Sequence Analysis of an 11.2 Kilobase, Near-terminal, BamHI Fragment of Fowlpox Virus

By FIONA TOMLEY,* MATTHEW BINNS, JOAN CAMPBELL AND MIKE BOURSNELL

AFRC Institute for Animal Disease Research, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA, U.K.

(Accepted 27 January 1988)

SUMMARY

The nucleotide sequence of an 11.2 kilobase fragment of the fowlpox virus genome is presented. The fragment comes from near one end of the genome and contains part of the terminal inverted repeat. Twenty open reading frames (ORFs) are predicted from the sequence and are classified into 13 major and seven minor ORFs. The 100 base pairs immediately upstream of each ORF are up to 83% AT-rich, with some motifs similar to those seen in vaccinia virus early gene promoters. The TTTTTNT element which has been identified as a termination signal for vaccinia virus early genes is also found downstream of several ORFs. Three ORFs are predicted to specify polypeptides with significant homology to proteins coded by genes near termini of orthopoxvirus genomes: the vaccinia virus 42K early gene and 32-5K host range gene, and the cowpox virus 38K red pock gene. In addition, there are two families of ORFs within the fragment which potentially encode related polypeptides. One of these, family B, contains three ORFs which are related to those of chicken and rat hepatic lectins.

INTRODUCTION

The development of techniques for the insertion of foreign DNA into vaccinia virus (VV) (Mackett et al., 1982; Panicali & Paoletti, 1982; Smith et al., 1983) has led to the expression of many foreign antigens in chimeric viruses and the possibility of using these as live recombinant vaccines (for review, see Mackett & Smith, 1986). However, because of its wide host range, there are potential problems that could result from the widespread use of VV for animal vaccination. It is therefore desirable to develop other poxviruses as species-specific vectors, for example fowlpox virus (FPV) for avian vaccines.

The avipoxviruses are a separate genus from the orthopoxvirus group, which includes VV. Their genomes are considerably larger than those of the orthopoxviruses. For example, FPV contains approximately 50% more DNA than VV, and by restriction site mapping FPV shows no similarity to the orthopoxviruses (Müller et al., 1978). However, more recently, the nucleotide sequences of two FPV early genes, those for thymidine kinase (TK) and DNA polymerase (Boyle et al., 1987; Binns et al., 1987), and the sequence of a 3.1 kb fragment that hybridizes to the VV HindIII J fragment (Drillien et al., 1987) have been determined and found to be closely related at the amino acid level to the homologous orthopoxvirus sequences.

The orthopoxviruses, particularly VV, have been well studied and much is known about their genomic structure and organization. In contrast, little is known about the genomic organization and transcriptional control sequences of FPV. In order to develop a vector system based on FPV it will be necessary to identify suitable promoters and regions in its genome where foreign genes can be inserted. As an aid to this we are sequencing selected areas of the FPV genome and present here the sequence of an 11.2 kb fragment that is located near one end of the virus genome and includes part of the terminal inverted repeat (TIR).
METHODS

Virus and cells. Fowlpox virus (Avipoxvirus gallinae) isolate HP-438 (Munich) was a gift from Dr A. Mayr. The virus was maintained by low multiplicity passage in chick embryo fibroblasts (CEFs) and purified as described by Mockett et al. (1987), each passage being numbered consecutively from HP-438.

Cloning. DNA isolation was essentially as described by Mackett & Archard (1979). Briefly, virus from 25 flasks (125 cm^3) of CEFs was lysed in 10 ml of 50 mM-Tris-HCl pH 7.8, 1 mM-EDTA, 27% (w/v) sucrose, 1% (w/v) SDS, 100 mM-2-mercaptoethanol and 500 μg/ml proteinase K (Boehringer), incubated at 50°C for 1 h then kept overnight at 4°C. Deproteinized DNA was extracted by gentle swirling for several hours with an equal volume of Mockett phenol saturated with 10 mM-Tris-HCl pH 7.5 and 1 mM-EDTA, followed by two ether extractions and was redissolved in deionized water at 500 μg/ml. One μg of DNA was cut with BamHI (Bethesda Research Laboratories) and ligated into BamHI-cut phosphatase-treated pUC13 (Pharmacia). Following transformation of Escherichia coli strain TG1 using standard methods (Hanahan, 1983), white colonies from X-gal plates were transferred to nitrocellulose and probed with nick-translated FPV DNA. Plasmids which contained FPV inserts were analysed by agarose gel electrophoresis of minipreparations (Holmes & Quigley, 1981). A range of recombinant plasmids containing FPV DNA inserts was obtained. One of these, pMH23, containing an insert of approximately 11.2 kb was selected for sequencing.

