Tandem Repeated Sequences within the Terminal Region of the Fowlpox Virus Genome

By Joan I. A. Campbell,* Matthew M. Binns, Fiona M. Tomley and Michael E. G. Boursnell
AFRC Institute for Animal Health, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA, U.K.

(Accepted 10 October 1988)

SUMMARY
A 6·2 kb BamHI terminal fragment from fowlpox virus has been cloned and the nucleotide sequence was determined. The fragment was cloned by S1 digestion of viral DNA and therefore does not contain the covalently closed terminal loop. The cloned sequences comprise a short (230 bp) unique region at the terminal end, which is adjacent to a 3·87 bp long, AT-rich region consisting of sets of short tandemly repeated units, 32 and 56 bp long. The remainder of the fragment is composed of a 2·18 kb unique region containing three major open reading frames. The amino acid sequence encoded by one of these has some similarity to that of platelet-derived growth factor.

INTRODUCTION
Several studies involving the use of vaccinia virus as an expression vector have formed the basis for the development of live vaccine vectors (Mackett & Smith, 1986). As a result other poxviruses, such as fowlpox virus (FPV) (Boyle & Coupar, 1988), are being studied as potential vectors for veterinary vaccines.

The production of new FPV vectors requires detailed genomic and biological characterization of the virus for the identification and manipulation of non-essential regions where foreign genes can be inserted. The best studied poxvirus, the orthopoxvirus vaccinia, has a linear 185 kb dsDNA molecule with inverted terminal repeats (TIRs) that are more than 10 kb in length (Wittek et al., 1978a). TIRs have also been found in all the other poxviruses so far studied. In orthopoxviruses, the TIR possesses a cross-linked hairpin loop at its extremity (Baroudy et al., 1982), followed by two sets of tandem 30 to 70 bp repeats separated by a short unique region, and then regions coding for polypeptides (Wittek et al., 1980; Pickup et al., 1983).

Shope fibroma virus (SFV), a member of the leporipoxvirus group, differs from this orthopoxvirus pattern in that its TIR lacks the tandem repeats with the result that, except for its terminal 400 bp, it could completely code for proteins (Upton et al., 1987). However, experiments with cloned vaccinia virus and SFV terminal hairpins suggest that for both poxvirus groups, the TIRs joined head-to-head represent an intermediate state in viral DNA replication, since such head-to-head constructs in superinfecting plasmids could be resolved into linear structures by trans-acting virus-induced proteins (Delange et al., 1986).

Evidence from studies on the genetic organization of the orthopoxviruses (Mackett & Archard, 1979; Plucieniczak et al., 1985) and fowlpox virus (Drillien et al., 1987; Binns et al., 1988) suggests that whereas the unique internal regions of these viruses manifest a high conservation in gene content and order, the terminal regions show a high degree of variation. It has been postulated that this is due to a requirement of the central genes for viral replication whereas those located at the ends are involved in pathogenicity, thus allowing for more evolutionary flexibility in order to suit a particular host. Heterogeneity in viral isolates during routine virus passaging arising from rearrangement in terminal DNA has been reported (Wittek et al., 1978b). Similar size variations have been observed in plaque-purified isolates of FPV.
(Tomley et al., 1988). Another feature of the terminal region of vaccinia virus is that it contains many non-essential genes. For example, a 9 kb sequence of the vaccinia virus genome could be deleted from near-terminal DNA sequences to leave a virus still capable of growth in tissue culture (Moss et al., 1981).

We are studying various aspects of FPV with a view to developing it as a vaccine vector specific for poultry. Part of this work involves the elucidation of the viral molecular structure. The terminal fragment is of particular interest both for its relationship to the other characterized poxvirus terminal fragments and for its suitability for insertion of foreign DNA.

METHODS

Plasmid DNA preparation. Plasmid DNA was prepared as described by Maniatis et al. (1982).

