotide kinase and [γ-32P]ATP (Geliebter, 1987). RNA sequencing was carried out by the dideoxynucleotide chain termination method using avian myeloblastosis virus reverse transcriptase. Alternatively, the [32P]-labelled oligonucleotide primer was annealed to viral RNA and a primer extension reaction was carried out using reverse transcriptase. The resulting cDNA product was separated on an 8% denaturing polyacrylamide gel and sequenced by the method of Maxam & Gilbert (1977).

Computer analysis. Nucleotide sequences were compiled and analysed using an NEC PC9801 personal computer and the Microgenie and Genetyx suites of programs.

RESULTS AND DISCUSSION

Cloning of the SVDV genome

The extracted SVDV RNA, analysed by 1% agarose gel electrophoresis, showed a species of the expected size, and the biological integrity of the RNA was demonstrated by generating infectious SVDV (6 × 10^5 p.f.u./μg RNA). The viral RNA was converted to ds cDNA as described in Methods; the resulting cDNA was near full-length.
Fig. 1. Restriction cleavage sites on SVDV and strategy for sequence analysis. Genome organization and nucleotide length of the SVDV genome are shown at the top of the figure (scale bar is in kb). Restriction cleavage sites of SVDV cDNA and cDNA clones obtained are shown below the SVDV genome. Open triangles on SVDV cDNA lane indicate the site of synthetic oligonucleotide primers. Arrows in brackets indicate the regions where nucleotide sequences of cloned cDNA were analysed. The dots by the arrows represent the restriction sites used for subcloning, the bars represent the sites chosen for the stepwise deletion sets. The closed triangle at the bottom of the figure represents the nucleotide sequence determined by using primer-extended products.

The ds cDNA was tailed with dCMP residues and annealed to dG-tailed PstI-cleaved pBR322. The recombinant mixture was used to transform competent E. coli HB101 cells. Thirty-five ampicillin-sensitive colonies were obtained. The harboured plasmids were cut with PstI and the released fragments were analysed by Southern blot hybridization using SVDV ss cDNA as a probe. Finally, two clones (PBRS14 and PBRS19) were obtained whose inserts were 5.5 kb and 0.8 kb respectively. Clone PBRS19 contained the 3' poly(A) tail consisting of 47 AMP residues.

Clone MPS453 which spans the genomic gap between PBRS14 and PBRS19 was prepared by ligation of an SphI/SacI fragment of ds cDNA to the M13 cloning vector replicative form carrying compatible cohesive ends.

To obtain the clone MPS671 covering the 5' end of the viral genome, ds cDNA was prepared by the use of a synthetic oligonucleotide (nucleotides 665 to 684) as a primer. An EcoRI linker was added to each terminus of the ds cDNA and digested with EcoRI and PstI. The resulting fragments carrying one EcoRI cohesive end and one PstI cohesive end were ligated to the M13 cloning vector carrying compatible cohesive ends. These cDNA clones covered the whole viral genome except for 24 nucleotides at the extreme 5' terminus. These clones and ds cDNA were mapped with restriction enzymes and aligned using restriction sites and sequence data. The restriction map is shown in Fig. 1 together with the sequencing strategy.

**Nucleotide sequence**

Fragments generated from cDNA inserts were subcloned into M13 cloning vectors and sequenced using the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). To obtain the sequence of the extreme 5' terminal region of the genome, primer extension
sequencing was performed using a \(^{32}\)P-labelled synthetic primer (nucleotides 91 to 110) as described in Methods.

The complete nucleotide sequence and predicted amino acid sequence of SVDV genome are shown in Fig. 2. The total size of the SVDV genome excluding the poly(A) tract is 7400 nucleotides. The base composition of the SVDV genome without the poly(A) tail showed a high adenine content (A, 28.1%; G, 25.2%; U, 22.6%; C, 24.1%) and a low frequency of the dinucleotide C-G (3.7%), which are both consistent with other enteroviruses (Toyoda et al., 1984; Iizuka et al., 1987).

Computer-aided translation of the RNA sequence in all three reading frames revealed a single long open reading frame (ORF) consisting of 6555 nucleotides (2185 triplet codons). The ORF starts at an AUG codon (nucleotide position 743) and ends at a UAA codon (nucleotide position 7298).

