Two Early Vaccinia Virus Genes Encode Polypeptides Related to Protein Kinases

By SUSAN T. HOWARD AND GEOFFREY L. SMITH*†
Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, U.K.

(Accepted 25 July 1989)

SUMMARY

Vaccinia virus particles contain a protein kinase with an \( M_r \) of 62K calculated from sedimentation rate. We have sequenced the SalI G restriction fragment of the vaccinia virus genome near to the right inverted terminal repeat and have identified two genes which share 36% amino acid identity with each other and are related to the family of protein kinase genes. One gene, designated B1R, encodes a 34-2K protein which shares 27% identity with a protein kinase encoded by the herpes simplex virus type 1 US3 gene and contains conserved motifs characteristic of protein kinases of serine/threonine specificity. The second gene, B12R, encodes a protein of 33-3K which is poorly related to known protein kinases and lacks specific amino acids at several highly conserved key positions. The deduced partial amino acid sequence of a gene in the corresponding region of the cowpox virus genome is identical to B12R except for one conservative amino acid substitution. Both of the vaccinia virus genes are transcribed towards the right-hand end of the genome early during infection. It is possible that the product of either or both of these genes associates to form a homo- or heterodimer that represents the 62K virion-associated protein kinase.

INTRODUCTION

Vaccinia virus is a large dsDNA virus that replicates in the cytoplasm of infected cells and is the prototype orthopoxvirus. The virus particle contains many enzymes necessary for virus transcription or replication (for review, see Moss, 1985) and several of these have been shown to be virus-encoded. The genes encoding the virion enzymes topoisomerase I (Shuman & Moss, 1987), mRNA guanyltransferase (Morgan et al., 1984; Niles et al., 1986), DNA-dependent ATPase (Rodriguez et al., 1986; Broyles & Moss, 1987) and several RNA polymerase subunits (Broyles & Moss, 1986; Jones et al., 1987; Patel & Pickup, 1989) have been mapped and sequenced. DNA polymerase (Traktman et al., 1983; Jones & Moss, 1984; Earl et al., 1986), ribonucleotide reductase (Slabaugh et al., 1988; Schmitt & Stunnenberg, 1988; Tengelsen et al., 1988) and thymidine kinase (Weir et al., 1982; Hruby & Ball, 1982; Weir & Moss, 1983) are additional virus-encoded enzymes expressed in infected cells and the genes have also been mapped and sequenced. An alkaline protease (Arzoglou et al., 1978), nicking-joining enzyme (Lakritz et al., 1985; Reddy & Bauer, 1989) and protein kinase (Paoletti & Moss, 1972; Kleiman & Moss, 1973, 1975a, b) have been identified within the virus particle but not proved to be virus-encoded.

Other viruses package one or more protein kinases into the virus particle. Examples are frog virus 3, an iridovirus (Silberstein & August, 1976), vesicular stomatitis virus, a rhabdovirus (Beckes et al., 1989), African swine fever virus (Polatnick et al., 1974) and the alphaherpes-
viruses herpes simplex virus (HSV) and pseudorabies virus (Stevely et al., 1985; for review, see Leader & Katan, 1988). The alphaherpesviruses encode a protein kinase gene within the unique short region of the genome (McGeoch & Davison, 1986; Purves et al., 1987) and related protein kinase genes are found in the genomes of other herpesvirus groups (Chee et al., 1989; Smith & Smith, 1989). Several oncogenic retroviruses encode tyrosine protein kinases which are responsible for the transforming potential of the virus, e.g. Rous sarcoma virus src (Cooper, 1982).

To understand better the biology of vaccinia virus we have sequenced a large region of the genome adjacent to the right inverted terminal repeat (ITR), a region for which little genetic data are available. Within the 12-6 kb SalI G fragment we have identified two open reading frames (ORFs) which have homology to each other and to members of the protein kinase family. One gene, designated B1R, encodes a 34.2K protein which has the conserved domains and invariant residues characteristic of protein kinases of serine/threonine specificity. A temperature-sensitive mutation which has a DNA-negative phenotype has been mapped to this gene (P. Traktman, personal communication). The second gene, designated B12R, encodes a protein of 33.3K which shares 36% identity with B1R but is poorly conserved in comparison to known protein kinases and differs at several highly conserved positions. Northern blot analyses demonstrate that both of these genes are transcribed early during infection.

**METHODS**

**DNA sequencing.** The vaccinia virus SalI G fragment was subcloned from a cosmid containing vaccinia virus genomic DNA from a rifampicin-resistant mutant of the WR strain (Baldick & Moss, 1987). Cosmid DNA was digested with SalI and the fragments were cloned into SalI-cut pUC13. Plasmids containing the SalI G fragment were identified by digestion with HindIII and SalI followed by electrophoresis of DNA fragments through agarose gels. Random subclones of SalI G were prepared and sequenced according to the procedures described by Bankier et al. (1987) with slight modifications: both HindII-cut mp19 and Smal-cut mp18 were used as cloning vectors for the production of single-stranded recombinant phage DNA, and polyacrylamide gels were both prepared with and run in a buffer gradient from 0.5 to 2.5 × TBE (1 × TBE is 89 mM-Tris-borate, 89 mM-boric acid and 2 mM-EDTA).

