Vaccinia Virus Encodes a Family of Genes with Homology to Serine Proteinase Inhibitors

By GEOFFREY L. SMITH,* SUSAN T. HOWARD AND Y. SANG CHAN
Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QF, U.K.

(Accepted 10 May 1989)

SUMMARY
Nucleotide sequencing of a region of the vaccinia virus genome proximal to the right inverted terminal repeat (ITR) identified two open reading frames (ORFs) encoding proteins of 39K and 40K with amino acid homology to each other, to another vaccinia virus gene near the opposite end of the virus genome and to the superfamily of serine proteinase inhibitors (serpins). Serpins have now been found in poxviruses from the genera orthopox (cowpox and vaccinia viruses), leporipox (myxoma virus) and avipox (fowlpox virus). One of the vaccinia virus serpins identified here (B13R) shares 92% amino acid identity with the serpin from cowpox virus and 46% and 19% identity with vaccinia serpins B24R and K2L, respectively. The amino acid sequence of B13R reported here differs at 11 positions from a recently reported sequence and contains an additional three internal residues. The serpin genes near the right ITR are separated by 8 kb of DNA. Both genes contain early transcriptional termination signals just downstream of the ORFs and are transcribed in a rightward direction towards the end of the genome. Analysis of mRNAs from virus-infected cells demonstrated that all three vaccinia virus serpin genes are transcribed early during infection. The amino acid sequences at the active sites of these serpins suggest that they may inhibit serine proteinases of differing biochemical specificities. The possible functions of these genes are discussed.

INTRODUCTION
Vaccinia virus is a DNA virus that replicates in the cytoplasm of infected cells and is the most intensively studied member of the poxvirus family. Properties of this virus family include a large complex virus particle, a linear double-stranded DNA genome and possession of many virus-encoded enzymes that permit transcription and replication in the cytoplasm in a manner independent of host RNA and DNA polymerases (Moss, 1985).

The development of techniques permitting the construction of recombinant vaccinia viruses that express genes from heterologous pathogens has raised the possibility of re-using vaccinia virus as a vaccine against diseases other than smallpox (Smith et al., 1983; Panicali et al., 1983). For vaccinia virus to be used again as a human vaccine it is necessary to construct safer attenuated strains of the virus that do not cause the rare vaccine-related complications (Lane et al., 1969). This requires the identification and deletion of genes that are non-essential for virus replication but which influence the ability of the virus to spread and cause disease in the host.
The thymidine kinase (TK) (Buller et al., 1985), haemagglutinin (Flexner et al., 1987) and growth factor (Buller et al., 1988) have all been shown to influence vaccinia virus pathogenicity. To identify additional genes such as these, and to find other non-essential sites into which foreign DNA may be inserted, we sequenced the region of the 185 kb vaccinia virus genome adjacent to the right inverted terminal repeat (ITR). We chose this region because (i) in a similar region of the closely related cowpox virus genome there is a gene encoding a serine proteinase inhibitor (serpin) which is non-essential for cowpox virus replication and which is responsible for the haemorrhagic pock phenotype (Pickup et al., 1986) and (ii) the genomes of
orthopoxviruses are most variable within or adjacent to the ITRs (Mackett & Archard, 1979). Deletions and transpositions have been shown to occur in these regions of vaccinia (Panicali et al., 1981; Kotwal & Moss, 1988a), monkeypox (Esposito et al., 1981), rabbitpox (Moyer et al., 1980) and cowpox (Pickup et al., 1984) viruses, and therefore the genes encoded here are non-essential for virus replication. Consistent with this, most conditional lethal mutants of vaccinia virus map to within the central highly conserved region of the genome and not near the variable termini (Condit et al., 1983). The genes encoded near the termini are more likely to influence virus properties such as host range and virulence. Indeed a gene has been found near the left ITR that is required for virus replication in human cells (Gillard et al., 1986). Another gene encodes a secretory protein with homology to the C4B-binding protein (a complement control factor) and may serve to prevent complement-mediated lysis of virus-infected cells (Kotwal & Moss, 1988b).