Sequencing. Random fragments of pMH23 were generated by sonication of the plasmid (Deininger, 1983) and cloned into Smal-cut phosphatase-treated M13 mpi10 (Amersham). M13 clones with FPV inserts were identified by probing with nick-translated purified virus insert from pMH23 and single-stranded templates were prepared. Sequencing was carried out by the dideoxy method (Sanger et al., 1977; Bankier & Barrell, 1983) using [32P]dATP in the reactions, and the products were electrophoresed on buffer gradient gels (Biggin et al., 1983). Some sequence information was obtained by reverse sequencing (Hong, 1981).

Sequences were read directly into a BBC microcomputer using a sonic digitizer (Graf/Bar, Science Accessories Corporation) and the data were analysed on a VAX 11/750 using the programs of Staden (1982, 1984a, b, c). Comparisons with the National Biomedical Research Foundation (NBRF) protein identification resource and with our own compilation of published poxvirus sequences were made using the programs FASTP (Lipman & Pearson, 1985) and SEQHP (Kanehisa, 1982).

Mapping the end of the terminal inverted repeat. M13 clones to be used as probes were radiolabelled using sequencing reaction mixtures lacking the dideoxynucleotides and containing [32P]dCTP. Following filtration in 440 DNA blotted onto nitrocellulose (Southern, 1975).

RESULTS

Location of the fragment

A method for the rapid identification of genes within the FPV genome has been devised in this laboratory (Binns et al., 1987). It involves the generation of M13 clones containing random fragments of the FPV DNA genome. One of these clones, MFP354, was found to overlap the left end of the 11·2 kb sequence presented in this paper. This clone was used to probe Southern blots of BamHI-digested FPV DNA in order to determine which restriction fragment lies next to the 11·2 kb fragment. Three fragments of around 6 kb, 11 kb and 25 kb hybridized to MFP354 (Fig. 1). The 6 kb fragment was shown, by snap-back analysis (J. Campbell et al., unpublished data) to contain cross-links typical of the terminal fragments of other poxviruses. It was the only snapped-back fragment observed in BamHI digests and thus should be present at both ends of the genome. It is known that poxviruses contain TIRs which vary in size from virus to virus and which undergo rearrangements in variants (Moyer et al., 1980; Dumbell & Archard, 1980; Esposito et al., 1981; Archard et al., 1984; Pickup et al., 1984). It appears that in this strain of FPV each TIR is over 6 kb in length and contains a BamHI site. At one end of the genome the terminal 6 kb BamHI fragment is adjacent to the 11·2 kb fragment sequenced here but at the other end it is adjacent to a 25 kb fragment. By using M13 clones from along the length of the 11·2 kb fragment to probe Southern-blotted FPV DNA, the internal end of the TIR could be mapped onto the sequence by looking for the position at which the 25 kb fragment no longer hybridized to the probe. By using this approach the junction between the TIR and the unique sequence was mapped to between 3715 and 3983 bp from the left end (Fig. 1) of the 11·2 kb sequence. The TIR in FPV is therefore approximately 9·8 kb in size.
Fowlpox virus near-terminal sequence

Fig. 1. Mapping of the end of the TIR onto the 11.2 kb sequence using M13 clones probed onto Southern-blotted FPV DNA on strips of nitrocellulose. Clones were taken from along the length of the 11.2 kb fragment as described in Methods and the point where the 25 kb fragment failed to hybridize was noted. Lane 1, clone MFP 354 which overlaps the 11.2 kb fragment and the terminal 6 kb fragment at one end of the genome and the 25 kb and the terminal 6 kb fragment at the other. Lane 2, clone GF107 which extends from 3715 to 3890. Lane 3, clone GF45 which extends from 3983 to beyond 4200. Lane 4, clone GE23 which extends from 4046 to beyond 4300. From this, the end of the TIR maps somewhere between 3715 and 3983 bases from the left end of the 11.2 kb sequence.