**Computer analyses.** Random sequence data were assembled into continuous sequence using the programs DBAUTO and DBUTIL (Staden, 1982). ORFs were identified using ANALYSEQ (Staden, 1986) and ORFFILE (Mike Boursnell, Institute of Animal Health, Houghton, U.K.) and examined for codon usage (ANALYSEQ) consistent with that found in previously sequenced vaccinia virus genes. Amino acid and nucleotide sequence comparisons were made using the FASTP program (Lipman & Pearson, 1985) and FASTA (Pearson & Lipman, 1988), respectively. Multiple alignment of protein sequences were obtained using the MULTALIGN program (Barton & Sternberg, 1987).

**RNA preparation.** Human TK-143 cells were infected with the vaccinia virus strain WR at 30 p.f.u. per cell in Glasgow modified Eagle’s medium (GMEM) in the presence or absence of 100 μg/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 μg/ml cycloheximide. After a further 5 h at 37 °C the cells were rinsed twice in ice-cold phosphate-buffered saline and lysed in 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 blot analysis.** RNA samples (10 μg) were denatured and electrophoresed on formaldehyde gels (1% agarose) according to the procedures of Maniatis et al. (1982). RNA gels were rinsed in water and then blotted overnight onto nitrocellulose membranes followed by baking for 2 h at 80 °C. Membranes were prehybridized overnight at 42 °C in a solution consisting of 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. Radiolabelled ssDNA probes were incubated with nitrocellulose membranes overnight at 42 °C in 50% deionized formamide, 1 × Denhardt’s solution, 5 × SSC, 0.1% SDS, 20 mM-sodium phosphate pH 6.5, 100 μg/ml boiled calf thymus DNA and 5% dextran sulphate. Membranes were washed three times in 2 × SSC, 0.1% SDS and three times in 0.1 × SSC, 0.1% SDS at room temperature for 10 to 15 min before drying and autoradiography.

**Preparation of prime-cut probes.** M13mp19 clones containing vaccinia virus DNA inserts, which mapped entirely within the coding regions of either B1R or B12R, were used as templates for the synthesis of single-stranded probes complementary to the coding strand. Universal sequencing primer was hybridized to the M13 template for 15 min...
at 60 °C in 20 mM-Tris-HCl pH 8.5, 10 mM-MgCl₂ and then extended at room temperature using the Klenow enzyme in the presence of 0.2 mM-dGTP, -dTTP and -dATP and 20 μCi of [³²P]dCTP. The reaction was allowed to proceed for 15 min and then 0.2 mM unlabelled dCTP was added. After a further 10 min, the reaction was stopped by heating for 5 min at 65 °C. Inserts containing vaccinia virus DNA were released by digestion with the restriction enzymes *EcoRI* and *HindIII*, resolved by electrophoresis on a 6% polyacrylamide gel, and located by brief autoradiography. The band was cut out of the gel and the DNA was eluted by incubating the crushed acrylamide overnight at 37 °C in 500 mM-ammonium acetate, 10 mM-magnesium acetate, 1 mM-EDTA, 0.1% SDS and 10 μg/ml tRNA.

**RESULTS**

Fig. 1 shows the position of the *SalI G* restriction fragment, near the right end of the vaccinia virus (WR) genome. The complete sequence of the *SalI G* restriction fragment was determined by sequencing randomly generated fragments (Bankier *et al.*, 1987) and assembling these data into contiguous sequence by computer. The sequence of *SalI G* is 12564 nucleotides (nt) long and was determined to an average character density of 7.3. Translation of this nucleotide sequence in six frames identified 14 complete ORFs, two of which are the subject of this paper. These genes are designated B1R and B12R to indicate that they are the first and twelfth ORFs initiating within the *HindIII* B fragment and are transcribed in a rightward direction. The codon usage of these genes is similar to a collection of previously published vaccinia virus genes (data not shown).

The nucleotide sequence and deduced amino acid sequence of B1R are shown in Fig. 2. The sequence was completely determined on both DNA strands. The methionine codon presumed to initiate the translation of the 300 amino acid protein is located 344 nt from the left end of the *HindIII* B fragment. Inspection of the nucleotide sequence downstream of the preceding ORF and before translational initiation of B1R showed that the octanucleotide TAAAACTT was repeated three times within a region of 95 nt, twice in 23 nt. This sequence occurs at only one other position within the entire 12.6 kb *SalI G* fragment and this was on the opposite strand, complementary to the coding region of another gene. Searches for this sequence within the flanking 9-8 kb *SalI I* and 13-4 kb *SalI F* fragments (G. L. Smith, unpublished data) and within the *HindIII* D, K, I, H, L, J and C fragments (Niles *et al.*, 1986; Boursnell *et al.*, 1988; Schmitt & Stunnenberg, 1988; Rosel *et al.*, 1986; Broyles & Moss, 1986; Plucienniczak *et al.*, 1985; Weinrich & Hruby, 1986; Kotwal & Moss, 1988) revealed no other clusters although it occurs on average once every 10000 nt. The repetition of this sequence within the probable promoter region of B1R is unlikely to be simply by chance, in view of its rarity elsewhere in the vaccinia virus genome. However its significance is unknown.

