J. gen. Virol. (1988), **69**, 2995–3003. *Printed in Great Britain*

Key words: *vaccinia virus/serine protease inhibitors/non-essential genes*

Non-essential Genes in the Vaccinia Virus *Hin*dIII K Fragment: a Gene Related to Serine Protease Inhibitors and a Gene Related to the 37K Vaccinia Virus Major Envelope Antigen

By M. E. G. BOURSNELL,* I. J. FOULDS, J. I. CAMPBELL AND M. M. BINNS

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

(*Accepted 18 August 1988*)

SUMMARY

The complete nucleotide sequence of a cloned copy of the *Hin*dIII K fragment of the WR strain of vaccinia virus has been determined. Eight open reading frames (ORFs) have been identified, on the basis of size and codon usage. The predicted amino acid sequences of the putative genes have been compared to the Protein Identification Resource and to published vaccinia virus sequences. One gene, predicted to encode a 42·2K protein, is highly related to the family of serine protease inhibitors. It shows approximately 25% identity to human antithrombin III and 19% identity to the cowpox virus 38K protein gene which is also related to serine protease inhibitors. The product of another gene shows a similar high level of identity to the 37K vaccinia virus major envelope antigen. The existence of viable deletion mutants and recombinants containing foreign DNA inserted into both these genes indicates that they are non-essential.

INTRODUCTION

Vaccinia virus, the prototype member of the orthopoxvirus group, has been much used as a eukaryotic expression vector (for review, see Mackett & Smith, 1986). Certain features of vaccinia virus make it suitable for this purpose. First, it has a large dsDNA genome (approx. 187 kb) with no rigorous packaging requirements. Hence it can accommodate very large amounts of foreign DNA (Smith & Moss, 1983). Second, it has several sites in the genome at which insertion of foreign DNA does not seriously affect replication in tissue culture (Moss *et al.*, 1981; Panicali *et al.*, 1981; Panicali & Paoletti, 1982). Because of the multiplicity of such insertion sites, or non-essential regions, recombinant vaccinia viruses expressing several genes inserted at different loci have been constructed (Perkus *et al.*, 1985). The most commonly used insertion site is the viral thymidine kinase (TK) gene (Mackett *et al.*, 1982). Unlike the TK gene, whose function and sequence are well characterized (Weir & Moss, 1983), several other non-essential regions are not well characterized. One of the reasons why further characterization is important is the possibility that insertion of foreign DNA into certain genes may lead to recombinant vaccinia viruses with reduced virulence. For example, although insertion into the TK gene has no effect on replication in tissue culture, it leads to recombinants with reduced virulence in a mouse model system (Buller *et al.*, 1985). It has recently been shown that insertion into the vaccinia virus growth factor gene also reduces virus virulence (Buller *et al.*, 1988). Because of the complications known to be associated with the use of vaccinia virus as a vaccine (Lane *et al.*, 1969), there is great interest in the production of such attenuated recombinants. Identification of further genes which, when disrupted, produce attenuated viruses is a useful step towards this goal.

One of the regions of the vaccinia virus genome that appears to consist entirely of genes non-essential for growth in many tissue culture systems is that towards the left-hand end of the viral map (see Fig. 1) which is covered by *Hin*dIII restriction fragments N, M, K, F and E (Perkus *et al.*, 1986). Thirteen insertion sites, identified on an empirical basis, have been used. Five of these