In contrast to the detailed information about genes mapping near the left ITR (Kotwal & Moss, 1988a; Boursnell et al., 1988; Gillard et al., 1986), little information is available about the equivalent region at the other end of the virus genome.

We report here the nucleotide sequence and deduced amino acid sequence of two vaccinia virus genes with homology to serpins. The sequences of both of these genes have recently been reported (Kotwal & Moss, 1989) and one gene is copied at the opposite end of the virus genome in a transposition mutant (Kotwal & Moss, 1988a). Our sequence of one of them (termed B13R) contains 11 amino acid differences and an extra three internal amino acid residues. A third vaccinia virus serpin gene has also been identified and sequenced (Boursnell et al., 1988). The amino acid sequences at the serpin active sites indicate that the three proteins may inhibit serine proteinases of different specificity. Transcriptional analyses show that all these serpin genes are expressed early during infection.

METHODS

Nucleotide sequencing. Plasmids containing the vaccinia virus (strain WR) restriction fragments SalI and the leftmost HindIII–SalI fragment of HindIII B were kindly provided by Mike Merchlinsky and Bernard Moss (National Institutes of Health, Bethesda, Md., U.S.A.). The vaccinia virus WR SalI G fragment was cloned into plasmid pUC13 after SalI digestion of cosmid 6 which contains this region of the genome of a rifampicin-resistant mutant (Baldick & Moss, 1987). The SalI I and G fragments were excised from plasmid sequences by SalI digestion, resolved by electrophoresis through agarose gels, purified and self-ligated with T4 DNA ligase. Circularized molecules were randomly fragmented by sonication, the termini repaired by treatment with T4 DNA polymerase and Klenow enzyme and blunt-ended molecules greater than 200 bp in length were ligated into Smal-cut M13mp18. After transformation of Escherichia coli TG2, white plaques were picked and used to infect fresh E. coli TG2 cultures. Single-stranded DNA purified from culture supernatants 5 h after infection was sequenced by the dideoxy nucleotide chain termination method (Sanger et al., 1977) using 35S-labelled dATP and buffer gradient polyacrylamide gels (Biggin et al., 1983). For further details of these procedures see Bankier & Barrell (1983).

Computer analyses. Nucleotide sequence data were read directly into a computer from autoradiographs of dried gels using a sonic digitizer (Scientific Accessories Corporation). Random data were assembled into contiguous sequences using programs DBUTIL and DBAUTO (Staden, 1982) on a VAX 8350 computer. Open reading frames (ORFs) were identified using ANALYSEQ (Staden, 1982) and ORFFILE (kindly provided by Mike Boursnell, Institute of Animal Health, Houghton, U.K.) and the deduced amino acid sequences were compared with protein database SWISSPROT using the program FASTP (Lipman & Pearson, 1985). Alignment of multiple protein sequences was performed by the MULTALIGN program (G. Barton, Birkbeck College, London, U.K.).

RNA preparation. Human TK-143 cells were infected with vaccinia virus strain WR at 30 p.f.u./cell in Glasgow modified Eagle's medium (GMEM) in the presence or absence of 100 lqg/ml cycloheximide. After 1 h at 37 °C the virus inoculum was removed and replaced with fresh GMEM supplemented with 2.5% foetal bovine serum and, where appropriate, 100 lqg/ml cycloheximide. After a further 5 h at 37 °C the cells were rinsed twice in ice-cold phosphate-buffered saline and lysed by suspension and vortexing in a 4 M-guanidinium isothiocyanate, 20 mM-sodium acetate pH 5.2, 0.1 M-dithiothreitol and 0.5% Sarkosyl. The cell lysate was passed through a 20-gauge syringe needle to reduce viscosity and then layered onto an equal volume of 5.7 M-CsCl and centrifuged at 25000 r.p.m. overnight at 18 °C in a Beckman SW41 rotor. The RNA pellet was dissolved in 10 mM-Tris–HCl pH 7.4, 5 mM-EDTA, 1% SDS, recovered by ethanol precipitation and stored at −20 °C.