DNA sequencing

The sequence of the 11.2 kb BamHI fragment was obtained from randomly generated subclones of pMH23 inserted into M13 mp10. Ninety-two percent of the fragment was sequenced on both strands and for over 98% the sequence was obtained from at least three different M13 clones. The complete sequence of 11225 nucleotides is shown in Fig. 2. The sequence is 70% A + T. There was one position, base 2133, where the sequence showed an A in two M13 clones and a T in three others. This would result in either an isoleucine or lysine residue should it be translated within the potential open reading frame (ORF) at this position (ORF a in Fig. 3b). There is one region from bases 6612 to 6739 which contains two copies of a 59 bp direct repeat followed by a partial (10 bp) copy. The central 10 bp element results in the introduction of a termination codon into the potential ORF at this position (ORF 7 in Fig. 3b).

Analysis of open reading frames

When translated in all six ORFs, the sequence contains 20 potential ORFs which could code for polypeptides of between 58 and 418 amino acids. These are shown as boxes in Fig. 3(a). At
Table 1. Predicted translation products from minor and major ORFs

<table>
<thead>
<tr>
<th>ORF*</th>
<th>Initiation codon†‡</th>
<th>Termination codon†</th>
<th>Number of amino acids</th>
<th>$M_r$ ($\times 10^{-3}$)</th>
<th>Characterization§</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>416–418</td>
<td>1672–1674</td>
<td>418</td>
<td>48.2</td>
<td>Like VV 42K</td>
</tr>
<tr>
<td>2</td>
<td>2166–2168</td>
<td>2669–271</td>
<td>167</td>
<td>19.8</td>
<td>Family B</td>
</tr>
<tr>
<td>3</td>
<td>c 4055–4052</td>
<td>3608–3606</td>
<td>148</td>
<td>16.4</td>
<td>Like CPV 38K</td>
</tr>
<tr>
<td>4</td>
<td>4170–4172</td>
<td>4592–4594</td>
<td>140</td>
<td>16.5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>c 5138–5136</td>
<td>4821–4819</td>
<td>105</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>c 5974–5972</td>
<td>5519–5517</td>
<td>151</td>
<td>17.9</td>
<td>Family A</td>
</tr>
<tr>
<td>7</td>
<td>c 7906–7904</td>
<td>6674–6672</td>
<td>410</td>
<td>46.8</td>
<td>Like VV 32-5K</td>
</tr>
<tr>
<td>8</td>
<td>8025–8027</td>
<td>8374–8376</td>
<td>116</td>
<td>13.2</td>
<td>Family B</td>
</tr>
<tr>
<td>9</td>
<td>8632–8634</td>
<td>8835–8837</td>
<td>67</td>
<td>7.9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>c 9686–9684</td>
<td>8844–8842</td>
<td>280</td>
<td>33.0</td>
<td>Family A</td>
</tr>
<tr>
<td>11</td>
<td>c 10120–10118</td>
<td>9687–9685</td>
<td>143</td>
<td>16.6</td>
<td>Family B</td>
</tr>
<tr>
<td>12</td>
<td>c 10705–10703</td>
<td>10139–10137</td>
<td>188</td>
<td>22.4</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>c 11225–11223</td>
<td>10706–10704</td>
<td>172</td>
<td>&gt;19.5</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>2084–2086</td>
<td>2302–2304</td>
<td>71</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>2533–2535</td>
<td>2733–2735</td>
<td>66</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>3669–3671</td>
<td>3923–3925</td>
<td>84</td>
<td>9.1</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>6690–6692</td>
<td>6866–6868</td>
<td>58</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>6978–6980</td>
<td>7289–7291</td>
<td>103</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>c 8079–8077</td>
<td>7894–7892</td>
<td>61</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>10203–10205</td>
<td>10475–10477</td>
<td>90</td>
<td>10.4</td>
<td></td>
</tr>
</tbody>
</table>

* 1 to 13 are major ORFs; a to g are minor ORFs.
† Positions of initiating and terminating codons refer to numbering of nucleotides in Fig. 2.
‡ c denotes that the coding strand is complementary to that shown in Fig. 2.
§ Characterization comments refer to homologies as discussed in the text.

