Nucleotide sequence analysis of a 10.5 kbp \textit{HindIII} fragment of fowlpox virus: relatedness to the central portion of the vaccinia virus \textit{HindIII} D region

James Tartaglia,† Joseph Winslow,† Scott Goebel,† Gerard P. Johnson,† Jill Taylor† and Enzo Paoletti*†

Wadsworth Center for Laboratories and Research, New York State Department of Health, P.O. Box 509, Empire State Plaza, Albany, New York 12201-0509, U.S.A.

The nucleotide sequence of a 10465 bp \textit{HindIII} genomic fragment from fowlpox virus (FPV) is presented. Analysis of the nucleotide sequence revealed 10 potential major open reading frames (ORFs). Five of these ORFs are predicted to encode polypeptides with significant homology to hypothetical polypeptides derived from nucleotide sequence analysis of the vaccinia virus (VV) \textit{HindIII} D region. Interestingly, these homologous ORFs do not occur in the same tandem arrangement in the FPV genome as they do in the VV genome. These results are discussed.

Introduction

Fowlpox virus (FPV) is the prototype of the avipoxvirus genus belonging to the family \textit{Poxviridae}. Like vaccinia virus (VV), the most well studied poxvirus, FPV contains a large, linear, double-stranded DNA genome. The size of the FPV genome is approximately 280 kbp (Müller et al., 1978; C. Edbauer, J. Tartaglia & E. Paoletti, unpublished data). This affords the potential for encoding between 200 to 300 polypeptides. Unlike VV, little is known about the FPV genomic organization and the mechanisms governing viral gene expression.

To date, no restriction maps of the FPV genome have been reported. Recently, however, the nucleotide sequence of a 3-1 kbp fragment (Drillien et al., 1987), a 6-1 kbp fragment (Binns et al., 1988) and an 11-2 kbp \textit{BamHI} FPV genomic fragment (Tomley et al., 1988) were reported and certain predicted open reading frames (ORFs) were found to be homologous to VV ORFs. Analysis of the 3-1 kbp and 6-1 kbp fragments revealed extensive conservation between the FPV genomic segment and the VV \textit{HindIII} J region at the amino acid sequence level and in the organization of the ORFs (Drillien et al., 1987; Binns et al., 1988). An FPV homologue of the VV thymidine kinase (TK) gene was, however, lacking from the \textit{HindIII} J equivalent region. A TK gene was identified in another region of the FPV genome (Binns et al., 1988; Boyle et al., 1987).

Methods

\textit{Cells and virus.} FPV (strain FP-1) was propagated on primary chick embryo fibroblasts (CEF) derived from 10 to 11 day old embryonated eggs obtained from Spafas using conditions described previously (Taylor et al., 1988a, b). Virus was purified from host cell contaminants by sucrose gradient centrifugation using the method described by Joklik (1962).

\textit{Purification of DNA.} Genomic DNA was extracted from purified FPV by addition of an equal volume of lysyl buffer containing 54% sucrose, 2% Sarkosyl (NL97), 200 mm-NaCl, 100 mm-2-mercaptopethyl-
anol, 20 mM-Tris–HCl pH 8.0 and 2 mM-EDTA and incubated at 37 °C for 3 h. This was followed by multiple extractions with phenol/chloroform/isoamyl alcohol in the ratio of 25:24:1. DNA was recovered by ethanol precipitation. Plasmid DNA was isolated using standard procedures (Maniatis et al., 1982).

Cloning. The 10.5 kbp HindIII fragment was isolated from an agarose gel using GeneClean (Bio101) according to the manufacturer’s specifications. The isolated HindIII fragment was ligated to HindIII-digested pBS-SK (Stratagene). The ligation mixture was used to transform competent Escherichia coli cells (strain JM83) by standard protocols (Maniatis et al., 1982).

DNA sequence analysis. Sequence data for both strands of the 10.5 kbp FPV genomic HindIII fragment were obtained using the modified T7 enzyme Sequenase and standard protocols (U.S. Biochemicals). The reactions were performed using double-stranded plasmid DNA templates which were denatured by treatment with 0.4 M-NaOH. The M13 forward and reverse primers were used to obtain initial sequence data. Subsequent sequence reactions utilized custom synthetic primers (17-18-mer) that were prepared with the Biosearch 8700 or Applied Biosystems 380B oligonucleotide synthesizers.

