Key words: capripoxvirus/leporipoxvirus/TK gene

The Nucleotide Sequence around the Capripoxvirus Thymidine Kinase Gene Reveals a Gene Shared Specifically with Leporipoxvirus

By PAUL D. GERSHON AND DONALD N. BLACK*

AFRC Institute for Animal Health, Pirbright Laboratory, Ash Road, Pirbright, Woking, Surrey GU24 0NF, U.K.

(Accepted 18 November 1988)

SUMMARY

We have extended previous comparisons of genetic organization between poxvirus genera by sequencing a 2.5K genomic fragment from isolate KS-1 (Kenya sheep-l) of the genus capripoxvirus. The fragment is located in the central region of the capripoxvirus genome and contains three complete and two incomplete open reading frames (ORFs). One of the complete ORFs is a gene for thymidine kinase (TK). This gene, with one of the other two complete ORFs and both the incomplete ORFs, are homologous to four contiguous ORFs from the central region of vaccinia virus (VV) DNA. They also match four ORFs of fowlpox virus (FPV) DNA, three of which are contiguous and the fourth, the FPV TK gene, is located elsewhere on the FPV genome. The third complete ORF of the capripoxvirus DNA fragment is located between the TK gene and the capripoxvirus homologue of the ORF immediately downstream of the VV TK gene. We show that a homologue to this third ORF is absent from VV and FPV DNAs, but is present downstream of the TK gene on Shope fibroma virus DNA. The sequence immediately upstream of the capripoxvirus homologue of a VV late gene contains a motif which is required for VV late gene expression. The motif required for VV early gene transcription termination is present in eight positions in the capripoxvirus sequence, and five of these positions are consistent with the motif having an equivalent function in capripoxvirus to that in VV.

INTRODUCTION

The chordopoxvirinae subfamily is divided, by serological criteria, into six genera (Matthews, 1982). The evolutionary relationships between the genera are at present unclear, and are almost certainly too complex to be ascertained simply by phenotypic comparisons or by comparisons of genomic restriction enzyme sites. The inter-relationships between a number of members of the Herpesviridae have been investigated by comparisons of partial and complete DNA sequences of their genomes (Gompels et al., 1988 and references therein). Substantial sequence data exist for the genome of the orthopoxvirus vaccinia virus (VV), and comparisons with the more limited sequence data available for the genomes of Shope fibroma virus (SFV) of the genus leporipoxvirus, and fowlpox virus (FPV) of the genus avipoxvirus, have identified similarities and differences of genetic organization in specific regions of their genomes. Regions in which comparisons have been made include the genome termini of VV, SFV and FPV DNAs (Baroudy et al., 1982; Venkatesan et al., 1981, 1982; Upton & McFadden, 1986a, b; Upton et al., 1987a, b; W. Chang et al., 1987; Tomley et al., 1988), and the regions of SFV and FPV DNAs that correspond to all or part of the contiguous HindIII fragments L and J from the central region of VV DNA (Plucienniczak et al., 1985; Weir & Moss, 1983; Broyles & Moss, 1986; Rosel et al., 1986; Upton & McFadden, 1986c; Boyle et al., 1987; Drillien et al., 1987; Binns et al., 1988). These sequence data have shown that orthopoxviruses, leporipoxviruses and avipoxviruses possess genes in common and, in terms of genetic organization, resemble one another more closely in the regions corresponding to the centrally placed HindIII L and J fragments of VV.
DNA than they do at their genome termini. The genetic organizations of VV DNA HindIII L/J region and the corresponding regions of FPV and SFV DNAs differ, however, in the following respects: the locus of the thymidine kinase (TK) gene of FPV is not equivalent to that of the VV TK gene which is present in VV DNA HindIII J (Weir et al., 1982; Weir & Moss, 1983; Drillien et al., 1987; Birns et al., 1988); the sequences immediately downstream of the VV and SFV TK genes are not equivalent, despite the equivalence of the genes immediately upstream of the VV and SFV TK genes (Upton & McFadden, 1986c).

There is an absence of reported genomic sequence data for members of the genus capripoxvirus. We have sequenced a capripoxvirus genome fragment which contains a gene for TK. We show that this fragment is comparable in genetic organization to a part of the VV genome which covers the junction of HindIII fragments L and J. However, an additional gene is present immediately downstream of the capripoxvirus TK gene which matches the sequence immediately downstream of the TK gene of SFV. An equivalent to this gene is not present in the genomes of VV and FPV.