Comparisons among the genomes of CB1, CB3 and CB4 were carried out at the nucleotide sequence level for the non-coding regions, and at the amino acid sequence level for the coding regions, and these are shown schematically in Fig. 3 and 4 respectively. The predicted amino acid sequence of the coding region of SVDV exhibits a high degree of homology to CBs (90.0%, 88.1% and 88.3% to that of CB1, CB3 and CB4 respectively) and also shows homology to poliovirus type 1 (PV1) (57.5%). This high degree of homology has enabled us to predict probable cleavage sites between the majority of the viral proteins. All of the proteolytic cleavage sites in PV1 and CBs are conserved in SVDV with the exception of that at the 1D/2A and 2B/2C junctions (Table 1). For the 1D/2A junction, a T/G pair in SVDV is conserved in CB1 but not in PV1 and CB4. For the 2B/2C junction, Q/N in SVDV is conserved in CB1 and CB4 but not in PV1. Lindberg et al. (1987) have proposed that in CB3 the 2B/2C cleavage occurs either at a Y/G or a Q/N pair. We therefore conclude that in SVDV Q/N and T/G pairs probably function as the cleavage sites for 2B/2C and 1D/2A junctions, respectively, although N- and C-terminal protein sequence analysis is required in order to identify the cleavage sites precisely.

Fig. 2. The nucleotide sequence of the SVDV genome and the predicted amino acid sequence. The RNA sequence is deduced from the DNA sequence of cDNA. The putative cleavage sites determined by alignment with CBs and PV1 are shown by arrows.
Fig. 3. Alignment comparison of nucleotide sequence homology of the non-coding regions of SVDV and CBs genomes [the details for CB1 are taken from Iizuka et al. (1987), for CB3 from Lindberg et al. (1987) and for CB4 from Jenkins et al. (1987)]. The nucleotides are arranged to obtain the highest level of homology. Positions of common sequences among SVDV and CBs are indicated as closed boxes. Positions of common sequences only among CBs are indicated as open boxes. The positions of nucleotides that are common among CBs but not SVDV are indicated by small arrows. The uppermost band indicates homologous regions in which more than one nucleotide is common, the second more than three, the third six, the fourth nine and the fifth 12 nucleotides. Nucleotide numbers are indicated above the uppermost bands. The comparison was performed using the Microgenie program.

Table 1. Comparison of the predicted cleavage sites of the SVDV polyprotein with those of CB1, CB3, CB4 and PV1*

<table>
<thead>
<tr>
<th>Boundary</th>
<th>SVDV</th>
<th>CB1</th>
<th>CB3</th>
<th>CB4</th>
<th>PV1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A (VP4)/1B (VP2)</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
</tr>
<tr>
<td>1B (VP2)/1C (VP3)</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>1C (VP3)/1D (VP1)</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>1D (VP1)/2A (protease?)</td>
<td>T/G</td>
<td>T/G</td>
<td>Q/S?</td>
<td>Y/G</td>
<td>Y/G</td>
</tr>
<tr>
<td>2A (protease?)/2B</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>Q/G</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>
</tr>
<tr>
<td>3A/3B (VPg)</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>3B (VPg)/3C (protease)</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
<tr>
<td>3C (protease)/3D (polymerase)</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
<td>Q/G</td>
</tr>
</tbody>
</table>

* Details taken from: Iizuka et al. (1987) for CB1; Lindberg et al. (1987) for CB3; Jenkins et al. (1987) for CB4; Kitamura et al. (1981) for PV1.

The 5' non-coding region

The 5' non-coding region of SVDV consists of 742 nucleotides, a number comparable to that of PVs (742 to 744) and CBs (740 to 743). The overall homologies of this region to CBs and PVs are 82 to 83% and 68 to 70% respectively (Table 2). An alignment comparison of SVDV with three CBs indicates similar features to those shown by PVs (Toyoda et al., 1984). From these features, the 5' non-coding region can be divided into three classes of sub-sequence, namely
Fig. 4. Predicted amino acid sequence homology of polypeptides among SVDV and CB genomes, represented as in Fig. 3. Conserved amino acids among SVDV and CBs are indicated as closed boxes. Amino acid sequence blocks of homology only among CBs are indicated as open boxes. Positions of amino acids that are common to CBs but not SVDV are indicated by small arrows at the bottom. The uppermost band indicates homologous regions in which more than one amino acid is common, the second more than two, the third more than four, the fourth more than six, and the fifth more than eight. The nucleotide and amino acid residue numbers are indicated above and below the scale bar respectively. Proposed functionally important conserved sequences are shown below P2 and P3 (see text). Details are taken from the same source as those in Fig. 3.
Table 2. Nucleotide sequence homology between non-coding region of SV DV and other picornaviruses*