Sequences were entered into a Dell system 310 computer and the data were analysed using the IBI Pustell sequence analysis programs (International Biotechnologies). Homology searches were done with the Biosearch 8700 or Applied Biosystems 380B oligonucleotide synthesizers.

Results

Nucleotide sequence analysis of the 10.5 kbp FPV HindIII fragment

The 10.5 kbp FPV HindIII subgenomic fragment was identified when mapping the gene encoding a highly expressed 26K virus-induced polypeptide. Detailed description of the mapping data and expression of the FPV 26K gene product and analysis of the 26K promoter will be presented in a subsequent communication. Preliminary restriction mapping of the FPV genome has positioned this 10.5 kbp HindIII fragment approximately 70 kbp from one end of the FPV genome (C. Edbauer, J. Tartaglia & E. Paoletti, unpublished observation). Nucleotide sequence of this 10465 bp HindIII fragment was obtained from plasmid DNA containing the whole FPV subgenomic fragment. The entire HindIII fragment was sequenced in both directions with custom-synthesized oligonucleotides (see Methods). Fig. 1 shows the complete sequence. This fragment is 69-5% A : T-rich. Analysis of the nucleotide sequence in all six reading frames revealed 10 potential major ORFs (Fig. 1; Table 1). Eight of these ORFs are completely contained within the limited number of FPV genomic regions analysed (Drillien et al., 1987; Binns et al., 1988; Tomley et al., 1988).

Analysis of the amino acid sequences

All 10 major ORFs identified by translation of the nucleotide sequence in all six potential reading frames were compared to the SWISS PROT protein database using the computer program FASTP (Lipman & Pear-

<table>
<thead>
<tr>
<th>ORF</th>
<th>Translation</th>
<th>Amino acids</th>
<th>Mr × 10-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP</td>
<td>start</td>
<td>stop</td>
<td></td>
</tr>
<tr>
<td>FPD4</td>
<td>530</td>
<td>1183</td>
<td>218</td>
</tr>
<tr>
<td>FP14</td>
<td>1282</td>
<td>1668</td>
<td>129</td>
</tr>
<tr>
<td>FP18</td>
<td>2119</td>
<td>2610</td>
<td>164</td>
</tr>
<tr>
<td>FP26</td>
<td>3317</td>
<td>2661</td>
<td>219</td>
</tr>
<tr>
<td>FPD5</td>
<td>4258</td>
<td>6630</td>
<td>791</td>
</tr>
<tr>
<td>FPD6</td>
<td>6614</td>
<td>8428</td>
<td>605</td>
</tr>
<tr>
<td>FPD7</td>
<td>8504</td>
<td>8986</td>
<td>161</td>
</tr>
<tr>
<td>FPD9</td>
<td>10230</td>
<td>⍺</td>
<td>78</td>
</tr>
</tbody>
</table>

* Designates incomplete ORF at left end of sequence. Actual start of translation unknown.
† ND, Not determined.
‡ Designates incomplete ORF at right end of sequence. Precise location of translation termination is unknown.

2(a) and Table 1. The putative direction of transcription is indicated in Fig. 2(a).

The 10 major ORFs exist in a tightly clustered arrangement (Fig. 2a). The 3' end of the ORF FPD5 and the 5' end of ORF FPD6 overlap by five amino acids. There are two regions, however, where significant gaps occur between ORFs. Analysis of these regions revealed no significant ORFs and the occurrence of many termination codons in all six potential reading frames. ORFs FP14 and FP18 are separated by 448 bp and ORFs FP26 and FPD5 have a 941 bp gap between them (Fig. 1 and 2a). Tomley et al. (1988) have also observed large stretches of non-coding sequences within the 11.2 kbp near-terminal BamHI fragment.

Within the 10.5 kbp HindIII fragment, there exists a dominant direction of expression. All of the major ORFs, except ORF FP26, are oriented in a left to right direction with respect to the presented sequence (Fig. 2a). The arrangement of tightly clustered ORFs oriented in a dominant direction is reminiscent of the organization in many VV genomic regions (Niles et al., 1986: Rosel et al., 1986; Schmitt & Stunnenberg, 1988; Tamin et al., 1988; Kotwal & Moss, 1988, 1989; Boursnell et al., 1988) and in the limited number of FPV genomic regions analysed (Drillien et al., 1987; Binns et al., 1988; Tomley et al., 1988).