**METHODS**

**Materials.** Capripoxvirus isolate ‘Kenya sheep-l’ (KS-1), previously referred to as ‘Kenya sheep and goat pox’, is a vaccine strain (Kitching et al., 1987) derived by the attenuation of sheep isolate 0240 of Kenya sheep and goat pox (Davies, 1976; Davies & Otema, 1981). Restriction digests of the genome of KS-1 (Black et al., 1986, unpublished data) characterize it as type 3 (Gershon et al., 1989), as exemplified by the cattle isolate KC-1 (Gershon & Black, 1987). *Escherichia coli* hosts DH5α, DH5α F', and TG1 were used for cloning.

**Cloning and hybridization.** Plasmid clones of HindIII, PstI and SalI fragments of KS-1 DNA were prepared using standard procedures (Maniatis et al., 1982), and were dot blotted onto nylon membranes (Amersham) according to the manufacturer’s instructions. Alkaline Southern transfers onto nylon membranes were performed according to the protocols of Reed & Mann (1985) and Amersham. Mixed oligonucleotide probes were made using an Applied Biosystems 380A synthesizer, and were hybridized to dot blots or to Southern transfers using standard procedures (Maniatis et al., 1982; Amersham). Hybridizations were performed at 37 °C in 6 × SSC (1 × SSC is 150 mM-NaCl, 15 mM-sodium citrate). Membranes were washed with 2 × SSC at 37 °C, and were re-washed at progressively higher temperatures (increasing in steps of 10 °C).

**Sequencing.** KS-1 DNA HindIII fragment S was cloned in both orientations in the phagemid Bluescript KS plus (Stratagene Cloning Systems). Recombinant phagemid DNA was digested with SalI and BamHI, and nested unidirectional deletions were made using exonuclease III and nuclease S1 (Henikoff, 1984). Single-stranded DNA was generated by rescue of the deleted phagemid clones with M13KO7 [a modified M13 containing the kanamycin resistance locus transposon Tn903 (nucleotide 5825); J. Vieira, personal communication] in the presence of kanamycin, and was sequenced from the universal M13 primer using standard dideoxynucleotide procedures (Sanger et al., 1977; Tabor & Richardson, 1987). Sequences were assembled and analysed using the programs of Eliopoulis et al. (1982) were used to predict protein secondary structure.

**RESULTS AND DISCUSSION**

**Cloning and gene localization**

HindIII fragments of KS-1 DNA are referred to here as shown for KC-1 (Gershon & Black, 1987). Each of the 41 HindIII fragments of KS-1 DNA, except for fragments Q2, T1, V, W1, W2, Y, 1C, 1D1, 1D2, 1E and 1F, was cloned in plasmids pUC9 or pUC19 (Vieira & Messing, 1982). A number of the PstI and SalI fragments of KS-1 DNA were also cloned. These clones, together with the terminal fragment clone InSSE2 from isolate InS-1 (Gershon & Black, 1987), formed a bank representing 94% of the genome of capripoxvirus.

The gene for TK was localized by separately hybridizing two mixed oligonucleotides, 5'-GG(A/C/G/T)CCCATGTT(T/C)G(T/C)TTCA(G/C/T)G(T/C)TTCA(G/C/T)GG-3' and 5'-GA(C/T)GA(A/G)CA(A/G)CA(A/G)TTT(T/C)G(T/C)TG(T/C)TTCA(G/C/T)GG-3', to dot blots of the clone bank described above. These oligonucleotides represent a conserved amino acid sequence from the TK genes of mouse, human, VV (Upton & McFadden, 1986c) and SFV. Specific hybridization to the cloned KS-1 DNA HindIII fragment S was obtained (data not shown).
TK gene sequence of a capripoxvirus

Nucleotide sequence and orientation of KS-1 HindIII fragment S

The 2497 bp sequence of HindIII fragment S was determined (Fig. 1), 84% of the sequence being obtained from both strands and all of the remainder from at least two templates. A single PstI site 42 to 47 bp from the right-hand end of the sequence (Fig. 1) served to orientate the fragment with respect to the genome. The HindIII and PstI site maps of KS-1 DNA are identical to those of the genome of isolate KC-1 (P. D. Gershon & D. N. Black, unpublished data; Gershon & Black, 1987). Since HindIII fragment S of KC-1 DNA hybridizes strongly to PstI A and very weakly to PstI F (P. D. Gershon & D. N. Black, unpublished data), and PstI A lies immediately to the right-hand side of PstI F, the orientation of HindIII S shown here must be opposite to that on the capripoxvirus restriction site maps as depicted previously (Gerson & Black, 1987, 1988). Open reading frames (ORFs) CF6 to CF9 would therefore be orientated from right to left on the genomes of capripoxviruses as shown previously.