Identification of the terminal fragment. The FPV terminal fragment was identified using the snap-back method of Archard & Mackett (1979). A BamHI restriction digest of 10 µg FPV DNA was extracted once with phenol, ether-extracted and the DNA was precipitated with ethanol. The DNA pellet was dried and resuspended in 20 µl TE (10 mM-Tris–HCl pH 8.0, 1 mM-EDTA). One-half of the DNA solution was then denatured by the addition of 40 µl of 95% deionized formamide, 50 mM-Tris–HCl pH 7.5 and incubated at 60 °C for 10 min. The sample tube was then rapidly cooled on wet ice containing NaCl for 10 min before analysis of the DNA by agarose gel electrophoresis.

Cloning of the end fragment of FPV DNA. The cloning of the BamHI-generated end fragment of FPV DNA was carried out as follows. The terminal cross-links were digested with S1 nuclease (Boehringer Mannheim) in 250 µl of 30 mM-sodium acetate pH 4.5, 0.3 M-NaCl, 1 mM-ZnSO4, 5% (v/v) glycerol at 37 °C for 20 min. The DNA was end-terminated with both Escherichia coli DNA polymerase I (Klenow fragment) and T4 DNA polymerase, digested with BamHI and ligated into plasmid pBG819 (Spratt et al., 1986) cut with BamHI and Smal. Screening for recombinants that contained the end fragment was carried out by standard procedures as described by Maniatis et al. (1982) and Holmes & Quigley (1981).

Digestion of DNA with restriction enzymes. For restriction mapping the enzymes were used as recommended by the manufacturers. Partial digestions were carried out by a modification of the procedure of Smith & Birnstiel (1976). For the latter, the reaction mixtures contained approximately 5 µg (10⁶ to 3 × 10⁶ c.p.m., Cerenkov) of [3²P]dNTPs and E. coli DNA polymerase I (Klenow fragment) and ϕX174 HaeIII marker fragments were end-labelled with T4 polynucleotide kinase (Maniatis et al., 1982).

Nucleotide sequence analysis. Random fragments of the FPV end fragment clone were generated by sonication and end-repaired with T4 DNA polymerase and Klenow fragment. The fragments were then cloned into Smal-cut M13mp10 (Amersham). DNA sequencing was by the dideoxynucleotide chain termination method (Sanger et al., 1977) using [α-3²P]dATP and buffer gradient gels (Biggin et al., 1983). Sequence data were read into a BBC microcomputer using a sonic digitizer (Graf/Bar, Science Accessories Corporation) and analysed on a VAX 11/750 using the programs of Staden (1982, 1984). Protein comparisons with data in the National Biochemical Research Foundation (NBRF) Protein Identification Resource (George et al., 1986) were carried out using the program FASTP (Lipman & Pearson, 1985). The diagram of the open-reading frames (ORFs) in Fig. 6 was drawn with a computer program (SIXFRAMES) written for the VAX 11/750.

RESULTS AND DISCUSSION

FPV terminal fragment

FPV DNA treated or untreated with S1 exonuclease was digested with BamHI and one-half of either digest was subjected to the snap-back treatment as outlined in Methods. The presumed terminal fragment renatured rapidly only in the native FPV DNA and was detected as a major band of approximately 6.3 kb as shown in Fig. 1, lane 4. There is only one BamHI-derived snap-back fragment, because the entire FPV TIR extends 3-7 to 3.9 kb internal to this site (Tomley et al., 1988). The deduced size (6.3 kb) of the snap-back fragment corresponds fairly well to a
Terminal region of fowlpox virus. DNA

Fig. 1. Identification of the FPV terminal fragment. SI-treated and native FPV DNAs were digested with BamHI and one-half of each digest was subjected to the snap-back treatment. Lane 1, SI-treated FPV DNA BamHI digest; lane 2, snap-back-treated DNA from lane 1; lane 3, BamHI digest of native FPV DNA; lane 4, snap-back-treated DNA from lane 3. The terminal fragment was observed as the major band in the snap-back lane of native DNA and has been indicated with the arrow.

BamHI restriction fragment as shown in lane 3 of Fig. 1. The snap-back fragment appears to run slightly faster in lane 4 than the corresponding fragment in lane 3; this may be due to different buffer conditions.