<table>
<thead>
<tr>
<th>Nucleotide region</th>
<th>CB1</th>
<th>CB3</th>
<th>CB4</th>
<th>PV1</th>
<th>PV2</th>
<th>PV3</th>
<th>BEV</th>
<th>HRV2</th>
<th>HRV14</th>
<th>HAV</th>
<th>EMCV</th>
<th>TMEV</th>
<th>FMDV</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' Non-coding</td>
<td>82.9</td>
<td>83.6</td>
<td>82.0</td>
<td>67.9</td>
<td>69.0</td>
<td>69.7</td>
<td>56.1</td>
<td>53.7</td>
<td>54.0</td>
<td>42.5</td>
<td>-</td>
<td>-</td>
<td>31.8</td>
</tr>
<tr>
<td>3' Non-coding</td>
<td>79.8</td>
<td>76.9</td>
<td>81.9</td>
<td>37.9</td>
<td>36.9</td>
<td>35.9</td>
<td>33.7</td>
<td>24.0</td>
<td>16.5</td>
<td>33.0</td>
<td>36.6</td>
<td>31.2</td>
<td>35.8</td>
</tr>
</tbody>
</table>

* Sequence homology is expressed as percentages. The percentage homology among CBs and PVs of the 5' non-coding regions are 87.2 ± 5.7 and 83.1 ± 2.3 respectively; 3' non-coding regions are 89.6 ± 3.5 and 98.1 ± 0.8 respectively. Details of the other virus sequences are taken from: Iizuka et al. (1987) for CB1; Lindberg et al. (1987) for CB3; Jenkins et al. (1987) for CB4; Kitamura et al. (1981) for PV1; Earle et al. (1988) for BEV; Callahan et al. (1985) for HRV14; Skern et al. (1985) for HRV2; La Monica et al. (1986) for PV2; Stanway et al. (1984b) for PV3; Fors et al. (1984) for FMDV; Palmenberg et al. (1984) for EMCV; Pevear et al. (1987) for Theiler's murine encephalomyelitis virus (TMEV); Najarian et al. (1985) for hepatitis A virus (HAV).

highly conserved regions (nucleotides 1 to 90, 430 to 490 and 510 to 564), highly variable regions (nucleotides 120 to 170 and 675 to 742) and the remainder, relatively variable regions (Fig. 3).

The highly conserved regions demonstrate much higher homology than the overall homology and are considered to be important to some aspects of viral replication (Toyoda et al., 1984; Kuge & Nomoto, 1987). The highly variable regions show a low degree of homology between SVDV and CBs. The region next to the VP4 coding region (nucleotides 675 to 742) has no SVDV-specific nucleotides that are different from the nucleotides conserved among CBs (indicated by small arrows at the bottom of Fig. 3). The corresponding region is low in homology among the genomes of CBs and PVs and is missing in the genome of human rhinoviruses (HRVs) (Callahan et al., 1985; Skern et al., 1985; Stanway et al., 1984a). This region of the genome has been considered to be a possible spacer (Toyoda et al., 1984). Mutations (insertion or deletion) in this region had no effect on viral replication and therefore it may not play an essential role in replication in cell culture (Kuge & Nomoto, 1987). Another highly variable region (nucleotides 120 to 170) includes several SVDV-specific nucleotides. In the relatively conserved region most of the common sequences are conserved as short blocks. However, comparison of homology among CBs except SVDV shows additional blocks of homology. These blocks are indicated as open boxes in Fig. 3 and seem to represent differences between SVDV and CBs. The relatively variable region containing these blocks may determine some specific characteristics that distinguish SVDV from CBs.

Two specific nucleotide sequences, the U-rich region (nucleotides 562 to 576) and the eight base consensus sequence (CUUAUGGU; nucleotides 587 to 594) are conserved in this region and might act as important signals to maintain the rates of viral replication (Kuge & Nomoto, 1987).

The predicted stem–loop structure close to the 5' end of the RNA of CB3 (Tracy et al., 1985) is conserved in SVDV. More RNA secondary structure was predicted by a computer search. There are nine potential translation start codons positioned before the codon that initiates the large ORF at nucleotide position 743. However, the potential translation products initiated from these AUGs are unlikely to function in viral replication because of the short ORF (two to 50 amino acids), the lack of conservation in the size and position among other picornaviruses, and the absence of efficient flanking sequences for ribosome recognition (Kozak, 1986).