A + T content

The sequence has an A + T content of 72.4% and is thus more A + T-rich than the equivalent regions of VV, SFV and FPV, for which values of 66.4%, 61.6% and 68.6%, respectively, were calculated from published sequences. The value for KS-1 DNA HindIII S approaches the theoretical maximum A + T content of 78.8% calculated for a DNA sequence that would be able to code for the amino acid sequence of the KS-1 TK gene (see later). A high A + T content for the whole genome of KS-1 is indicated by the relatively high frequency of A + T-rich restriction enzyme sites, such as that for DraI (TTTAAA) which occurs on average once in every 300 to 1000 bp (P. D. Gershon & D. N. Black, unpublished data), and the relatively low frequency of G + C-rich sites such as those for PstI (CTGCAG) and SalI (GTCGAC) (Gershon...
Since the average frequencies of sites for the A + T-rich cutter *HindIII* (AAGCTT) on the whole genomes of KS-1, VV and SFV are one in every 3-6, 17-8 and 12-5 kb respectively (from Gershon & Black, 1987; Mackett & Archard, 1979; Cabirac et al., 1985), the whole genome of KS-1 appears more A + T-rich than the genomes of VV or SFV.

The high A + T content of KS-1 DNA is not a reflection of the A + T content of the host genome, since analyses of selections of sheep and goat genes give values for A + T content of around 50%. In addition, the parapoxviruses, which have the same host-range as the capripoxviruses, have a genomic A + T content of around 37% (Wittek et al., 1979). The high A + T content of KS-1 DNA may reflect virally controlled aspects of DNA synthesis, such as the characteristics of the capripoxvirus DNA polymerase, or the levels of expression of virally encoded enzymes involved in the biosynthesis of A or T nucleotides, such as TK. In the latter respect, the genomes of poxviruses reported to possess a TK gene are A + T-rich, whereas the presence of a TK gene has not been reported on the G + C-rich genomes of the parapoxviruses.

**Genetic organization of *HindIII* S**

The sequence of *HindIII* S possesses one incomplete and three complete ORFs whose predicted protein products are longer than 90 amino acids. The nucleotide sequences of these four ORFs were screened for matches within the sequence of the 33 kb central region VV DNA covering *HindIII* fragments L, J, H and D (Plucienniczak et al., 1985; Weir & Moss, 1983; Broyles & Moss, 1986; Rosel et al., 1986; Weinrich & Hruby, 1986; Niles et al., 1986). The VV ORFs designated F1 to F12 by Plucienniczak et al. (1985) are referred to here as VF1 to VF12. Two of the complete ORFs and the incomplete ORF of KS-1 *HindIII* S match VV ORFs VF7, VF8 and VF9, respectively, in both sequence and length, and thus are referred to as CF7 to CF9.

By analogy with VV, ORF CF8 of KS-1 is a gene for TK. The two mixed oligonucleotide probes which hybridized to KS-1 *HindIII* S each match sequences within CF8 in up to 16 out of their 17 nucleotides.

The 3' end of VV ORF VF6 overlaps the 5' end of VF7 in a different phase (Plucienniczak et al., 1985). Since the carboxy-terminal 50 amino acids of the predicted product of VF6 match the equivalent phased amino acid sequence translated from the CF7 end of KS-1 *HindIII* S, we have assumed that this 50 amino acid KS-1 sequence represents the 3' end of a larger capripoxvirus ORF, which we have designated CF6. The amino acid sequences of the predicted products of CF6 to CF9 are shown in Fig. 1.

Two KS-1 ORFs, whose putative products are shorter than 90 amino acids, occur on the opposite strand to that containing the ORFs described above. We consider it unlikely that these two short ORFs code for a polypeptide, since their codon usage differs from that of ORFs CF6 to CF9 and their predicted products do not match the translation products of similarly placed 'small' ORFs from VV sequences (Plucienniczak et al., 1985).