DNA sequence of the FPV terminal fragment

The FPV terminal BamHI fragment was cloned in plasmid pBGS19 as described in Methods and recombinant plasmid pB3 was isolated. Comparison of restriction digests of the fragment cloned in pB3 with the terminal fragment from FPV virion DNA showed that the SI treatment had removed approximately 45 bp of sequence from the terminus. Sequence data from this fragment were obtained as described in Methods and the sequences are presented in Fig. 2. Over 92% of the non-repeated sequence was determined on both strands and all the sequence has been determined from at least two separate M13 clones. Upon computer analysis of the nucleotide sequence determined from this clone, two major sets of contiguous sequences were obtained. The first consisted of 230 bp of unique sequence followed by 180 bp of sequence containing many tandem repeats. The second, larger, contiguous block started with similar repeated units, followed by 2-18 kb of unique sequence. Restriction analysis of the clone pB3 showed that these repeated regions were approximately 3.87 kb in length, and the numbering of the sequence downstream of the repeats in Fig. 2 is based on this estimate. Although during the shotgun sequencing many individual M13 clones containing sequences from the repeated regions were
Fig. 2. Nucleotide sequence of the cloned 6.2 kb BamHI terminal fragment. The sequence starts from the extreme terminus of FPV DNA after St I treatment and extends to the BamHI site within the TIR. The major ORFs have been translated using single-letter amino acid code. Only the repeated sequence determined in contiguous blocks with unique sequences are presented.
Terminal region of fowlpox virus. DNA

Fig. 3. *Ava*I, *Sst*I, *Xho*I, *Asu*II and *Hind*III complete digestion of recombinant plasmid pB3 (lanes 1, 2, 4, 6 and 8). The fragments were resolved in 2.4% low gelling temperature agarose gels. The sizes of the fragments are indicated on the figure in bp and were determined by comparison with φX174 *Hae*III markers in lanes 3, 5, 7 and 9.

(a) *Hind*III

L (56 bp)  

<table>
<thead>
<tr>
<th>L</th>
<th>S</th>
<th>S</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>603</td>
<td>310</td>
<td>234</td>
<td>194</td>
</tr>
<tr>
<td>310</td>
<td>234</td>
<td>194</td>
<td>153</td>
</tr>
<tr>
<td>234</td>
<td>194</td>
<td>153</td>
<td>118</td>
</tr>
<tr>
<td>194</td>
<td>153</td>
<td>118</td>
<td>72</td>
</tr>
</tbody>
</table>

(b) *Asu*II/

L (56 bp)  

<table>
<thead>
<tr>
<th>L</th>
<th>S</th>
<th>S</th>
<th>S</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>220</td>
<td>153</td>
<td>120</td>
<td>88</td>
<td>72</td>
</tr>
<tr>
<td>153</td>
<td>120</td>
<td>88</td>
<td>603</td>
<td>310</td>
</tr>
<tr>
<td>120</td>
<td>88</td>
<td>603</td>
<td>310</td>
<td>234</td>
</tr>
<tr>
<td>88</td>
<td>603</td>
<td>310</td>
<td>234</td>
<td>194</td>
</tr>
</tbody>
</table>

| *Asu*II/*Sst*I/*Xho*I cleavage |
| *Asu*II/*Hind*III cleavage |

Fig. 4. (a) DNA sequence of the two basic repeat units, long and short. The restriction enzyme sites have been underlined and the sequence variations observed in some repeats are shown underneath the main sequences. (b) Diagram showing how the observed fragment sizes (bp) are obtained by digestion with the restriction enzymes.

obtained, attempts to order these sequences using the computer proved unsuccessful because very similar, but not identical, repeats were matched incorrectly.