several places in the sequence there are gaps without termination codons; potentially these could code for proteins of 50 amino acids or longer but are not translated into proteins because there is no initiating ATG codon available in the correct reading frame. The ORFs have been classified into two groups, major and minor, on the basis of three considerations. Firstly, the codon bias of each ORF was calculated using ANALYSEQ (Staden, 1984b, c), and compared with the known codon bias for all the sequenced VV genes, the FPV TK and DNA polymerase genes and all of the other ORFs sequenced here. Secondly, the spatial arrangement of the ORFs on the fragment was considered. Thirdly, the size of the predicted polypeptides was taken into account. Figure 3(b) shows the structural organization of the major ORFs denoted numerically along the fragment with arrows indicating the 3′ ends. The minor ORFs are denoted by letters. A summary of the features of the ORFs is shown in Table 1.

One of the major ORFs, ORF 9, would on the basis of size alone be assigned to the minor group. However, its codon usage corresponds to that of our test panel of VV and FPV genes, and it does not overlap other ORFs. ORFs in the minor group show poor codon bias, overlap other ORFs considerably, or are contained within ORFs consigned to the major group. This does not preclude their expression in vivo but they are not considered further here. The 13 major ORFs constitute a non-overlapping set along the length of the fragment. At the right end ORFs 9 to 13 are closely spaced with gaps of 9, 3, 19 and 1 bp between adjacent termination and initiation codons. This is similar to the spatial arrangements seen in the central region of the VV genome (Plucienniczak et al., 1985; Weinrich et al., 1986; Niles et al., 1986; Rosel et al., 1986). At the left end of the fragment however, the ORFs are well separated with gaps of 495, 939, 116, 229, 381, 700, 119 and 258 bp. ORFs 1 and 2 are contained entirely within the TIR, and the junction of TIR with the unique sequence is located somewhere within ORF 3. There are no obvious structural features within ORF 3 which indicate the boundary.

Possible translational and transcriptional sequences

The sequence close to the 5′ end of each ORF was compared with favourable contexts for eukaryotic gene translation initiation (Kozak, 1983) and with the consensus sequences derived for VV early and late gene transcriptional promoters. Table 2 shows sequences from −100 to +6
**Fowlpox virus near-terminal sequence**

1031

FOWLPOX VIRUS NEAR-TERM WIN SEQUENCE

YPINNTIFNPNNYISDLSRRRAVOLEGFLICNEVROM

[ORF 7]

MEEGK

[ORF 8]

PRRSASAVLUMRLPCOSIIIIVKVSIFVILSTRPDPKK

[ORF 10]

SMFYENIGCIKIDYNFLCFLYNIIS

[ORF 11]

TLMKLINQIIIRFCSNPFITKINMKELLTDVLTYVIAINOE

[ORF 12]

MDDLKTLIMWRIEIPLTYWKNVLNANMINNIELLMELYK

[ORF 13]

NSSLYEC

[ORF 14]

MIDLLDFLSLOQFSNGDYIKFTKMREIEEKIINICSSLYEC
Fig. 2. Nucleotide sequence of the FPV 11.2 kb BamHI fragment. The sequence is from left to right from the end of the genome. The major ORFs have been translated using the single-letter amino acid code.

(a)

(b) kb

Fig. 3. Identification of ORFs. The nucleotide sequence was translated in all six frames using the ANALYSEQ program of Staden (1982). Boxes in (a) represent all potential ORFs of 50 amino acids or more and vertical lines indicate translational stop codons. (b) Arrangement along the fragment of potentially protein-coding ORFs. Major ORFs are numbered 1 to 13 and minor ORFs lettered a to g. Arrows indicate size and direction of ORFs.

with respect to the A (+1) of the initiator ATG for 12 of the major ORFs (the 5' end of ORF 13 is beyond the end of the fragment). According to Kozak's rules for favourable translational initiation contexts, the nucleotides at positions -3 and +4 are of greatest significance. Of the 12 sequences, 10 have a purine at -3 which is highly favoured for expression. In addition, five of these have a G at +4 which further favours expression.