Structural proteins

The P1 region of the polyprotein consists of four capsid proteins, 1A (VP4), 1B (VP2), 1C (VP3) and 1D (VP1). The amino acid sequence similarities between the predicted protein of SVDV and those of CBs, PVs and other picornaviruses are shown in Table 3. The comparative amino acid sequence homology of the whole polyprotein between SVDV and CBs is shown in Fig. 4. The homology of the P1 region to other enteroviruses is lower than that of the P2 or P3 regions. VP4 is the most conserved of the structural proteins; it is not exposed on the outer surface of the virion and there is probably little immune pressure to mutate. In the other three
Table 3. Amino acid sequence homology between the proteins of SVDV and other picornaviruses*

<table>
<thead>
<tr>
<th>Protein</th>
<th>CB1</th>
<th>CB3</th>
<th>CB4</th>
<th>PV1</th>
<th>PV2</th>
<th>PV3</th>
<th>BEV</th>
<th>HRV2</th>
<th>HRV14</th>
<th>HAV</th>
<th>EMCV</th>
<th>TMEV</th>
<th>FMDV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A (VP4)</td>
<td>92.8 ± 0.3</td>
<td>91.3</td>
<td>91.3</td>
<td>65.2 ± 0.7</td>
<td>65.2</td>
<td>66.7</td>
<td>50.7</td>
<td>56.5</td>
<td>58.6</td>
<td>-</td>
<td>14.6</td>
<td>16.3</td>
<td>12.5</td>
</tr>
<tr>
<td>1B (VP2)</td>
<td>80.6 ± 0.6</td>
<td>80.6</td>
<td>82.8</td>
<td>50.2</td>
<td>55.4</td>
<td>52.6</td>
<td>52.7</td>
<td>52.8</td>
<td>65.2</td>
<td>-</td>
<td>25.8</td>
<td>31.0</td>
<td>24.4</td>
</tr>
<tr>
<td>1C (VP3)</td>
<td>84.5 ± 0.4</td>
<td>82.8</td>
<td>78.6</td>
<td>52.9</td>
<td>54.6</td>
<td>54.6</td>
<td>46.5</td>
<td>50.2</td>
<td>48.5</td>
<td>-</td>
<td>29.9</td>
<td>30.6</td>
<td>26.7</td>
</tr>
<tr>
<td>1D (VP1)</td>
<td>77.1 ± 0.7</td>
<td>73.5</td>
<td>70.1</td>
<td>43.6</td>
<td>44.5</td>
<td>42.5</td>
<td>35.9</td>
<td>37.9</td>
<td>36.6</td>
<td>-</td>
<td>17.8</td>
<td>15.9</td>
<td>16.8</td>
</tr>
<tr>
<td>2A (protease?)</td>
<td>91.3 ± 0.4</td>
<td>82.1</td>
<td>86.7</td>
<td>56.3</td>
<td>57.6</td>
<td>58.3</td>
<td>62.0</td>
<td>34.7</td>
<td>42.1</td>
<td>12.7</td>
<td>10.6</td>
<td>14.4</td>
<td>7.3</td>
</tr>
<tr>
<td>2B</td>
<td>94.9 ± 0.3</td>
<td>93.9</td>
<td>93.9</td>
<td>48.5</td>
<td>51.5</td>
<td>50.5</td>
<td>63.6</td>
<td>39.4</td>
<td>55.6</td>
<td>14.6</td>
<td>18.2</td>
<td>20.5</td>
<td>18.8</td>
</tr>
<tr>
<td>2C</td>
<td>97.3 ± 0.6</td>
<td>96.4</td>
<td>97.0</td>
<td>60.9</td>
<td>63.0</td>
<td>62.4</td>
<td>51.9</td>
<td>45.8</td>
<td>58.3</td>
<td>25.2</td>
<td>30.9</td>
<td>30.4</td>
<td>29.6</td>
</tr>
<tr>
<td>3A</td>
<td>91.0 ± 0.6</td>
<td>92.1</td>
<td>91.0</td>
<td>48.3</td>
<td>50.6</td>
<td>49.4</td>
<td>48.3</td>
<td>33.7</td>
<td>51.7</td>
<td>15.7</td>
<td>12.8</td>
<td>16.0</td>
<td>12.4</td>
</tr>
<tr>
<td>3B (VPg)</td>
<td>92.5 ± 0.3</td>
<td>100.0</td>
<td>90.9</td>
<td>100.0</td>
<td>77.3</td>
<td>72.7</td>
<td>72.7</td>
<td>34.8</td>
<td>40.9</td>
<td>52.2</td>
<td>12.0</td>
<td>36.4</td>
<td>27.3</td>
</tr>
<tr>
<td>3C (protease)</td>
<td>93.9 ± 0.3</td>
<td>92.9</td>
<td>95.6</td>
<td>54.9</td>
<td>61.2</td>
<td>60.7</td>
<td>59.8</td>
<td>48.1</td>
<td>53.0</td>
<td>16.8</td>
<td>24.2</td>
<td>19.8</td>
<td>21.2</td>
</tr>
<tr>
<td>3D (polymerase)</td>
<td>95.5 ± 0.3</td>
<td>95.0</td>
<td>96.3</td>
<td>73.6</td>
<td>74.2</td>
<td>74.0</td>
<td>70.0</td>
<td>56.3</td>
<td>64.7</td>
<td>24.4</td>
<td>30.9</td>
<td>32.2</td>
<td>30.9</td>
</tr>
</tbody>
</table>