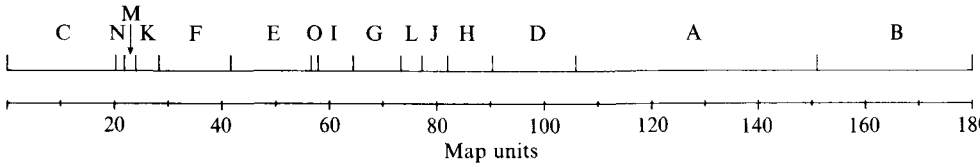

Fig. 1. *Hin*dIII restriction map of the vaccinia virus genome (DeFilippes, 1982)

```
   1 AAGCTTTTCAGCTGCTTAGACTTCCAAGTATTAATTCGTGACAGATCCATGTCTGAAACGAGACGCTAATTAGTGTATATTTTTTCATTTTTTATAATTTTGTCATATTGCACCAGAATT
     F  S  K  L  Q  K  S  K  W  T  N  I  R  S  L  D  M  [ORF KO]

 121 AATAATATCTCTAATAGATCTGATTAGTAGATACATGGCTATCGCAAAACAACATATACACATTTAATAAAAATAATATTTATTAAGAAAATTCAGATTTCACGTACCCATCAATATAAA

 241 TAAAATAATGATTCCTTACACCGTACCCATATTAAGGAGATTCCACCTTACCCATAAACAATATAAATCCAGTAATATCATGTCTGATGATGAACACAAATGGTGTATTAAATTCCAGTT
                                   *  P  S  E  V  K  G  M  F  L  I  F  G  T  I  D  H  R  I  I  F  V  F  P  T  N  F  E  L  K

 361 TTTCAGGAGATGATCTCGCCGTAGCTACCATAATAGTAGATGCCTCTGCTACAGTTCCTTGTTCGTCGACATCTATCTTTGCATTCTGAAACATTTTATAAATATATAATGGGTCCCTAG
     E  P  S  S  R  A  T  A  V  M  I  T  S  A  E  A  V  T  G  Q  E  D  V  D  I  K  A  N  Q  F  M  K  Y  I  Y  L  P  D  R  T

 481 TCATATGTTTAAACGACGCATTATCTGGATTAAACATACTAGGAGCCATCATTTCGGCTATCGACTTAATATCCCTCTTATTTTCGATAGAAAATTTAGGCAGTTTAAGATTGTACACTT
     M  H  K  F  S  A  N  D  P  N  F  M  S  P  A  M  M  E  A  I  S  K  I  D  R  K  N  E  I  S  F  K  P  L  K  L  N  Y  V  K

 601 TATTCCCTAATTGAAACGACCAATAGTCTAATTTTGCAGCCGTAATACAATCTGTGAAATGGGTCATATTATCACCTATTGCCAGGTACATACTAATATTAGCATCCTTATACGGAAGGC
     N  G  L  Q  F  S  W  Y  D  L  K  A  A  T  I  S  D  T  F  H  T  M  N  D  G  I  A  L  Y  M  S  I  N  A  D  K  Y  P  L  R

 721 GTACCATATCATATTCTTCGTCATCGATTGTGATTGTATTTCCTTGCAATTTAGTAACTACGTTCATCATGGGAACCGTTTTCGTACCGTACTTATTAGTAAAACTAGCATTGCGTGTTT
     V  M  D  Y  E  E  D  D  I  T  I  T  N  G  Q  L  K  T  V  V  N  M  M  P  V  T  K  T  G  Y  K  N  T  F  S  A  N  R  T  K

 841 TAGTGATATCAAACGGATATTGCCATATACCTTTAAAATATATAGTATTAATGATTGCCCATAGAGTATTATTGTCGAGCATATTAGAATCTACTACATTAGACATACCGGATCTACGTT
     T  I  D  F  P  Y  Q  W  I  G  K  F  Y  I  T  N  I  I  A  W  L  T  N  N  D  L  M  N  S  D  V  V  N  S  M  G  S  R  R  E

 961 CTACTATAGAATTAATTTTATTAACCGCATCTCGTCTAAAGTTTAATCTATATAGGCCGAATCTATGATATTGTTGATAATACGACGGTTTAATGCACACAGTATTATCTACGAAACTTT
     V  I  S  N  I  K  N  V  A  D  R  R  F  N  L  R  Y  L  G  F  R  H  Y  Q  Q  Y  Y  S  P  K  I  C  V  T  N  D  V  F  S  Q

1081 GATAAGTTAGATCAGTGTACGTATATTTAGATGTTTTCAGCTTAGCTAATCCTGATATTAATTCTGTAAATGCTGGACCCAGATCTCTTTTTCTCAAATCCATAGTCTTCAATAATTCTA
     Y  T  L  D  T  Y  T  Y  K  S  T  K  L  K  A  L  G  S  I  L  E  T  F  A  P  G  L  D  R  K  R  L  D  M  T  K  L  L  E  I

1201 TTCTAGTATTACCTGATGCAGGCAATAGCGACATAAACATAGAAAACGAATAACCAAACGGTGAGAAGACAATATTATCATCTTGAATATTTTTATACGCTACTATACCGGCATTGGTAA
     R  T  N  G  S  A  P  L  L  S  M  F  M  S  F  S  Y  G  F  P  S  F  V  I  N  D  D  Q  I  N  K  Y  A  V  I  G  A  N  T  F

1321 ATCCTTGCAGACGATAGGTAGACACTGAACACGTTAACGATAGTATCAATAACGCAATCATGATTTTATCGTATTAATAATTAACCTTATTTTTATGTTCGGTATAAAAATTATTGATGT
     G  Q  L  R  Y  T  S  V  S  C  T  L  S  L  I  L  L  A  I  M  [ORF K1]                                    *  Q  H

1441 CTACACATCCTTTTGTAATTGACATCTATATATCCTTTTGTATAATCAACTCTAATCACTTTAACTTTTACAGTTTTCCCTACCAGTTTATCCCTATATTCAACATATCTATCCATATGC
     R  C  M  R  K  Y  N  V  D  I  Y  G  K  T  Y  D  V  R  I  V  K  V  K  V  T  K  G  V  L  K  D  R  Y  E  V  Y  R  D  M  H

1561 ATCTTAACACTCTCTGCCAAGATAGCTTCAAAGTGAGGATAGTCAAAAAGATAAATATATAGAGCATAATCCTTCTCGTATACTCTGCCCTTTATTACATCACCCGCATTGGGCAACGAA
     M  K  V  S  E  A  L  I  A  E  F  H  P  Y  D  F  L  Y  I  Y  L  A  Y  D  K  E  Y  V  R  G  K  I  V  D  G  A  N  P  L  S

1681 TAACAAAATGCAAGCATCTTGTTAACGGGCTCGTAAATTGGGATAAAAATTATGTTTTTATATCTATTTTATTCAAGAGAATATTCAGGAATTTCTTTTTCCGGTTGTATCTCATCGCAG
     Y  C  F  A  L  M  [ORF K2]                                          *  E  L  S  Y  E  P  I  E  K  E  P  Q  I  E  D  C

1801 TATATATCATTTGTACATTGTTTCATATTTTTTAATAGTCTACACCTTTTAGTAGGACTAGTATCGTACAATTCATAGCTGTATTTTGAATTCCAATCACGCATAAAAATATCTTCCAAT
     Y  I  D  N  T  C  Q  K  M  N  K  L  L  R  C  R  K  T  P  S  T  D  Y  L  E  Y  S  Y  K  S  N  W  D  R  M  F  I  D  E  L

1921 TGTTGACGAAGACCTAATCCATCATCCGGTGTAATATTAATAGATGCTCCACATGTATCCGTAAAGTAATTTCCTGTCCAATTTGAGGTACCTATATACGCCGTTTTATCGGTTACCATA
     Q  Q  R  L  G  L  G  D  D  P  T  I  N  I  S  A  G  C  T  D  T  F  Y  N  G  T  W  N  S  T  G  I  Y  A  T  K  D  T  V  M

2041 TATTTGGCATGGTTTACCCTAGAATACGGAATGGGAGGATCAGCATCTGGTACAATAAATAGCTTTACTTCTATATTTATGTTTTTAGATTTTAGCATAGCGATAGATCTTAAAAAGTTT
     Y  K  A  H  N  V  R  S  Y  P  I  P  P  D  A  D  P  V  I  F  L  K  V  E  I  N  I  N  K  S  K  L  M  A  I  S  R  L  F  N

2161 CTCATGATAAACGAAGATCGTTGCCAGCAACTAATCAATAGCTTAACTGACACTTGTCTGTCTATAGCGGCTCTTCTTAATTCATCTTCTATATAAGGCCAAAACAAAATATTGCCTGCC
     R  M  I  F  S  S  R  Q  W  C  S  I  L  L  K  V  S  V  Q  R  D  I  A  A  R  R  L  E  D  E  I  Y  P  W  F  L  I  N  G  A

2281 TTCGAATAAATAATAGGGATAAAGTTCATAACAGATACATAAACGAATTTACTCGCATTTCTGATACATGACAATAAAGCGGTTAAATCATTGGTTCTTTCCATAGTACATAGTTGTTGC
     K  S  Y  I  I  P  I  F  N  M  V  S  V  Y  V  F  K  S  A  N  R  I  C  S  L  L  A  T  L  D  N  T  R  E  M  T  C  L  Q  Q

2401 GGTGCAGAAGCAATAAATACAGAGTGTGCAACGCCGCTTACGTTAATACTAAGAGGATGATCTGTATTATAATACGACGGATAAAAGTTTTTCCAATTATATGGTAGATTGTTAACTCCA
     P  A  S  A  I  F  V  S  H  P  V  G  S  V  N  I  S  L  P  H  D  T  N  Y  Y  S  P  Y  F  N  K  W  N  Y  P  L  N  N  V  G

2521 AGATACCAGTATACCTCAAAAATTTGAGTGAGATCCGCTGCCAAGTTCCTATTATTGAAGATCGCAATACCCAATTCTTTGACCTGAGTTAGTGATCTCCAATCCATGTTAGCGCTTCCT
     L  Y  W  Y  V  E  F  I  Q  T  L  D  A  A  L  N  R  N  N  F  I  A  I  G  L  E  K  V  Q  T  L  S  R  W  D  M  N  A  S  G

2641 AAATAAATATGTGTATTATCAGATATCCAAAATTTTGTATGAAGAACTCCTCCTAGGATATTTGTAATATCTATGTATCGTACTTCAACTCCGGCCATTTGTAGTCTTTCAACATCCTTT
     L  Y  I  H  T  N  D  S  I  W  F  K  T  H  L  V  G  G  L  I  N  T  I  O  I  Y  R  V  E  V  G  A  M  Q  L  R  E  V  D  K

2761 AATGGTTTGTTAGATTTATTGACGGCTACTCTAACTCGTACTCCTCTTTTGGGTAATTGTACAATCTTGTTTAATATTATCGTGCCGAAATTCGTACCCACTTCATCCGATAAACTCCAA
     L  P  K  N  S  K  N  V  A  V  R  V  R  V  G  R  K  P  L  Q  V  I  K  N  L  I  I  T  G  F  N  T  G  V  E  D  S  L  S  W  Y
```