The 100 bp upstream of each ATG are from 70 to 83% AT with all but two (ORFs 2 and 12) being 76% or higher. There are no immediately obvious similarities to early or late promoter consensus sequences of VV, which are also AT-rich, in particular the TAAAT motif which is well conserved in VV late promoters is absent. There are some conserved motifs within the 12 FPV sequences, for example the sequence TTTATTA appears within 100 bp upstream of the ATG for ORFs 1, 4, 5, 7 and 8 and is very similar to the TTTATTTG at the mRNA start sites of the VV 7.5K early and TK early genes (Weir & Moss, 1987). There are also individual sequences which are similar to specific sequences in orthopoxvirus promoters (Fig. 4a).
<table>
<thead>
<tr>
<th>Reading frame</th>
<th>-100</th>
<th>-3</th>
<th>+4</th>
<th>A + T (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AATATATCTAGTTTATAGCGTATTTATATATAAAAAATATAAAAACAAATCGTATTTATGATGCTACGGCGGTCTTTAACAATTACG</td>
<td>G</td>
<td>GAG</td>
<td>76</td>
</tr>
<tr>
<td>2</td>
<td>CTTAGGGTTTTTTATAGTGCTACAGGTGGTTTCTCGCTATATACAAAAATCGGAAAAATTATAAACAGAAATGATTTTGGCAATTACG</td>
<td>G</td>
<td>CAATG</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>AAGGTATATCTATATGTCTATTCTTATATCAATATAGAAGCTATATATTTTATTTATTATAAAAATAATATTTATTTTATTTTTTTTTTTTTTTT</td>
<td>A</td>
<td>AGATG</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>GTATATAAACATTATACGAAAGAATAGTGTTTTTTATATAGGAGTTTTTACGCGTATATATAATACGGTTTATAAAATAAATAATATATAT</td>
<td>A</td>
<td>AATG</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>TCGAGCTCTGCTGATGAAACTATATTTATTTATTAAATTGAGCATACTTTAAATATAATATTTTATTTTATTTTATTTTTTTTTTTTTTTT</td>
<td>A</td>
<td>AAATG</td>
<td>83</td>
</tr>
</tbody>
</table>
| 6            | GTTTGTATTGTTTTTTAATAATAGAAGACATAATAATAGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
Another interesting feature is the presence of TTTTTNT sequences downstream from several ORFs, since this sequence has been identified as a transcriptional termination signal for VV early genes (Yuen & Moss, 1986; Rohrmann et al., 1986) and it is also found 60 bp downstream from the FPV DNA polymerase and 349 bases downstream from the FPV TK gene (Binns et al., 1987; Boyle et al., 1987). The sequence occurs downstream from ORFs 1, 2, 4, 5 and 7 at distances of 400, 14, 52, 34 and 65 bp beyond the translational termination codons but is not present within the coding sequences of these ORFs. Over the entire fragment, on both strands, the TTTTTNT motif appears a total of 45 times but only four times within the coding sequence of a major ORF (ORF 3, 8 or 9). This is most striking between bases 9686 and 11225 where ORFs 10 to 13 are situated. Not one TTTTTNT sequence appears on the coding strand in this region whereas 10 can be found on the non-coding strand. It is therefore possible that the TTTTTNT sequence, which is also found downstream from a cowpox virus (CPV) gene (Pickup et al., 1986) and two leporipox virus genes (Upton & McFadden, 1986; Upton et al., 1987), is used as a transcriptional termination signal for early genes by most poxviruses.

**Analysis of predicted polypeptides**

All 20 ORFs were translated and the predicted amino acid sequences compared to the NBRF protein library and to all the published poxvirus sequences using the program FASTP. Three of the 13 major ORFs share significant homology with previously published poxvirus sequences. (These are referred to in Table 1.)

Figure 5 shows the match between ORF 1 and the VV 42K early gene sequence (Venkatesan et al., 1982). The homology extends throughout the length of the VV sequence, FPV having an extra 29 amino acids at the amino terminus and 63 at the carboxy terminus. Twenty-four percent of amino acids are identical and a further 42% are substituted by similar amino acids. The FPV 48K sequence lies within the TIR as does the VV 42K sequence, but its direction appears to be reversed. This may be the result of sequence rearrangements in this region, which have been described for several orthopoxviruses. In VV the 42K gene is located next to the growth factor 19K gene within the TIR (Venkatesan et al., 1982). ORF 2 would be a good candidate for being a FPV growth factor gene since it is of a similar size and lies within the TIR, but it shows no homology with the VV growth factor sequence.
Fig. 5. Comparison of the predicted polypeptide from ORF 1 with the VV 42K early protein (Venkatesan et al., 1982) using the program FASTP. Double dots show identical amino acids, single dots show conserved substitutions.