Northern analyses. Ten lqg samples of total infected cell RNA were denatured in 22 M-formaldehyde, 50% deionized formamide and 0.5 x MOPS buffer (20 mM-MOPS pH 7.0, 0.5 mM-EDTA, 5 mM-sodium acetate) at 55 °C for 15 min and then electrophoresed through a 1% agarose gel in MOPS buffer containing 22 M-
formaldehyde. RNAs were transferred and fixed to nitrocellulose membranes by overnight blotting followed by baking at 80 °C for 2 h. Membranes were then prehybridized in 50% deionized formamide, 5 × Denhardt's solution, 5 × SSC, 0-1% SDS, 50 mM-sodium phosphate pH 6-5 and 200 μg/ml boiled calf thymus DNA for 4 h at 42 °C. Hybridization with single-stranded DNA probes was performed overnight at 42 °C in 50% deionized formamide, 5 × SSC, 20 mM-sodium phosphate pH 6-5, 1 × Denhardt's solution, 100 μg/ml boiled calf thymus DNA and 5% dextran sulphate. Filters were washed twice in 2 × SSC, 0-1% SDS and twice in 0-1 × SSC, 0-1% SDS at room temperature for 15 min before drying and exposing against X-ray film.

Preparation of prime-cut probes. M13 clones were chosen that contained vaccinia virus DNA from entirely within the serpin genes and of coding strand polarity. Single-stranded DNA from these clones was hybridized with universal sequencing primer in 20 mM-Tris–HCl pH 8-5, 10 mM-MgCl2 and an elongation reaction with Klenow DNA polymerase was carried out in the presence of 32P-labelled dCTP and unlabelled 0-2 mM-dGTP, -dTTP and -dATP. After 15 min unlabelled 0-2 mM-dCTP was added and the incubation was continued for a further 10 min. The product was digested with restriction enzymes SalI and EcoRI that cut on either side of the vaccinia virus sequences within the M13 polylinker, and the digests were resolved by electrophoresis in a 6% polyacrylamide gel. After brief autoradiography the labelled single-stranded DNA fragment was excised from the gel and eluted by incubating the crushed acrylamide at 65 °C in 500 mM-ammonium acetate, 10 mM-magnesium acetate, 1 mM-EDTA, 0-1% SDS and 10 μg/ml tRNA for 2 h. The radioactivity in the supernatant was used directly as the single-stranded probe.

RESULTS

Fig. 1 shows a HindIII restriction map of the vaccinia virus genome and the positions of SalI fragments near the right ITR that have been sequenced in this study. The complete sequences of the SalI 1 and G fragments will be presented elsewhere. The shotgun method of DNA sequencing resulted in the sequence of the I and G fragments being determined to a character density of 5-7 and 7-3, respectively. Because the coding region of one serpin gene crosses the SalI site separating the SalI G and I restriction fragments, it was necessary to sequence a clone that spanned this site. This was done by cloning a 7-9 kb EcoRI fragment directly from purified virus DNA into pUC13 and then subcloning a BamHI–PstI fragment into PstI- and BamHI-cut M13mp18 (Fig. 1). The PstI site is only 54 nucleotides from the SalI site permitting the sequence at the SalI junction to be easily determined. This analysis demonstrated that there are no extra sequences between the ends of the SalI G and I restriction fragments and, therefore, that the sequence of the serpin gene crossing this junction obtained by joining the data from the two restriction fragments was correct. The positions and direction of transcription of these genes and the third vaccinia virus serpin gene are indicated in Fig. 1. The serpin genes sequences in this study are 8 kb apart.

The nucleotide sequences of the two serpin genes and the deduced amino acid sequences are shown in Fig. 2 and 3. The nomenclature used for these genes, B13R and B24R, reflects their positions within the HindIII B restriction fragment and the direction of transcription (Earl & Moss, 1987). ORFs initiating within the HindIII B fragment are numbered sequentially from left to right and the letter R indicates transcription in the rightward direction. The serpin gene within the HindIII K fragment (Boursnell et al., 1988) is referred to as K2L since it is the second ORF initiating in the HindIII K fragment and is transcribed leftwards.