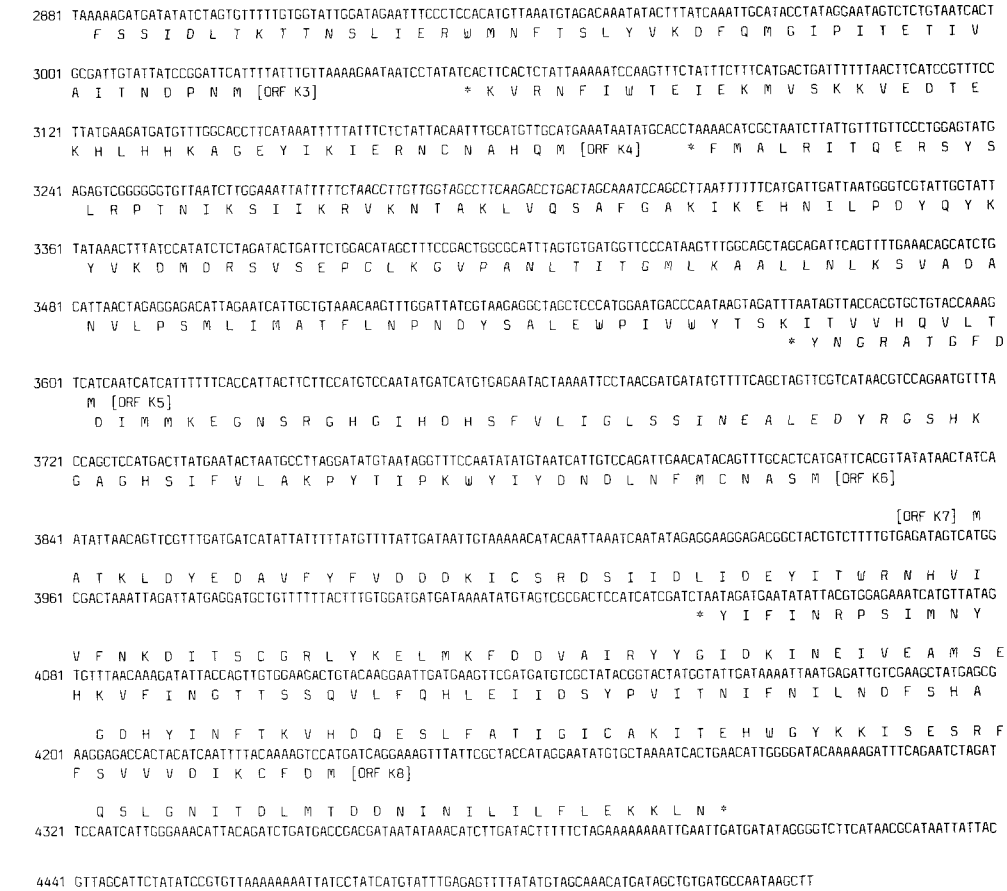

Fig. 2. The nucleotide sequence of the *Hin*dIII K fragment of the WR strain of vaccinia virus. Translation in single-letter amino acid code of eight selected ORFs (K1 to K8) is shown.