Analysis of the amino acid sequences

All 10 major ORFs identified by translation of the nucleotide sequence in all six potential reading frames were compared to the SWISS PROT protein database using the computer program FASTP (Lipman & Pear-
Fig. 1. Nucleotide sequence of the FPV 10.5 kbp fragment. The major ORFs have been translated using the single-letter amino acid code. The amino acids underlined in ORFs FP26 and FPD5 denote ATP/GTP motif A binding sites (Walker et al., 1982).
Table 2. Homology between FPV ORFs and ORFs within VV HindIII D* region

<table>
<thead>
<tr>
<th>FPV ORF</th>
<th>VV ORF</th>
<th>Alignment score (initial/optimized) †</th>
<th>Homology (%) ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPD4</td>
<td>D4</td>
<td>743/752</td>
<td>53.7/216</td>
</tr>
<tr>
<td>FPD5</td>
<td>D5</td>
<td>995/2427</td>
<td>54.0/789</td>
</tr>
<tr>
<td>FPD6</td>
<td>D6</td>
<td>1023/2426</td>
<td>74.1/607</td>
</tr>
<tr>
<td>FPD7</td>
<td>D7</td>
<td>530/537</td>
<td>59.7/160</td>
</tr>
<tr>
<td>FPD9</td>
<td>D9</td>
<td>105/149</td>
<td>50.8/59</td>
</tr>
</tbody>
</table>

* Sequence of VV HindIII D region reported previously (Niles et al., 1986).
† Sequences were aligned using the FASTP algorithm of Lipman & Pearson (1985) using the PAM250 weighting matrix (Schwartz & Dayhoff, 1978). Scores are presented before (initial) and following (optimized) gap insertion for alignment optimization.
‡ Percentage homology is expressed as % identity/no. of amino acid overlap.

Discussion

Analysis of the nucleotide sequence of a 10465 bp HindIII fragment from the FPV genome has identified 10 major ORFs. Eight of these ORFs are completely contained within this sequence (Fig. 1 and 2a). The ORFs are situated in a tightly clustered configuration with a dominant direction of orientation similar to previously reported nucleotide sequences from various VV and FPV genomic regions (Drillien et al., 1987; Binns et al., 1988; Tomley et al., 1988; Niles et al., 1986; Rosel et al., 1986; Schmitt & Stunnenberg, 1988; Tamin et al., 1988; Kotwal & Moss, 1988, 1989; Boursnell et al., 1988). Only two regions within the 10465 bp sequence have any significant stretches that lack potential ORFs. These gaps exist between ORFs FP14 and FP18 (447 bp) and ORFs FP26 and FP5 (931 bp). Additionally, only ORF FP26 is oriented in a right to left direction with respect to the reported sequence.

Amino acid analysis has demonstrated that five of the ORFs present in this fragment have a high degree of homology to hypothetical VV ORFs previously identified by sequence analysis of the HindIII D region (Niles et al., 1986). FPV ORFs FPD4, FPD5, FPD6, FPD7 and FPD9 were found to be homologous to VV ORFs D4, D5, D6, D7 and D9, respectively (Table 2). The other five FPV ORFs (FP left, FP14, FP18, FP31 and FP26) did not demonstrate significant homology to any amino acid sequence within the SWISS PROT protein database. FPV ORFs homologous to the VV ORFs D3 and D8 were not identified within this sequence (Fig. 2). It is not known whether FPV homologues of the D3 and D8 ORFs exist elsewhere in the FPV genome or whether FPV lacks these homologues. Since the VV D8 ORF has been found to encode a non-essential transmembrane protein that has a high degree of homology with the mammalian enzyme carbonic anhydrase (Niles & Seto, 1988), it would not be surprising if it were absent from
the FPV genome. No information is presently available on the expression and function of the VV D3 ORF.

At the present time, evidence exists that only the FP26 ORF is expressed by FPV. The FPV 10-5 kbp HindIII fragment, reported in this communication, was identified by mapping the gene encoding the FPV 26K polypeptide using a cDNA probe prepared with polyclone mRNA extracted from FPV-infected cells at a time when the 26K polypeptide was the predominant translation product (J. Tartaglia & E. Paoletti, unpublished results). It is not known whether any of the other genes are expressed in FPV-infected cells. However, it is known that the VV genes represented by ORFs D4, D5, D6, D7 and D9 are all expressed in VV-infected cells. Results have been reported from transcriptional analysis of the HindIII D region that demonstrate the expression of these genes (Lee-Chen & Niles, 1988a, b; Lee-Chen et al., 1988). VV genes D4, D5, D7 and D9 all appear to be expressed early after infection whereas D6 is transcribed at late times post-infection (Lee-Chen & Niles, 1988a, b; Lee-Chen et al., 1988).