The proteins encoded by genes B13R and B24R have 345 and 353 amino acids, respectively. The predicted Mr values are 38537 and 40467. The amino acid sequence of B24R is identical to that reported by Kotwal & Moss (1989) although a slight difference in Mr is computed. In contrast, our nucleotide sequence for B13R contains 16 base alterations and an additional nine internal nucleotides. After translation this results in 11 amino acid substitutions, three extra amino acids and an additional possible site for N-linked glycosylation (Fig. 5). There are also three more nucleotide differences outside the coding region of B13R. These surprising differences might have been explained by sequencing errors of either group, or by sequencing different strains of vaccinia virus. The sequence of B13R was determined to an average character density of eight using the shotgun sequencing method and we are confident it is correct. However, the sequence of B13R to the left of the SalI site at amino acid 278 was derived from DNA cloned from a rifampicin-resistant mutant of vaccinia virus (Baldick & Moss, 1987), while the published sequence is derived from wild-type (wt) virus. The genetic lesion conferring
Fig. 1. Vaccinia virus genome. The positions of HindIII restriction fragments are marked above the line. Open boxes represent the 10 kb ITRs. The stippled box represents the SalI restriction fragments and arrows indicate the position and direction of serpin genes. ○, SalI; □, BamHI; △, PstI.

Fig. 2. Nucleotide sequence and deduced amino acid sequence of the vaccinia virus B13R gene. Positions of potential early transcriptional termination signals are underlined. Nucleotides are numbered starting 120 upstream of the methionine codon initiating translation.

Fig. 3. Nucleotide sequence and deduced amino acid sequence of the vaccinia virus B24R gene. Positions of potential early transcriptional termination signals are underlined. Nucleotides are numbered starting 120 upstream of the methionine codon initiating translation.
rifampicin resistance maps 60 kb away from B13R and consists of only a single nucleotide alteration in a 485 bp restriction fragment (Balick & Moss, 1987). In view of this it seemed most improbable that the 11 amino acid alterations and the presence of an extra three amino acids in the serpin gene were attributable to virus strain differences. Nonetheless to exclude this possibility we sequenced the corresponding DNA from wt virus obtained from B. Moss. A 1-4 kb SphI–SalI fragment from the rightmost part of SalI G was purified and rendered blunt-ended by incubation with T4 DNA polymerase and Klenow enzymes. The DNA was digested with HinclI and AluI and the fragments were cloned into the HinclI site of M13mp19 and sequenced. The data obtained were identical to our original sequence derived from the rifampicin-resistant mutant.

The hydrophobicity profiles of these genes and the third serpin gene, K2L (Boursnell et al., 1988), are shown in Fig. 4. It is evident that all three genes contain a relatively hydrophobic sequence near the amino terminus. However, careful analysis of these regions indicates that only the K2L hydrophobic sequence is likely to function as a signal sequence according to the parameters defined by McGeoch (1985). This must be determined experimentally. The K2L protein also contains a hydrophobic sequence near the carboxy terminus that might serve as a membrane anchor domain. B13R and B24R contain four and five potential N-linked glycosylation sites, respectively.

Comparisons of the amino acid sequences of genes B13R and B24R with the protein database SWISSPROT using the FASTP program established that these proteins belong to the family of serpins. Both genes have between 25 and 35% amino acid identity with several members of this superfamily (alpha-1-antitrypsin, antithrombin III, plasminogen activator inhibitor). In addition, alignment of the vaccinia virus genes with members of this family (data not shown) indicate that at most positions at which there is high amino acid conservation among the serpins, the vaccinia virus proteins contain the same amino acid. Fig. 5 shows the alignment of six serpin genes taken from poxviruses of three different genera. These come from myxoma virus (a leporipoxvirus) (Upton et al., 1986), fowlpox virus (an avipoxvirus) (Tomley et al., 1988), cowpox virus (an orthopoxvirus) (Pickup et al., 1986) and vaccinia virus (an orthopoxvirus) (Boursnell et al., 1988; Kotwal & Moss, 1988a, 1989; and this communication). There are 33
positions (indicated with an asterisk) at which there is perfect amino acid homology among all these sequences (although the sequence of the fowlpox virus gene is truncated). The most closely related proteins are the cowpox 38K protein and the vaccinia virus B13R protein which share 46% amino acid identity. The separate identity of the three vaccinia virus serpins is demonstrated by comparisons of the degree of homology with 92%, 46% and 19% amino acid identity. The next most closely matched pair are the two vaccinia virus genes the cowpox virus gene. The B13R, B24R and K2L gene products share 92%, 46% and 19% amino acid identity with the cowpox virus protein, respectively. This argues against these genes having evolved relatively recently by gene duplication, a contention supported by the different hydrophobicity profiles and the different sequences at the active sites.