sites either flank, or are within, the 4·5 kb *Hin*dIII K fragment. In this paper we present the nucleotide sequence of this fragment, revealing the presence of eight genes, two of which can have a function tentatively assigned to them.

METHODS

The vaccinia virus *Hin*dIII K fragment originally from vaccinia virus strain WR, cloned in pBR322, was a gift from R. Wittek. The DNA was transformed into *Escherichia coli* strain TG1 and CsCl gradient-purified DNA was prepared. For shotgun sequencing, subclones of the plasmid were obtained by sonicating the DNA (Deininger, 1983) and ligation into *Sma*I-cut M13mp10 (Amersham). Bacterial colonies containing M13 with inserts were grown, transferred to nitrocellulose filters and probed with labelled purified viral insert. Single-stranded templates were prepared from M13 clones identified by probing as containing viral DNA. Dideoxy sequencing (Sanger *et al.*, 1977; Bankier & Barrell, 1983) was carried out using [α-^{35}S]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 minicomputer using the programs of Staden (1982, 1984*a*,*b*). Comparisons against the Protein Identification Resource (George *et al.*, 1986) were made with the program FASTP (Lipman & Pearson, 1985). Optimal alignments between individual sequences were obtained using the University of Wisconsin program GAP (Devereux *et al.*, 1984).

RESULTS

The complete sequence of the *Hin*dIII K fragment has been obtained. Over 92% of the sequence was obtained from both strands and over 98·7% of the sequence was obtained from at least two separate M13 clones. The sequence of 4536 bp is shown in Fig. 2. The sequence has

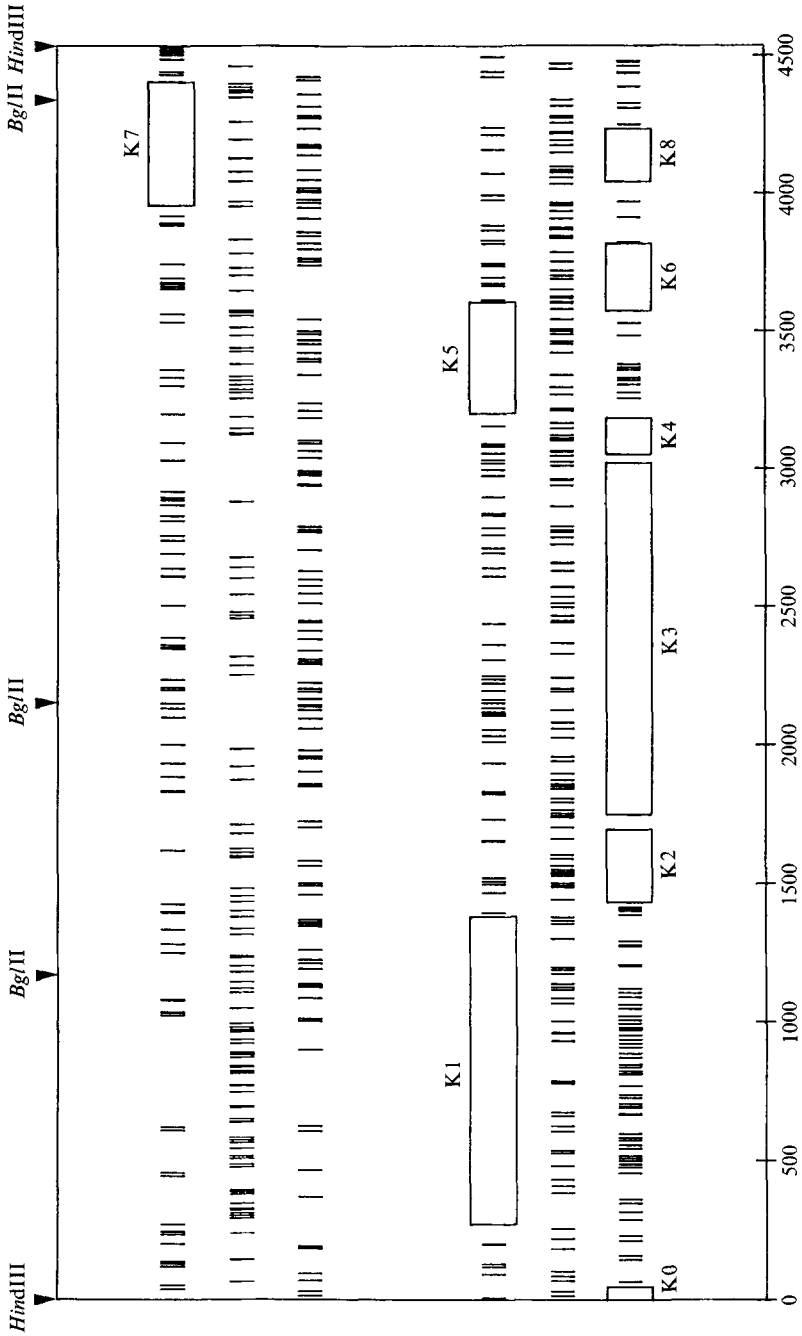

Fig. 3. Diagram showing the positions of the eight selected ORFs. Numbered boxes show the ORFs and vertical lines show stop codons in all six possible reading frames. The top three frames are shown 5′ to 3′ and the bottom three are shown 3′ to 5′. The black triangles at the top of the diagram show the positions of the two terminal *Hin*dIII sites and the three internal *Bgl*II sites which have been used as insertion sites (Perkus *et al.*, 1986).