To resolve this problem the following approach was adopted: individual repeated units generated by digestion with *Asu*II and *Hind*III were cloned and sequenced, and then the order of these different units was determined by performing partial digestion of end-labelled DNA. An examination of the repeated sequences revealed that although enzymes *Asu*II, *Sst*I and *Xho*I cleave frequently within the repeated sequence (every 32 or 56 bp), *Hind*III and *Asu*II cut less frequently (approximately every 87 bp). Thus, fragments corresponding to sizes of 32, 56 and 90 bp were resolved in 2.4% agarose gels with *Asu*II, *Sst*I and *Xho*I (Fig. 3, lanes 1, 2 and 4) and those of 56, 90, 120, 153 and 220 bp were observed with *Asu*II and *Hind*III (Fig. 3, lanes 6 and 8).
AsulI digestion of pB3Q. Lanes 1 and to the right of lane 7, \( \lambda \) HindIII and \( \phi X174 \) HaeIII marker fragments with sizes in kb. Lanes 2 and 3, AsulI partial digest of pB3Q linearized with EcoRI and end-labelled, the reactions being terminated at 10 min and 20 min respectively. Lanes 5 and 6, AsulI partial digest of pB3Q linearized with XbaI and end-labelled, the reactions being terminated at 10 min and 20 min respectively. Lanes 4 and 7, AsulI complete digests of the EcoRI- and XbaI-linearized plasmids respectively.

(The extra 190 bp fragment observed in lane 6 is due to an additional AsulI site occurring within flanking unique sequences.) These fragments were purified from low gelling temperature agarose, cloned into M13 vectors and sequenced. This revealed that they are composed entirely of two types of sequence motif, a long (L) one of 56 bp and a short (S) one of 32 bp. These two types of repeat are shown in Fig. 4(a). It can be seen that the L repeat contains the sequences of the S repeat as well as having an extra 24 bases at the left-hand end. These extra 24 bases however do not constitute an autonomous repeat unit as they never occur except in conjunction with the right-hand 32 bases. Some sequence variation occurs within these repeat units, as shown beneath the main sequences in Fig. 4(a). Also shown are the positions of the HindIII site in the L repeat and the AvaI site in the L and S repeats. The derivation of the restriction fragments seen in Fig. 3 is shown in Fig. 4(b). The 90 bp fragment seen in Fig. 3 lanes 1, 2 and 4 is derived from an L fragment lacking the AvaI/SstI/XhoI sites adjacent to an S fragment. A pair of L fragments separated by four S fragments was not seen.

The tandem repeats in vaccinia virus are separated into two regions by a short unique sequence (Baroudy et al., 1982). No evidence of such a separating unique sequence was found in FPV; all the fragments generated by digestion with the enzymes that cut within the repeated
region contained only the two repeated motifs L and S. In addition no anomalous bands, which might represent a separating unique sequence, can be seen after digestion with enzymes that cut many times within the repeated sequences (Fig. 3). Comparison of the repeating sequences from FPV with those of the orthopoxviruses vaccinia virus and cowpox virus shows that although there is no large scale similarity between the repeat units of the different genera, there is a 12 bp sequence (AAAAACTTTTTTA) conserved between all three viruses.

**Arrangement of repeat sets**

The arrangement of the motifs within the repeated region was determined by mapping the internal *AsuII* sites using a modification of the Smith & Birnstiel (1976) partial digestion procedure. In order to reduce interference from the larger unique region, the DNA between the *BglII* and *BamHI* sites in the plasmid pB3 was deleted (see Fig. 6a). The plasmid thus created was designated pB3Q.

![Diagram](image-url)
152

J. I. A. CAMPBELL AND OTHERS

PDGF-A

EF1

117 138

172

WLPVCKRTGCNTSSVK

WLPSCVCKRTMTDFTK

Fig. 7. Alignment of the regions of amino acid similarities between platelet-derived growth factor subunit A and open reading frame EF1. The sequences are numbered from their respective N termini and conserved residues are boxed.

The cloned FPV terminal fragment in pB3Q was labelled from either end separately by first digesting with XbaI (which cuts within the vector sequence on the right-hand side of the FPV DNA insert), and Klenow filling-in. The end-labelled DNA fragments were then partially digested with AsuII. This procedure created ladders of end-labelled fragments extending in both directions from the labelled XbaI site. Because the ladder containing the pBGS19 vector sequences only started at 4 kb, the arrangement of approximately the first 3-4 kb of the repeats could be deduced correctly without interference from the higher ladder. This higher ladder can be seen in Fig. 5, lanes 5 and 6. Using EcoRI (which cuts to the left of the FPV sequences) as the labelled site, the arrangement of the repeats coming in from the other direction could be deduced. Repeated units seen in the labelled EcoRI ladder (lane 2) can be seen reversed in the XbaI ladder. The two patterns in fact overlap substantially and hence the pattern of repeats for the entire repeated region can be determined. The deduced pattern is shown diagrammatically in Fig. 6 (a).