The third complete ORF of KS-1 *HindIII* S, which is situated between CF8 and CF9 and designated CF8a, did not match any region of the 33 kb central region of VV DNA. The only sequence in the nucleic acid databases GENBANK 52.0 and EMBL 12 that had greater than 52% sequence identity to CF8a was that present downstream of the TK gene of SFV (Upton & McFadden, 1986c). This sequence comprises the 5' 156 nucleotides of a previously unidentified ORF (referred to here as SF8a), which has 63.2% and 54% homology with the 5' end of CF8a at the nucleotide and amino acid levels respectively. There are no sequences in the database PIR 15-0 with greater than 20% homology to the amino acid sequence predicted from CF8a. In addition, sequences matching CF8a were not found in the FPV DNA sequences directly upstream of FF9, or downstream of the translocated TK gene (Boyle et al., 1987), or in a library of sequences from randomly cloned FPV fragments (Binns et al., 1987; M. M. Binns & M. E. G. Boursnell, personal communication).

To determine whether a homologue of CF8a is present anywhere on the genomes of VV and FPV, mixed oligonucleotides were synthesized representing the amino acid sequence GDSYGCTI, which is completely conserved between the predicted products of the 5' regions of CF8a and SF8a. The 32P-labelled oligonucleotides were hybridized to blots of *NcoI*-digested FPV DNA (strain CM7, kindly provided by M. M. Binns and M. E. G. Boursnell, Institute for...
**TK gene sequence of a capripoxvirus**

Animal Health, Huntingdon, Cambridgeshire, U.K.), HindIII-digested VV strain WR DNA, and HindIII-digested KS-1 DNA. These oligonucleotides hybridized to HindIII fragment S of KS-1 DNA; however, we were unable to detect hybridization to fragments of VV or FPV DNAs under the hybridization and washing conditions employed (data not shown).

The genetic organizations of KS-1 HindIII S and the equivalent regions of VV, SFV and FPV DNAs, shown in Fig. 2, thus differ in the divergent location of the TK gene of FPV (Boyle et al., 1987; Drillien et al., 1987; Binns et al., 1988), and in the presence of an additional ORF, F8a, downstream of ORF F8 of KS-1 and SFV DNAs.

**ORF CF8a**

The predicted 197 amino acid, protein (M, 23048) encoded by CF8a is hydrophilic in character, and contains two possible N-linked glycosylation sites (NXS/T; Marshall, 1972) at amino acids 94 to 96 and 153 to 155. The secondary structure of CF8a protein was predicted by eight methods (Eliopoulos et al., 1982), the consensus of which, in the distribution of \( \alpha \) helix, \( \beta \) strand and random coil, is shown in Fig. 3. The distribution of charges on the protein backbone is also shown in Fig. 3. The protein contains a high proportion of acidic residues, particularly towards its carboxy terminus, with 23% of its amino acids, including 19 of its 26 carboxy-terminal amino acids, being either aspartate or glutamate.

The function of the predicted product of CF8a is not known. The low homology matches in the sequence databases to CF8a or CF8a protein, or the 26 carboxy-terminal amino acids of CF8a protein were unrelated to one another, and included the 17K subunit of yeast ubiquinol-cytochrome c reductase, trout high mobility group proteins, rat spermine-binding protein and the RAD6 gene product of *Saccharomyces cerevisiae*. The acidic region of CF8a protein may permit a pH-dependent conformational change to occur (Van Loon et al., 1984), such a change being from \( \alpha \) helix (Fig. 3) in the uncharged (low pH) state, to random coil in the charged (high pH) state. Other proteins with acidic domains have been shown or postulated to bind either to a positively charged substrate such as a polyamine (Reynolds et al., 1985), or a basic protein subunit such as a histone (Pentecost et al., 1985; Haggren & Kolodrubetz, 1988; C. Chang et al., 1987; Van Loon et al., 1984). In addition, transcription by cellular RNA polymerase II can be activated by proteins with acidic domains (Hope et al., 1988 and references therein), such as the herpes simplex virus protein VP16 (Triezenberg et al., 1988).
Table 1. Homology at the nucleotide/amino acid sequence* levels between the ORFs of KS-1 HindIII fragment S and the equivalent ORFs of VV, SFV and FPV

<table>
<thead>
<tr>
<th>Virus</th>
<th>F6</th>
<th>F7</th>
<th>F8 (TK)</th>
<th>F8a</th>
<th>F9</th>
</tr>
</thead>
<tbody>
<tr>
<td>VV</td>
<td>67.3/56.9†</td>
<td>63.2/53.1</td>
<td>68.4/64.6</td>
<td>GNP†</td>
<td>72.5/76.4†</td>
</tr>
<tr>
<td>SFV</td>
<td>sna§</td>
<td>60.2/56.3</td>
<td>66.3/53.9</td>
<td>62.8/53.8†</td>
<td>sna</td>
</tr>
<tr>
<td>FPV</td>
<td>53.6/42.9†</td>
<td>50.9/39.2</td>
<td>60.3/50.6</td>
<td>GNP</td>
<td>59.3/53.9†</td>
</tr>
</tbody>
</table>

* Sequences were aligned using program GAP (Devereux et al., 1984) and identical amino acids only were scored. Sequence data is from Plucienniczak et al. (1985), Broyles & Moss (1986), Rosel et al. (1986), Upton & McFadden (1986c), Drillien et al. (1987) and Binns et al. (1988).
† Incomplete ORF in one or both sequences.
‡ GNP, Gene not present.
§ SNA, Sequence not available.