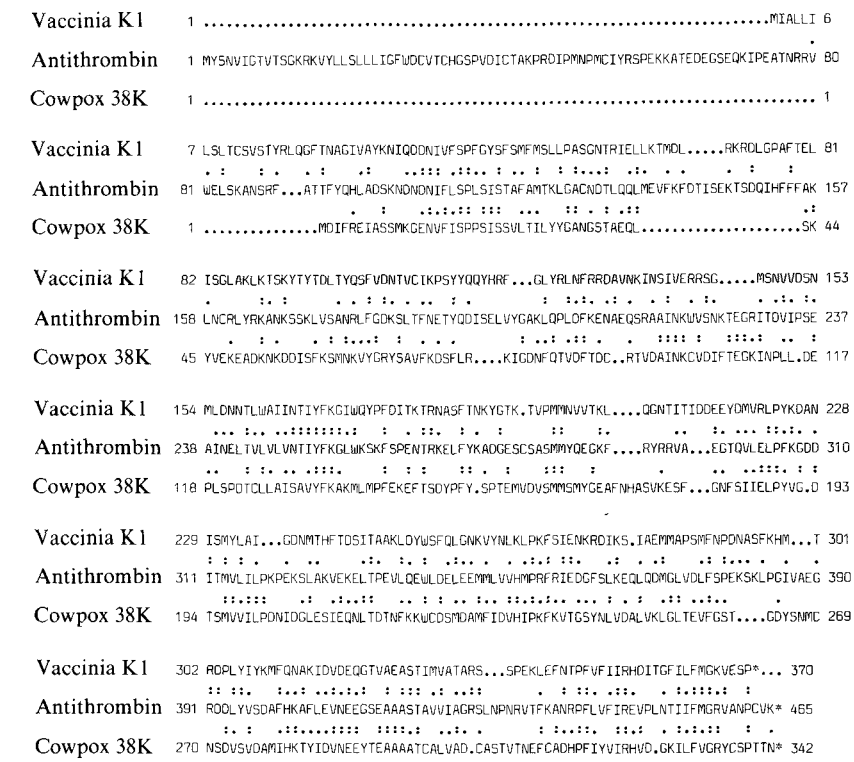

Fig. 4. A comparison of the predicted amino acid sequences of the vaccinia virus K1 gene, human antithrombin III, and the cowpox virus 38K gene. The optimal alignment was achieved using the program GAP (Devereux *et al.*, 1984). Dots within the sequences show where gaps have been introduced to achieve alignment. Colons between the sequences show identical amino acids and dots show conservative changes.

been analysed for open reading frames (ORFs) and Fig. 3 shows diagrammatically the main ORFs, selected by size and by codon usage (Staden, 1984*b*; Staden & McLachlan, 1982). ORFs that are contained within other larger ORFs, either on the same strand or on the other strand, are not shown except when they conform to the codon usage derived from a large selection of previously sequenced vaccinia virus genes. This is the case for ORFs K7 and K8. Small ORFs, such as K4, are also only selected if they exhibit strong vaccinia virus codon usage. Apart from the ORFs that overlap the terminal *Hin*dIII sites there are eight ORFs shown which have been designated K1 to K8. Translations of these ORFs, in single-letter amino acid code, are shown on Fig. 2.

Codon usage analysis of the ORFs in this fragment reveals some reading frames that are open but have no AUG methionine codon at the start. An example of this is the sequence upstream of K4. This, combined with the observation that in the C-terminal 18 amino acids of K5 the vaccinia virus codon usage drops off dramatically, suggests that K4 should be joined to K5 in one continuous ORF. However, the sequence at all such points has been carefully checked and appears to be correct.

The predicted amino acid sequences of ORFs K1 to K8 were compared to the Protein Identification Resource (George *et al.*, 1986) using the computer program FASTP (Lipman & Pearson, 1985). Two highly significant similarities were found. ORF K1 has a high level of identity with the large super-family of serine protease inhibitors, notably human alpha-1-antitrypsin and antichymotrypsin, chicken gene Y and gene X proteins, and human antithrombin III. When the amino acid sequences of ORF K1 and human antithrombin were compared, with suitable gaps inserted to achieve optimal alignment (Devereux *et al.*, 1984) it was found that, within the aligned regions, 25% of amino acids matched perfectly, with a further 25% showing conservative changes. This alignment is shown in Fig. 4. In the same diagram the