It is possible to gain some insight into the specificity of serine proteinases that may be inhibited by a particular serpin by an analysis of the amino acid sequence at the serpin active site. A highly conserved glutamic acid residue at position 342 in human alpha-1-antitrypsin is usually 17 or 18 amino acids upstream of the active centre (Carrell et al., 1982) and is usually serine but occasionally threonine. The following residue (P1') is usually serine but occasionally threonine. All the vaccinia virus serpins contain serine at this position. The positions of these residues within the vaccinia virus serpins are indicated in Fig. 5. Thus, serpin K2L might inhibit a serine proteinase that cleaves after arginine, possibly with a preceding alanine; serpin B13R might inhibit an enzyme cleaving after alanine, perhaps preceded by cysteine; serpin B24R might inhibit a proteinase cleaving after phenylalanine, possibly preceded by asparagine. The active site of B24R is also a potential N-linked glycosylation site. These are only rough guides and there is no direct evidence that the gene products described here do actually function as serpins. However, since B13R has 92% amino acid identity and a perfectly matched active site with the cowpox virus 38K serpin, a protein that is known to interfere with blood coagulation (Pickup et al., 1986), it is probable that this vaccinia virus protein does function as a serpin.
Inspection of the nucleotide sequences upstream of the serpin ORFs showed the absence of the conserved motif TAAAT(G) that is present upstream of late vaccinia virus genes (Rosel et al., 1986). The motif TTTTTNT, which functions to terminate early vaccinia virus transcription (Yuen & Moss, 1987), is present downstream of all these serpin genes and not within the coding sequences in a position that might cause truncation of mRNAs for these genes. Therefore, it is likely that these genes are transcribed early, before DNA replication. To test this, RNA was prepared from vaccinia virus-infected cells early (6 h post-infection in the presence of cycloheximide) or late (8 h post-infection) during the infectious cycle. These RNAs were used in Northern analyses as described in Fig. 6. Single-stranded radioactive probes that map entirely within the serpin gene-coding regions and which are of negative polarity (complementary to mRNA), were used. For serpin K2L, a major early mRNA of approximately 1500 nucleotides and a minor early mRNA of 2100 nucleotides are found (lane 4). The sizes of these early mRNAs correspond to the size of transcripts predicted to be formed by initiation just upstream of the serpin-coding region and termination 50 nucleotides downstream of TTTTTNT motifs present 1200 and 2200 nucleotides downstream. Allowances are made for the addition of poly(A) and the slightly different mobility of the double-stranded DNA M, markers. There are late mRNAs of heterogeneous length running through this region (lane 5). These may initiate from upstream genes and are not necessarily originating from the serpin promoter. Primer extension and S1 nuclease analyses are required to establish their origin.
For the B13R serpin gene, a major early mRNA of 1400 nucleotides and a minor early mRNA of 2300 nucleotides are found (lane 6). These are somewhat larger than the sizes of the predicted transcripts initiating at the promoter region and terminating downstream of TTTTTNT motifs that are present 1120, 1700 and 1750 nucleotides downstream. However, the exact site of transcriptional initiation and the length of poly(A) tail are unknown. There is very little late transcription through this gene in the rightward direction (lane 7).