* Sequence homologies are expressed as percentages. The values in parentheses represent the percent homology among CBs or PVs. Details of the other virus sequences are taken from the same sources as indicated in Table 2.

In spite of divergence in the sequence of this region, highly conserved regions are distributed as blocks in the variable region. A similar distribution of highly conserved blocks is observed in the comparison between SVDV and PV1 (Fig. 5c). This pattern is very similar to the alignment comparison among PVs (Toyoda et al., 1984). The amino acid sequence of SVDV is aligned with that of PV1 to give maximal sequence homology. The hydrophilicities of SVDV and PV1 polyprotein sequences are represented together in Fig. 5 and are very similar. Computer-predicted protein secondary structure of SVDV is similar to that of PV1 determined by crystallography by Hogle et al. (1985). The β-sheet structures which are well conserved in SVDV and PV1 almost totally overlap on the highly conserved blocks and hydrophobic valleys of the hydrophilicity profile. The structurally essential regions, especially the β-sheet forming β-barrel structures which form the viral core structure, seem to be common to SVDV and PV1. Thus the SVDV virion seems to have a very similar structure to PV1. From the similarity of the structures, we predicted the antigenic determinant sites of SVDV on the basis of the secondary structure of PV1. These regions, which have hydrophilic and poorly conserved properties, are located on the outer surface of virion and at loop-out sequences. The antigenic sites are marked as 'cl'; other structurally important regions, such as the canyon floor which is situated near the receptor, and the hinge which connects each subunit, are also indicated in Fig. 5.
Fig. 5. Comparison of predicted structure of capsid proteins between SVDV and PV1 Mahoney (Kitamura et al., 1981). (a) Comparison of hydrophilicity between SVDV and PV1 (GENETYX:HYDO program). Solid line indicates SVDV and dotted line indicates PV1. Positive values represent hydrophilic residues and negative values represents hydrophobic ones. (b) The amino acid numbers of PV1 and SVDV are indicated above and below the line respectively. They are arranged to show amino acid homology and similarity between SVDV and PV1. Comparison was performed by the Microgenie program. (c) Predicted amino acid sequence homology of the P1 regions of SVDV and PV1. The figure is represented as in Fig. 4. Homology analysis was performed by the Microgenie program. (d) The protein secondary structure of PV1 (solid line, β-sheet; broken line, α-helix). The line represents the X-ray crystallographic analysis by Hogle et al. (1985). These data are aligned according to amino acid homology between SVDV and PV1. Structurally and antigenically important sites are also indicated. The abbreviation ‘cl’ means cluster of neutralizing antibody determinants as described by Hogle et al. (1985). Alphabetic characters indicate the β-sheet structures as described in Rossmann et al. (1985).
Non-structural proteins

The P2 and P3 regions are highly conserved among enteroviruses, especially between SVDV and CBs. Except for 2A, all the proteins have more than 90% homology (Table 3). Protein 2A has been shown to function as a protease responsible for cleavage of the polyprotein 1D/2A at the Y/G cleavage site in PV (Toyoda et al., 1986). The homology of 2A between SVDV and CB1 is higher than that of SVDV to CB3 or SVDV to CB4. This may mean that CB1 and SVDV have the same substrate cleavage site at T/G, although this is not the case for comparisons of CB4 and CB3, and CB4 and PV1 (Jenkins et al., 1987).