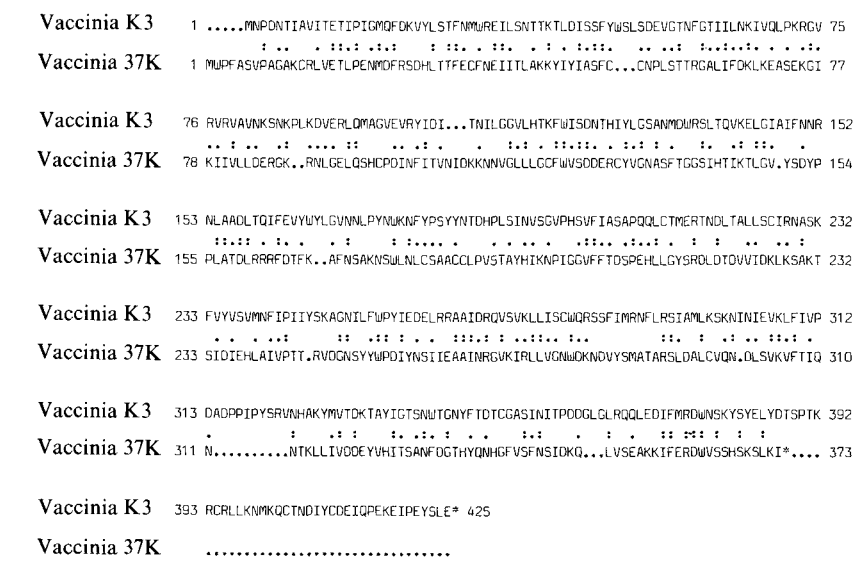

Fig. 5. A comparison of the predicted amino acid sequences of the vaccinia virus K3 gene and the 37K major envelope antigen. The optimal alignment was achieved using GAP (Devereux *et al.*, 1984). Dots within the sequences show where gaps have been introduced to achieve alignment. Colons between the sequences show idential amino acids and dots show conservative changes.

match of antithrombin III to another poxvirus gene is shown, namely to the cowpox virus gene that causes the red pock phenotype (Pickup *et al.*, 1986). ORF K3 shows a similar high level of identity (28% identity plus 29% conservative changes) with a previously sequenced vaccinia virus gene which has been shown to encode a major 37K envelope antigen (Hirt *et al.*, 1986). This match is shown in Fig. 5. A third match, which is not shown, is that of ORF K7 to the gene designated T3A in the terminal region of Shope fibroma virus (Upton *et al.*, 1987) with 16% identical amino acids and a further 53% conservative changes.

Five sites within this fragment have been used as insertion points for foreign genes (Perkus *et al.*, 1986). These are the *Hin*dIII sites at the ends of the fragment, and three *Bgl*II sites within the fragment. These *Bgl*II sites at positions 1160, 2146 and 4343 fall within K1, K3 and K7 (Fig. 3). The left-hand *Hin*dIII site falls within the ORF shown as K0 in Fig. 3. This is the N terminus of the host range gene, which allows the virus to replicate in human cells (Gillard *et al.*, 1986). The right-hand *Hin*dIII site appears to fall within a gene which continues into *Hin*dIII F.

DISCUSSION

Analysis of the nucleotide sequence of the *Hin*dIII K fragment of vaccinia virus strain WR has identified eight ORFs. Examination of the sequence shows that three of these ORFs have been used as sites for insertion of foreign DNA (Perkus *et al.*, 1986). Two of these genes, K1 and K3, can have functions tentatively assigned to them on the basis of high degrees of similarity to previously sequenced genes. A third ORF, K7, has similarity to an ORF from the terminal regions of Shope fibroma virus.

At present it is not known which, if any, of these genes are transcribed in tissue culture. Belle Isle *et al.* (1981) have looked at the *in vitro* translation products from early and late mRNAs selected using cloned *Hin*dIII fragments. No late mRNAs were selected from the *Hin*dIII K fragment, which is consistent with the fact that analysis of the upstream sequences suggests that none of the genes in this fragment are late genes. Five protein products were translated from early mRNA, of sizes 46K, 30K, 16K, 11K and 9K. The 30K protein was also translated from mRNA selected by *Hin*dIII M, the next fragment to the left. Thus this is probably the 32·5K host range gene (Gillard *et al.*, 1986) most of which is in *Hin*dIII M. The eight possible genes in

Table 1. *Percentage identity after optimal alignment between the vaccinia virus K1 gene, the cowpox virus 38K gene, human antithrombin III, human alpha-1-antitrypsin and the fowlpox virus ORF3 gene*

	Vaccinia K1	Cowpox 38K	Antithrombin	Antitrypsin	Fowlpox ORF3
Vaccinia K1	100·0	19·2	25·5	24·2	22·2
Cowpox 38K	19·2	100·0	30·0	25·7	23·6
Antithrombin	25·5	30·0	100·0	27·3	22·1
Antitrypsin	24·2	25·7	27·3	100·0	32·4
Fowlpox ORF3	22·2	23·6	22·1	32·4	100·0

*Hin*dIII K, K1 to K8, encode proteins of predicted sizes 42·2K, 10·5K, 48·8K, 5·3K, 15·1K, 9·1K, 17·4K and 7·4K respectively. Several of these show good agreement with the sizes of the *in vitro* polypeptides.

ORF K1, potentially encoding a protein of 42·2K, is very similar to members of the superfamily of serine protease inhibitors. Several of the members of this family are present in blood plasma and act as anticoagulants. For example antithrombin III inhibits thrombin, a protein involved in blood clotting. Other poxvirus genes with similarity to serine protease inhibitors have been found in cowpox virus (Pickup *et al.*, 1986) and in fowlpox virus (Tomley *et al.*, 1988). The cowpox gene, whose nucleotide sequence predicts a protein of size 38K, has been shown to be involved in the haemorrhaginous lesions, or red pocks, caused by the virus (Pickup *et al.*, 1986). Lack of the gene causes white pock variants, presumably because inhibition of proteins involved in blood clotting no longer occurs.