Serpin B24R contains a major early mRNA of 1300 nucleotides and minor early species of 2300 and 3900 nucleotides (lane 1). There is a TTTTTNT motif 1080 nucleotides downstream of the promoter region that could give rise to the observed band of 1300 nucleotides. The sequence of the SalI I fragment we have determined extends a further 565 nucleotides beyond this TTTTTNT motif and there are no more TTTTTNT sites in this region. However, the sequence presented by Kotwal & Moss (1989) shows additional TTTTTNT sites a further 790 and 815 nucleotides downstream that could give rise to a larger mRNA of 2300 nucleotides. The origin of the 3900 nucleotide RNA is unclear but might represent RNA running into the serpin gene from upstream, or RNA terminating at a TTTTTNT site further downstream. There is very little late mRNA running through this gene.

In summary, all three serpin genes are transcribed early producing mostly RNAs of predicted sizes. There is evidence of late transcription through the K2L serpin but not the other genes. Late mRNAs running through the K2L gene are not necessarily translated into a K2L gene product since they may initiate far upstream at other late gene(s).

**DISCUSSION**

Vaccinia virus has a family of genes that code for proteins with strong homology to members of the serpin family. The nucleotide sequences of two of the serpin genes are presented here together with the alignment of all known poxvirus serpins, the possible specificity of the target serine proteinases and the transcriptional mapping of the three vaccinia virus serpin genes.

Serpins have now been found in four different poxviruses of three different genera. The serpin from fowlpox virus (Tomley et al., 1988) is truncated having lost a large region at the amino terminus and is probably non-functional. This truncation has probably resulted from genomic recombination since the position of truncation corresponds to the junction of the left ITR and unique sequences. It is possible that fowlpox virus contains a complete serpin at the other end of the genome or elsewhere. The presence of serpins in divergent poxviruses suggests that they play some important role in the virus life cycle. Nonetheless, several of them are non-essential for virus replication in vitro. Mutants of cowpox virus lacking the 38K serpin are able to grow in tissue culture and on the chorioallantoic membrane (CAM) of avian embryos (Pickup et al., 1986). Viable deletion mutants of vaccinia virus have been described that have lost the K2L gene (Perkus et al., 1986) and we have constructed a vaccinia virus deletion mutant lacking the B13R vaccinia virus serpin gene (G. L. Smith, unpublished data). The only phenotypic change reported for serpin-deleted viruses is from cowpox virus, where the deletion mutant forms white pocks rather than the usual haemorrhagic pocks, on the avian CAM. This was of interest because members of the serpin family perform important regulatory functions in biological cascades such as blood coagulation and complement activation. The cowpox virus serpin is a major, early, intracellular virus protein, and so probably interferes with blood coagulation after breakage of virus-infected cells. Prevention of blood coagulation might confer an advantage to the virus in vivo since the site of infection would not be ‘walled off’ by fibrin and so virus dissemination would be enhanced.

The subcellular locations of the vaccinia virus serpins remain to be determined experimentally. However, the presence of a potential signal sequence at the amino terminus and a hydrophobic sequence near the carboxy terminus of serpin K2L suggest a secretory or membrane-associated protein. This serpin is likely to inhibit enzymes that cleave after arginine residues such as thrombin, urokinase, complement component C1 and endothelial plasminogen activator (for review see Carrell et al., 1987). Potentially it could interfere with the processes of blood coagulation and the immune response to virus infection. The serpin (B13R) with 92% amino acid identity with the cowpox virus protein is likely to interfere with blood coagulation
Vaccinia virus serpins


We express our thanks to David Brown and Adrian Terry for help with computing, to Mike Boursnell and Matthew Binnie for computer program ORFFILE, for M13 clones within the K2L serpin gene and for making their sequence data available to us before publication, to Mike Merchlinsky and Bernard Moss for plasmids, to Mary Wright for typing the manuscript, and to Robin Carrell for helpful advice about alignment of serpin sequences. This work was supported by the Medical Research Council project grant G8703796CA. S.T.H. holds a Commonwealth Scholarship Commission Research Studentship and G.L.S. is a Lister Institute-Jenner Research Fellow.

REFERENCES


Vaccinia virus serpins


(Received 30 March 1989)