The extremely high levels of sequence conservation at the amino acid level between the FPV ORFs, FPD4, FPD5, FPD6, FPD7, FPD9 and the corresponding VV ORFs D4, D5, D6, D7 and D9 suggest a biological significance for the encoded proteins in the replication of poxviruses. It will be of interest to determine whether the FPV ORFs are essential for viral replication. The role they play in the replicative scheme, however, remains unknown. Speculation can only be made pertaining to the function of the gene product encoded by ORF FPD5 based upon results with its homologue in the VV system. It was found that ts17 maps to ORF D5 (Roseman & Hruby, 1987; Evans & Traktman, 1987). Further analysis of this VV mutant by temperature-shift experiments has suggested that the ORF D5 gene product has a role in DNA replication, in particular, elongation of DNA (Evans & Traktman, 1987). In support of such a function for the FPD5 and VV D5 homologue is the occurrence of an ATP/GTP binding site (Walker et al., 1982) at amino acids 507 to 512 (Fig. 1) and 502 to 509 (Niles et al., 1986), respectively. Such consensus patterns are associated with ATP-utilizing proteins such as helicases (McGeoch et al., 1988), a function consistent with an involvement in the elongation of DNA during replication.

It will be of interest to determine whether the FPV homologues are under the same temporal regulation as their VV counterparts. Sequence extending to 100 bp upstream from the initiation codon for the FP ORFs exhibits 72% to 84% A:T content, a characteristic of defined VV promoter elements (reviewed in Tartaglia et al., 1990). No other striking similarities were detected between these upstream regions and those of other poxviruses. Definition of the upstream regulatory elements awaits the mapping of the transcription start site(s) and a directed genetic approach to determine which sequences are involved in regulating gene expression.

Other sequences known to affect VV gene expression were noted in the context of the FPV ORFs reported here. For instance, the sequence motif TAAATG has been found to be essential for promoter activity of certain VV transcription units (Hänggi et al., 1986). The ATG in this sequence acts as the initiation codon of a particular ORF. Both the FPD5 and FPD6 ORFs contain this sequence at the beginning of the ORF (Fig. 1). Similarly, the VV D5 and D6 are preceded by this sequence and transcription is initiated at the TAAAT sequence (Lee-Chen & Niles, 1988a, b; Lee-Chen et al., 1988).

The T_s NT sequence motif has been shown to signal the termination of early transcription 50 to 75 nucleotides downstream from its occurrence in VV-infected cells (Yuen & Moss, 1986, 1987; Rohrmann et al., 1986). Transcription termination by this sequence, however, is not absolute as sequence environment affects the overall efficiency of termination (Yuen & Moss, 1987). A total of 43 T_s NT elements occur in the 10-5 bp sequence when both orientations are evaluated. Significantly, only one occurs in the orientation of expression of the nine hypothetical ORFs. Most of these occur downstream of ORFs and may provide transcription stop signals for the putative transcription units. Three T_s NT elements appear within ORF FPD6 (Fig. 1). This may not affect transcription of this gene, since in VV the FPD6 homologue is expressed at late times post-infection (Lee-Chen & Niles, 1988a, b). A T_s NT element also occurs within the FPD4 coding sequence (Fig. 1). By analogy with the VV D4 ORF, the ORF FPD4 would be expressed at early times post-infection and thus the T_s NT sequence could affect transcription. It is presently not known whether FPD4 is expressed with a temporal regulation different from its VV counterpart or whether this element provides a means to down-regulate FPD4 expression levels in FPV-infected cells.

From the nucleotide sequence provided in this communication and others reported previously (Drillien et al., 1987; Binns et al., 1988; Tomley et al., 1988), it is apparent that FPV and VV are closely related in at least some genomic regions. Previous reports have found that FPV contains a genomic region closely related to a portion of the VV HindIII J region (Drillien et al., 1987;
Binns et al., 1988). The most notable difference demonstrated was that FPV lacked a TK gene in this region.

These and similar studies with members of the poxvirus family will provide useful information on the biological relevance of certain gene products in the replication of these viruses and may yield insight into functional domains of the gene products. This information will enable the evaluation of the evolutionary pathways that members of the poxvirus family have taken to replicate efficiently in different species.

We acknowledge H. Martinez for the preparation of the oligonucleotides. We would also like to thank K. Dombrowski for the excellent preparation of this manuscript. The sequence retrieved has EMBL Accession No. X17202.

References