The vaccinia virus K1 gene maps at 25 kb from the left end of the genome whereas the cowpox virus 38K gene maps at approximately 32 kb from the right-hand end of the cowpox virus genome. Hence both of these genes map in that central region of the orthopoxvirus genome which appears to be highly similar between vaccinia and cowpox viruses, by restriction mapping (Mackett & Archard, 1979) and probably at a sequence level (Pickup *et al.*, 1986). The positioning and degree of identity of these two genes suggests therefore that they are not direct equivalents. In fact the cowpox virus 38K protein and the vaccinia virus 42K K1 protein are slightly less closely related to each other than either is to the serine protease inhibitor superfamily as a whole. Table 1 shows the percentage of identical amino acids found in individual pairwise comparisons between several of these proteins. It seems likely, however, that a cowpox virus equivalent of the vaccinia virus K1 gene exists in the corresponding place in the cowpox virus genome and that a vaccinia virus equivalent of the cowpox virus 38K gene exists in the corresponding place in the vaccinia virus genome.

The 38K cowpox virus gene is one of the most highly expressed early genes in the virus. An 11 bp sequence, GAAAATATATT, has been found upstream of the gene, and is also upstream of the vaccinia virus 7·5K gene (Pickup *et al.*, 1986; Venkatesan *et al.*, 1981). The vaccinia virus 42K K1 gene also appears to be an early gene, in that it does not have the characteristic late TAAAT motif (Hanggi *et al.*, 1986; Rosel *et al.*, 1986) immediately before the ATG codon, and has a potential early TTTTTAT termination of transcription signal (Rohrmann *et al.*, 1986) 78 bases downstream from the termination codon. However, it does not share any sequence homology with sequences upstream of the 7·5K gene.

Another gene from the vaccinia virus *Hin*dIII K fragment, ORF K3, potentially encoding a 48·8K protein, shows extensive similarity along most of its length to the p37K major envelope antigen of vaccinia virus (Hiller *et al.*, 1981; Hirt *et al.*, 1986). Twenty-five per cent of residues are identical and a further 23% are conservative changes. At 41·7K, the predicted size of the p37K envelope protein is somewhat smaller than that of the K3 48·8K predicted product. After optimal alignment (see Fig. 5) the K3 gene extends for 37 extra amino acids at the C-terminal end. The p37K protein resides in the viral envelope present on extracellular vaccinia virus but not on virus purified from within cells (Hiller *et al.*, 1981). Hirt *et al.* (1986) highlight two

relatively hydrophobic regions of the sequence (130 to 157 and 175 to 192) which they suggest could represent putative membrane anchors. The sequence of the K3 gene also predicts hydrophobic domains, which are not, however, in an equivalent position (for example 190 to 255 and 275 to 312). Neither of the polypeptide sequences contains an N-terminal region characteristic of a signal peptide (McGeoch, 1985). Both monoclonal and polyclonal antibodies to the p37K protein are highly specific (Hiller *et al.*, 1981; Hirt *et al.*, 1986) and no other immunologically related polypeptide which could be the K3 product can be seen. However, although p37K appears to be the major new protein present on extracellular virus (Hiller *et al.*, 1981) other proteins with sizes ranging from 20K to 210K have been detected as new extracellular products (Payne, 1978). The nearest of these in size to the 48·8K predicted for the K3 protein is a 42K glycosylated polypeptide.

The p37K product is a late gene, and it has the characteristic TAAATG motif which is present in most vaccinia virus late genes. The sequences upstream of the ATG codon of the K3 gene have no TAAAT sequence, but rather a TAAAATG sequence. Combined with the fact that there is a possible early termination of transcription signal, TTTTTAT, downstream of the gene, this suggests, rather surprisingly, that this may be an early gene.

In conclusion, the sequence of the 4·5 kb vaccinia virus *Hin*dIII K fragment has revealed the presence of eight potential genes, all of which are non-essential in certain tissue culture systems (Perkus *et al.*, 1986). Any of these genes may encode factors that affect the virulence of the virus. Large deletions in the vaccinia virus genome (Dallo & Esteban, 1987) and insertions into individual genes (Buller *et al.*, 1985) have been found to produce attenuated variants of vaccinia virus. The complete nucleotide sequence of this fragment now allows the possibility of directed insertions into each of the genes in order to determine whether specific disruption of any of them leads to viruses with altered pathogenicity or reduced virulence. In particular, the K1 gene may be a suitable candidate, as it shows a high level of identity with a cowpox virus gene whose product is known to affect the growth of cowpox virus *in vivo*.

We thank Philip Green for excellent technical assistance, and Dr Riccardo Wittek for donating the cloned *Hin*dIII K fragment.

REFERENCES

BANKIER, A. & BARRELL, B. G. (1983). Shotgun DNA sequencing. In *Techniques in the Life Sciences* (*Biochemistry*), vol. B5, *Techniques in Nucleic Acid Biochemistry*, pp. 1–34. Edited by R. A. Flavell. Amsterdam: Elsevier.

BELLE ISLE, H., VENKATESAN, S. & MOSS, B. (1981). Cell-free translation of early and late mRNAs selected by hybridization to cloned DNA fragments derived from the left 14 million to 72 million daltons of the vaccinia virus genome. *Virology* **112**, 306–317.

BIGGIN, M. D., GIBSON, T. J. & HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 3963–3965.

BULLER, R. M. L., SMITH, G. L., CREMER, K., NOTKINS, A. & MOSS, B. (1985). Decreased virulence of recombinant vaccinia virus expression vectors is associated with a thymidine-kinase negative phenotype. *Nature, London* **317**, 813–815.

BULLER, R. M. L., CHAKRABARTI, S., COOPER, J., TWARDZIK, D. & MOSS, B. (1988). Deletion of the vaccinia virus growth factor gene reduces virus virulence. *Journal of Virology* **62**, 866–874.

DALLO, S. & ESTEBAN, M. (1987). Isolation and characterization of attenuated mutants of vaccinia virus. *Virology* **159**, 408–422.