The region near the C-terminal of the 2A protein, PGDCGXLXCHG (residues 107 to 119), which is similar to the possible active site of protease 3C, is completely conserved among SVDV, PVs, CBs, bovine enterovirus (BEV) and HRVs (Fig. 4).

A comparison between the 2C protein of SVDV and CBs shows a remarkably high degree of homology (96 to 99%) (Table 3). It is thought that the 2C protein attaches to vesicular membranes and is associated with viral RNA synthesis (Bienz et al., 1983; Takegami et al., 1983), although its exact function is obscure.

The VPg (3B) sequence of SVDV is identical to those of CB1 and CB4 at the amino acid level. It differs from those of CB5 and CB3 by only one and two amino acids respectively (Lindberg et al., 1987).

The 3C polypeptide is a protease that cleaves at certain Q/G sites in the polyprotein. These are highly conserved sequences in the SVDV 3C region in comparison with other picornaviruses. In particular, two stretches of amino acids, GQCGGV (residues 145 to 150) and HVGGNG (residues 161 to 166), in the C-terminal region of the protein are conserved in the proteases of PVs, CBs, BEV and HRV14 (the other HRVs have YQCGGV in place of the former sequence). The Cys 147 and His 161 residues (SVDV and PV1) are conserved in picornaviruses and plant comoviruses. These residues are possibly part of the active site of the protease, which is considered to be a cysteine protease (Argos et al., 1984).

In the polymerase region, 3D, the homologies between SVDV and CBs are 95 to 96%, compared to 73-6% between SVDV and PV1. The homologous regions are distributed along the polymerase region. The amino acid sequence YGDD (residues 327 to 330) is conserved among picornaviruses including SVDV, and in plant comoviruses. The element GDD is also identified in other RNA viruses, for example vesicular stomatitis virus, the haemagglutinating virus of Japan, flaviviruses, plant and bacterial viruses and retroviruses (Kamer & Argos, 1984; Ishihama & Nagata, 1988). These stretches might be a possible active site or nucleic acid recognition region for the polymerase. Two other sequences, PSD (residues 228 to 290) and FLKR (residues 374 to 377), are conserved among picornaviruses and cowpea mosaic virus.

3' non-coding region

The translation of the polyprotein RNA is terminated at position 7298 by the sequence UAA, which is followed by a non-coding region of 102 nucleotides prior to the poly(A) tract. The 75 nucleotides next to the 3' poly(A) of the SVDV genome have been reported by Porter et al. (1978) and are identical in our strain of SVDV. The sequence homology of SVDV to other picornaviruses, except CBs, is generally low (Table 2). This is partly a reflection of the difference in sequence length. For example, the number of nucleotides of the 3' non-coding region are 70 to 71 for PVs, 42 to 46 for HRVs, and 125 for encephalomyocarditis virus (EMCV). Although the sequence homologies of SVDV to CBs having nearly the same size of 3' non-coding region (100 to 102 nucleotides) are high (76-9 to 81.9%), the homology value among the different CB serotypes (89-6%) is higher than that of SVDV to CBs.

The alignment comparison of the 3' non-coding region is shown in Fig. 3. There are some SVDV-specific regions that are different from CBs; these are indicated as open boxes as shown in the 5' non-coding region.

In the 3' non-coding region of enteroviruses, a characteristic RNA secondary structure has been reported by many investigators. For SVDV, a number of RNA secondary structures are possible in this region including part of the 3D region (about 100 nucleotides) as shown in Fig. 6. This region is composed mainly of five independent stem–loop structures (Fig. 6a). These
Fig. 6. (a) Computer-predicted RNA secondary structure of the 3' terminal region of the SVDV genome. The structure was calculated by the program GENETYX: SECST ΔG = -376.1 kJ/mol. (b) Alignment of 3' terminal region of several picornaviruses (details are taken from the sources indicated in Table 2). Deletions are marked by a dash (-) and a blank space indicates no change from the base in the SVDV. Boxes around sequences indicate bases involved in formation of the predicted stems. Sequence homology analysis was performed by the program GENETYX: MAXHAM and the RNA secondary structures are calculated by the program GENETYX: SECST.
Table 4. Free energy of computer-predicted RNA secondary structure of the 3' terminal region