Open reading frames

The DNA sequence of the terminal fragment was translated in all six reading frames and potential ORFs were identified. Fig. 6 (b) shows diagrammatically the positions of stop codons and ORFs. A small ORF (representing 51 amino acids) was predicted within the smaller unique region while five possible ORFs encoding more than 67 amino acids were predicted in the long non-repeated region. Translation of the tandem repeats does not produce any significant ORFs. The ORFs are designated EF1, EF2, EF3, EFa, EFb and EFc, and are presented as boxes in Fig. 6 (b). The corresponding amino acid sequences of the putative polypeptides are presented alongside the coding DNA sequence in Fig. 2. The ORFs have been classified into major and minor groups depending on their sizes and how they conform to the codon bias as compiled from known FPV genes in the near-terminal 11.2 kb fragment (Tomley et al., 1988). Thus, ORFs EF1, EF2 and EF3 (26K, 11.5K and 13.9K products respectively) and the terminal 51 amino acid ORF (EFa), were highly typical while ORFs EFb (8.6K) and EFc (8.8K) were not. Interestingly the predicted FPV codon bias of EF1 continues in another reading frame at the 3' end of EF1. To determine whether this shift of reading frame was due to a cloning artefact, DNA fragments covering this region were cloned directly from viral DNA into M13mp10 and sequenced. All the clones had the same sequence at this point. It is possible that this region of the genome is non-essential and that a mutation causing the frameshift (or removing an ATG) has occurred. Another possibility is that a ribosomal frameshift occurs at this point to allow the product of the downstream ORF to be fused to that of EF1. However no frameshifting signals similar to those detected in other systems have been detected (Jacks & Varmus, 1983; Brierley et al., 1987).

Comparison of the predicted ORFs in this fragment to those in the Protein Identification Resource and to known poxvirus genes indicated that none of these FPV ORFs are highly homologous to known protein sequences. However, the search of the NBRF database did reveal that a string of 23 amino acids towards the C terminus of EF1 had 50% similarity to a conserved region in subunit A of human platelet-derived growth factor, a known mitogen (Waterfield et al., 1983; Bersholtz et al., 1986) (Fig. 7). The former group reported finding a related gene in a retrovirus. Growth factor-like genes have been reported in vaccinia virus (Brown et al., 1985) and SFV (Chang et al., 1986) but these are more related to epidermal growth factor. FPV infection in vitro has been reported to induce marked but self limiting hyperplasia (Cheevers & Randall, 1968), which could be caused by a growth factor-like gene.
The FPV terminal fragment studied in this work shares the general pattern of nucleotide sequence arrangement with the characterized orthopoxviruses, i.e. terminal cross-linking, high AT content and internal redundancy. The DNA sequences of the repeated subunits are different (although there is a 12 bp conserved region) and there is considerable evidence presented here that the repeats are not partitioned by a central unique region, though the entire sequence of the repeated region was not determined in one contiguous block. Another difference from orthopoxviruses is the presence of a possible protein-coding sequence in the short unique region distal to the repeats (Baroudy et al., 1982; Parsons & Pickup, 1987). The larger non-repeated region of the FPV terminal fragment codes for a putative protein with a region of homology to platelet-derived growth factor. Mutation by insertion of the vaccinia virus growth factor gene has been found to reduce the virulence of the virus (Buller et al., 1988). Recombination studies aimed at inserting foreign DNA into the FPV growth factor-like gene are in progress to determine what significance, if any, this may have on infection by FPV.

We thank Mark Elsdon for excellent technical assistance, Stuart Hodgson for preparing the figures and Helen Vickery Jr for her patience in typing the manuscript.

REFERENCES


(Received 4 July 1988)