DEFILIPPES, F. M. (1982). Restriction enzyme mapping of vaccinia virus DNA. *Journal of Virology* **43**, 136–149.

DEININGER, P. L. (1983). Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Analytical Biochemistry* **129**, 216–223.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

GEORGE, D. G., BARKER, W. C. & HUNT, L. T. (1986). The protein identification resource (PIR). *Nucleic Acids Research* **14**, 11–15.

GILLARD, S., SPEHNER, D., DRILLIEN, R. & KIRN, A. (1986). Localization and sequence of a vaccinia virus gene required for multiplication in human cells. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 5573–5577.

HANGGI, M., BANNWARTH, W. & STUNNENBERG, H. G. (1986). Conserved TAAAT motif in vaccinia virus late promoters: overlapping TATA box and site of transcription initiation. *EMBO Journal* **5**, 1071–1076.

HILLER, G., EIBL, H. & WEBER, K. (1981). Characterization of intracellular and extracellular vaccinia virus variants: N1-isonicotinyl-N2-3-methyl-4-chlorobenzoylhydrazine interferes with cytoplasmic virus dissemination and release. *Journal of Virology* **39**, 903–913.

HIRT, P., HILLER, G. & WITTEK, R. (1986). Localization and fine structure of a vaccinia virus gene encoding an envelope antigen. *Journal of Virology* **58**, 757–764.

LANE, J. M., RUBEN, F. L., NEFF, J. M. & MILLAR, J. D. (1969). Complications of smallpox vaccination, 1968. National surveillance in the United States. *New England Journal of Medicine* **281**, 1201–1208.

LIPMAN, D. J. & PEARSON, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435–1441.

McGEOCH, D. J. (1985). On the predictive recognition of signal peptide sequences. *Virus Research* **3**, 271–286.

MACKETT, M. & ARCHARD, L. C. (1979). Conservation and variation in *Orthopoxvirus* genome structure. *Journal of General Virology* **45**, 683–701.

MACKETT, M. & SMITH, G. (1986). Vaccinia virus expression vectors. *Journal of General Virology* **67**, 2067–2082.

MACKETT, M., SMITH, G. L. & MOSS, B. (1982). Vaccinia virus: a selectable eukaryotic cloning and expression vector. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 7415–7419.

MOSS, B., WINTERS, E. & COOPER, J. A. (1981). Deletion of a 9,000 base pair segment of the vaccinia genome that codes for non-essential polypeptides. *Journal of Virology* **40**, 387–395.

PANICALI, D. & PAOLETTI, E. (1982). Construction of poxviruses as cloning vectors: insertion of the thymidine kinase gene from herpes simplex virus into the DNA of infectious vaccinia virus. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 4927–4931.

PANICALI, D., DAVIS, S. W., MERCER, S. R. & PAOLETTI, E. (1981). Two major DNA variants present in serially propagated stocks of the WR strain of vaccinia virus. *Journal of Virology* **37**, 1000–1010.

PAYNE, L. (1978). Polypeptide composition of extracellular enveloped vaccinia virus. *Journal of Virology* **27**, 28–37.

PERKUS, M. E., PICCINI, A., LIPINKSKAS, B. R. & PAOLETTI, E. (1985). Recombinant vaccinia virus: immunization against multiple pathogens. *Science* **229**, 981–984.

PERKUS, M. E., PANICALI, D., MERCER, S. & PAOLETTI, E. (1986). Insertion and deletion mutants of vaccinia virus. *Virology* **152**, 285–297.

PICKUP, D. J., INK, B. S., HU, W., RAY, C. A. & JOKLIK, W. K. (1986). Hemorrhage in lesions caused by cowpox virus is induced by a viral protein that is related to plasma protein inhibitors of serine proteases. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 7698–7702.

ROHRMANN, G., YUEN, L. & MOSS, B. (1986). Transcription of vaccinia virus early genes by enzymes isolated from vaccinia virions terminates downstream of a regulatory sequence. *Cell* **46**, 1029–1035.

ROSEL, J., EARL, P. J., WIRE, J. & MOSS, B. (1986). Conserved TAAATG sequence at the transcriptional and translational initiation sites of vaccinia virus late genes deduced by structural and functional analysis of the *Hin*dIII genome fragment. *Journal of Virology* **60**, 436–449.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SMITH, G. & MOSS, B. (1983). Infectious poxvirus vectors have capacity for at least 25,000 base pairs of foreign DNA. *Gene* **25**, 21–28.

STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731–4751.

STADEN, R. (1984*a*). Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Research* **12**, 521–538.

STADEN, R. (1984*b*). Measurements of the effects that coding for a protein has on DNA sequence and their use for finding genes. *Nucleic Acids Research* **12**, 551–567.

STADEN, R. & McLACHLAN, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Research* **10**, 141–157.

TOMLEY, F., BINNS, M., CAMPBELL, J. & BOURSNELL, M. (1988). Sequence analysis of an 11·2 kilobase, near-terminal, *Bam*HI fragment of fowlpox virus. *Journal of General Virology* **69**, 1025–1040.

UPTON, C., DELANGE, A. M. & McFADDEN, G. (1987). Tumorigenic poxviruses: genomic organization and DNA sequence of the telomeric region of the Shope fibroma virus genome. *Virology* **160**, 20–30.

VENKATESAN, S., BAROUDY, B. M. & MOSS, B. (1981). Distinctive nucleotide sequences adjacent to multiple initiation and termination sites of an early vaccinia virus gene. *Cell* **25**, 805–813.

WEIR, J. P. & MOSS, B. (1983). Nucleotide sequence of the vaccinia virus thymidine kinase gene and the nature of spontaneous frameshift mutations. *Journal of Virology* **46**, 530–537.

(*Received 28 June 1988*)