Gene homologies

The degrees of homology between the ORFs of KS-1 HindIII S and the equivalent ORFs of VV, FPV and SFV DNAs at the nucleotide and amino acid levels are shown in Table 1. These values indicate that FPV is more distantly related to the mammalian poxviruses than the latter are to one another, which is consistent with the view of Boyle et al. (1987) that the poxviruses diverged from one another in synchrony with host speciation.

Regulatory sequences

Capripoxviruses are able to express transiently genes coupled to the VV early/late promoter P7.5 (P. D. Gershon & D. N. Black, unpublished data), which indicates that they can recognize some or all of the transcriptional regulatory elements of VV.

Motifs with a consensus (A/T)(A/T)TAAAT(A/G)(A/G), in which the triplet shown in bold letters can be a translational initiation codon, have been implicated in the expression and 5' modification of VV late mRNA (De Magistris & Stunnenberg, 1988) and are highly conserved between orthopoxvirus and FPV late genes (Hanggi et al., 1986; Patel & Pickup, 1987; Drillien et al., 1987; Binns et al., 1988). VV ORFs VF6 and VF7 are late genes, and this motif is present at their immediate 5' ends. By analogy with VV we would expect KS-1 ORFs CF6 and CF7 to be late genes. The 5' end of CF6 is not present within HindIII S; however the 5' end of CF7 is present and possesses the sequence AATAAATGGA, which conforms to the consensus. The other immediate 5' ends of ORFs present within HindIII S do not possess such a site. The
presence of the motif in capripoxvirus suggests conservation within the poxviridae of some features of late gene expression. Matches to possible VV late gene modulatory elements (Miner et al., 1988) were not found upstream of CF7.

VV ORFs VF8 and VF9 are transcribed early in infection (Mahf & Roberts, 1984). By comparison with VV, we would expect KS-1 ORFs CF8 and CF9 to be early genes and we presume that CF8a is also transcribed early since the sequence TAA does not occur immediately 5' to its initiation codon.

Transcription termination of VV early genes has been shown to require a motif with the consensus (A)TTTTTNT 20 to 50 nucleotides upstream of the transcription termination site (Yuen & Moss, 1987). Such motifs are present in the 3' regions of all but one of the known early genes in a large sequenced region of VV DNA, the single exception being a TTTTTCT 140 bp from the initiation codon of ORF VF9 (Yuen & Moss, 1987). Matches to the consensus are also present upstream of early genes of SFV (Upton et al., 1987a, b) and some ORFs of FPV (Binns et al., 1987; Tomley et al., 1988). The positions of matches to the consensus in approximately 15 kb sequence of the genome of capripoxvirus KS-1 other than HindIII S (P. D. Gershon & D. N. Black, unpublished data), are consistent with TTTTTNT having a regulatory role in capripoxvirus. However, three of the eight TTTTTNT motifs present in HindIII S are within presumed early genes and more than 50 bp upstream of the stop codon. The three motifs are TTTTTCT within CF8, and TTTTTCT and TTTTTAT within the first 250 bp of CF9 (Fig. 1). Although these three motifs would presumably cause premature termination of early transcripts of CF8 and CF9, it is of interest that the TTTTTCT motif within CF9 is in an equivalent position to its counterpart in VF9, 140 bp from the initiation codon (Yuen & Moss, 1987). Further work is required to determine the mechanisms of transcriptional regulation by capripoxvirus. The non-coding regions immediately downstream of both CF8 and CF8a, the two presumed early genes whose 3' regions are contained within HindIII S, include the sequence TAAATATTAA. This sequence does not occur anywhere else in HindIII S and may have a regulatory function.

We thank Chris Bostock for critical reading of the manuscript, Paul Thomas for technical advice, John McCauley for the synthesis of oligonucleotides, and David Ansell for technical assistance. This study was supported by the Overseas Development Administration, London, U.K.

REFERENCES


TK gene sequence of a capripoxvirus


(Received 12 September 1988)