<table>
<thead>
<tr>
<th>Stem-loop no.</th>
<th>SVDV</th>
<th>CB1</th>
<th>CB3</th>
<th>CB4</th>
<th>PV1</th>
<th>BEV</th>
<th>HRV2</th>
<th>HRV14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1†</td>
<td>-13-8</td>
<td>-13-8</td>
<td>-22-6</td>
<td>-31-8</td>
<td>-2-5</td>
<td>-51-9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2†</td>
<td>-102-1</td>
<td>-87-0</td>
<td>-93-3</td>
<td>-111-3</td>
<td>-40-6</td>
<td>-59-4</td>
<td>-34-7</td>
<td>-19-2</td>
</tr>
<tr>
<td>3†</td>
<td>-60-7</td>
<td>-0-8</td>
<td>-3-3</td>
<td>-18-4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4‡</td>
<td>-80-8</td>
<td>-56-5</td>
<td>-74-5</td>
<td>-25-9</td>
<td>-138-9</td>
<td>-135-6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5‡</td>
<td>-72-4</td>
<td>-41-0</td>
<td>-40-2</td>
<td>-77-8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* Each stem-loop no. correlates with Fig. 6(a). Free energy was calculated by the program GENETYX:SECST.
† 3' Non-coding region.
‡ 3D protein-coding region.

structures are also predicted for other enteroviruses and rhinoviruses. The alignment and the position of each predicted stem structure are represented in Fig. 6(b). The free energy of each stem–loop structure is shown in Table 4. The stem–loop no. 2 corresponds to the previously mentioned stem–loop structure and its free energy is very low. This structure is also observed in HRV but its free energy is not always low. The stem–loop no. 3 observed in SVDV is not obvious in CBs, and the corresponding stem–loop is deleted in PV1, BEV and HRVs. Stem–loop no. 4 and 5 in SVDV which are located in the 3D region are also observed in CBs but not in BEV, PV or HRV. The function of the 3' non-coding region has not been determined and the functional significance of its secondary structure is obscure. However, the secondary structure may contribute to RNA replication (Sarnow et al., 1986) and stabilization, such as resistance against attack by endonuclease. On the other hand, these proposed structures in the 3' non-coding region seem to reveal some differences among enteroviruses. In the picornaviruses deletions or insertions that contribute to evolutionary divergence may occur at a stem–loop structure.

SVDV has been found to be closely related to CB5 in its physicochemical, biochemical and serological characteristics, while immunodiffusion, neutralization and RNA hybridization have shown differences between them (Harris et al., 1977). CB5 shows considerable variation (Brown & Wild, 1974), but the variation of SVDV is smaller (Harris & Brown, 1975). Graves (1973) suggested that SVDV was originally derived from human CB5. Garland & Mann (1974) were critical of this hypothesis because they failed to infect SVDV-susceptible pigs with CB5. Knowles et al. (1979) proposed that SVDV should be considered a porcine strain of CB5 on the basis of production of c.p.e. and antigenic analysis. Stålhandske et al. (1984) have also considered SVDV as a variant of CB5. However, the evidence that SVDV is a variant of CB5 is inconclusive.

In this paper we have reported the complete nucleotide sequence of the genome of an SVDV strain of low pathogenicity, and have demonstrated that the genomic organization is similar to that of other picornaviruses. Comparisons of nucleotide and predicted amino acid sequence between SVDV and other picornaviruses indicate that SVDV is closely related to the other enteroviruses and rhinoviruses. The homology between SVDV and CBs is remarkably high and large parts of the sequence are conserved. Although the complete nucleotide sequence of CB5 has not been determined, our results indicate a close relationship between SVDV and CBs.

Whether SVDV is a variant of CB5 will not be known until the complete nucleotide sequences of CB5 and other strains of SVDV, including a pathogenic one, are established. Our results also suggest the presence of SVDV-specific sequence blocks, which are conserved among CBs but not in SVDV. As nucleotide sequences, these blocks are located in relatively variable regions of the 5' non-coding region and in the 3' non-coding region. As amino acid sequences, they are observed in the 2A region. The SVDV-specific blocks may hold a clue to understanding not only the relationship between SVDV and CBs but also more precise molecular features of picornaviruses.
REFERENCES


(Received 9 August 1988)