Pickup et al. (1986) have reported the map position and sequence of a CPV gene which causes the production of haemorrhagic pocks on the chorioallantoic membranes of hens' eggs. This sequence, located between 31 and 32 kb from the end of the CPV genome, is related to plasma protein inhibitors of serine proteases. Open reading frame 3, which is located at the junction between the TIR and the unique region, is also related to this super-family of proteins, which includes chicken ovalbumin as well as the protease inhibitors. The predicted protein from ORF 3 is 18.4K in size, considerably smaller than the CPV protein, and matches the carboxy-terminal half of the larger protein. Between 25 and 30% of the residues are identical in the ORF 3 sequence, the CPV 38K sequence, chicken ovalbumin and the serine protease inhibitors, and a further 42 to 50% are conservative changes. Even though the smaller FPV ORF matches the carboxy-terminal half of the 38K protein, examination of the DNA sequences upstream of the initiating ATG reveals conservation of a stretch of sequence from −100 in CPV and −53 in FPV (Fig. 4b). Since ORF 3 lies at the junction of the TIR and the unique sequence it is possible
Fig. 6. Homologous predicted polypeptide sequences within the 11.2 kb FPV fragment. Two sets of ORFs specify proteins that are related to one another. (a) Comparison of the amino acid sequences of ORFs 6 and 10 (family A); (b) Comparison of the amino acid sequences of ORFs 2, 8 and 11 (family B); (c) Hydropathicity across the matching regions of ORF 2 (A), ORF 8 (B) and ORF 11 (C).
**Fig. 7.** Comparison of the predicted polypeptides from ORFs 2, 8 and 11 with regions of rat and chicken hepatic lectins. Positions where three or more sequences are identical are boxed.

that a rearrangement or deletion has occurred resulting in the loss of the amino-terminal portion of the coding region but retention of both upstream sequences and the carboxy-terminal coding region.

The predicted polypeptide from ORF 7 shows homology to the 32.5K host range gene described in VV by Gillard et al. (1986). This gene, which overlaps the HindIII M/K junction in VV, is located approximately 30 kb from the left end of the genome. The FPV ORF 7 product would be considerably larger than the VV protein and the best match covers 142 residues near the amino terminus of the FPV protein sequence. Using SEQHP, several matches of high significance can be found using different alignments of the two sequences against each other. This is caused by short repeated amino acid strings in both the VV and the FPV sequences which cause high scoring in several alignments. Upstream of the translational initiation site of both sequences there is a short conserved AT-rich region (Fig. 4c), shorter than that conserved between the CFV 38K ORF and ORF 3.

In addition to the matches found with previously published poxvirus sequences, two sets of ORFs within the 11.2 kb fragment could encode related proteins (Fig. 6). These are called families A and B in Table 1. Family A contains ORFs 6 and 10 which have a stretch of 144 amino acids where 26% were identical and a further 40% were conserved substitutions. Family B contains ORFs 2, 8 and 11, which show homology to each other over almost the entire length of the smallest ORF (ORF 8) with 21% to 28% of residues being identical between any two of the three and a further 44% to 53% being conserved substitutions. Even more striking is the similarity in hydrophobicity between the latter three proteins over the matching region (Fig. 6c). The region conserved between the three shows a high degree of homology with chicken and rat hepatic lectins (Fig. 7). These proteins are receptors that mediate the endocytosis of glycoproteins, binding either exposed galactose for the rat lectin (Holland et al., 1984) or N-acetylglucosamine for the chicken lectin (Drickamer, 1981). There is evidence that the hepatic lectins are transmembrane proteins with their amino termini located in the cytoplasm and their glycosylated carboxy termini located on the outside of the cell membrane. In this respect they belong to a class of membrane proteins which includes the influenza virus neuraminidase and the invariant gamma chain of class II histocompatibility antigens. The three proteins specified by ORFs 2, 8 and 11 are smaller than either of the published hepatic lectin sequences, matching part of the extracellular region but lacking the carbohydrate-binding site of the lectins. The matches are statistically significant with between 26% and 35% identical amino acids and a further 37% to 46% conserved substitutions in the overlapping regions.

**DISCUSSION**

Examination of the nucleotide sequence of this 11.2 kb DNA fragment of FPV revealed a large number of potential ORFs which could encode polypeptides of between 58 and 418 amino acids. Upstream sequences of 12 of the major ORFs are A + T-rich, and conserved motifs can be seen. In particular the sequence TTTATTA, which is similar to the TTTATTG found at the mRNA start site of two VV early genes (Weir & Moss, 1987), is found upstream of five major ORFs. Specific sequence stretches which were very similar to stretches in orthopoxvirus promoters were also observed. Since transcripts have not been mapped to the ORFs the significance of such sequence homologies is not known. A full study of the functional promoter activity of the sequences upstream of the 12 major ORFs in this fragment will be published separately (F. M. Tomley et al., unpublished data).
In contrast to the orthopoxviruses, very little is known about the genome organization of the avipoxviruses. Two FPV early genes, the TK (Boyle et al., 1987) and DNA polymerase (Binns et al., 1987) genes and a 3.1 kb fragment of FPV DNA which hybridizes to the VV HindIII fragment (Drillien et al., 1987) have been sequenced and found to resemble their VV counterparts closely at the amino acid level. These genes are located within the large central region of the VV genome which has been shown to be broadly conserved between all members of the orthopoxvirus group (Mackett & Archard, 1979). This region is known to contain many closely arrayed genes, several of which are involved in functions such as replication and transcription. Several more genes from the central region which share high levels of homology with their VV equivalents have been identified (M. M. Binns et al., unpublished data).

Unlike the central region, the terminal portions of orthopoxvirus genomes are less well conserved and can readily undergo deletions, duplications and sequence rearrangements which lead to considerable variation even within one species (Moyer et al., 1980; Dumbell & Archard, 1980; Esposito et al., 1981; Archard et al., 1984; Pickup et al., 1984). We have observed similar variations in plaqued isolates of FPV (F. M. Tomley, unpublished observation). In the orthopoxviruses, the terminal regions, of around 30 to 35 kb, are thought to encode genes which determine specific virus characteristics such as host range, virulence, tissue tropism and cytopathogenicity and are thus less likely to be conserved between different members of the poxvirus group. Several genes from the terminal portions of orthopoxvirus DNAs have been sequenced. Since the 11.2 kb FPV DNA fragment lies close to the end of the genome with between 3700 and 3900 bp at the left end contained within the TIR it was of interest to see whether any of the predicted polypeptides showed homology to any previously published poxvirus sequences.

Three of the predicted polypeptides were significantly homologous to other poxvirus sequences, two from VV and one from CPV. All three of the sequences come from the terminal region of their respective genomes. The highest homology is between the FPV ORF 1 product and the VV 42K gene, both of which lie within the TIR. The function of the latter gene is unknown but it is expressed early in infection. Two further matches were seen between the ORF 3 product and the CPV red pock gene and between the ORF 7 product and the VV host range gene. For both of these ORFs the predicted products are different in size to their orthopoxvirus equivalents and their physical locations are closer to the end of the genome. These differences could be the result of sequence deletions and rearrangements in the FVP genome, particularly for ORF 3 which retains homology with upstream non-coding sequences of the CPV red pock gene and with the carboxy terminus of the protein but lacks the intervening amino terminus sequences. A red pock phenotype is not observed for FPV. This strain, HP438, has lost the ability to grow in eggs but its parental strain, HP1, forms white non-haemorrhagic pocks as do other strains which have been examined. It is not known whether ORF 3 and ORF 7 genes are transcribed and translated in FPV infection but if they are it is likely that they are involved in virus-specific functions similar to those attributed to the terminally located genes of the orthopoxviruses.

One final point of interest is that two families of ORFs (A and B) encode polypeptides that are related to one another. The striking conservation of hydropathicity in family B throughout the region of matching would suggest a functional similarity and protein comparisons show that the proteins are homologous to hepatic lectins.

It is clear that the terminal region of FPV DNA contains genes which are related to others found in the orthopoxviruses. However, the exact physical locations of these genes vary between the two groups and the level of conservation is considerably lower than that between genes within central regions of the genomes.

We thank Stephen Laidlaw and Ian Foulds for excellent technical assistance.

REFERENCES


(Received 30 October 1987)