Fig. 3 shows the nucleotide sequence and deduced amino acid sequence of B12R. This ORF lies between nucleotides 8995 and 9844 from the left end of the *HindIII* B fragment. The sequence in this region was determined on both DNA strands except for 6 nt which were read from five separate clones of the same polarity. The predicted polypeptide is 283 amino acids long with an *Mᵣ* of 33.3K.
Fig. 2. Nucleotide sequence and deduced amino acid sequence of gene B1R. Potential transcriptional control signals are underlined. The repeated octanucleotide sequence TAAAAACTT is marked by asterisks.
Fig. 3. Nucleotide sequence and deduced amino acid sequence of gene B12R. Potential early transcriptional termination signals are underlined. The first 18 nucleotides and six amino acids of gene B13R, a serpin gene (Kotwal & Moss, 1989; Smith et al., 1989), are included.
Fig. 4. Northern blot analysis of mRNAs for genes B1R and B12R. RNA was purified from cells infected with vaccinia virus early (6 h post-infection in the presence of cycloheximide) or late (8 h post-infection) or from mock-infected cells. 10 μg samples of RNA were denatured and resolved through agarose gels containing formaldehyde, before being transferred to nitrocellulose by blotting. Filters were probed with 32P-labelled ssDNA complementary to the coding strand of gene B1R (lanes 1 to 3) or gene B12R (lanes 4 to 6). After washing, autoradiographs were prepared. Lanes 1 and 4, early RNA; lanes 2 and 5, late RNA; lanes 3 and 6, mock-infection. The positions of dsDNA Mr markers (kb) are indicated by arrows.

RNA analysis

Examination of the nucleotide sequence of ORFs B1R and B12R for potential transcriptional control signals showed that both genes contain the motif TTTTTNT within 35 nt of their 3' end and not within their coding region. This signal is also found 110 and 52 nt upstream of the initiating methionine codon of ORF B1R and ORF B12R, respectively. This sequence causes the termination of transcription approximately 50 nt downstream early during infection (Yuen & Moss, 1987). These data suggest that both genes might be transcribed early during infection. The sequence TAAAT, which is found at, or very close to the 5' end of late genes (Rosel et al., 1986), is present 70 nt upstream of ORF B1R. If late RNA transcription initiated from a region close to this sequence, it would contain an AUG codon in a different reading frame before the AUG codon initiating the B1R ORF. However, since the first AUG is followed four codons later by a termination codon, efficient translation from the AUG initiating the B1R ORF could still occur. Therefore, gene B1R might be transcribed and/or expressed late during infection as well.

In order to analyse the temporal expression of both genes, RNA was isolated from virus-infected cells early or late during infection, and examined by Northern blotting. Radiolabelled single-stranded probes were prepared from M13 clones previously sequenced and shown to contain vaccinia virus DNA that maps entirely within the coding region of B1R or B12R. The
probes were complementary to the sequence shown in Fig. 2 and 3. Fig. 4 shows the result of this analysis. For B1R (lanes 1 to 3) transcripts of 1200 and 3200 nt are found early during infection (lane 1). The sizes of these RNA molecules correspond to transcripts initiating just upstream of the ORF and terminating shortly after TTTTTNT motifs present 930 and 2650 nt downstream of the translational initiation codon [with allowances for a poly(A) tail and the uncertain site of transcriptional initiation]. RNA of heterogeneous length is found late during infection (lane 2). S1 nuclease or primer extension analyses are required to determine whether these originate from the TAAAT motif 70 nt upstream of the ATG codon, or from late genes further upstream. The haemagglutinin gene is less than 2 kb upstream of B1R (Shida, 1986; S. T. Howard, unpublished data) and is transcribed rightwards late during infection. RNA from uninfected cells (lane 3) does not hybridize to the probe.

For B12R, a major early transcript of 1100 nt and minor transcripts of 1400 and 2300 nt are detected (lane 4). The 1100 nt RNA would correspond to a transcript initiating just upstream of the ORF and terminating 50 nt downstream of the sequence TTTTTTTTAT which is present 870 nt downstream of the initiating methionine codon. This sequence, TTTTTTTTAT, represents three overlapping termination signals. There is no obvious signal to account for the 1400 nt minor transcript, but the 2300 nt RNA is of a size consistent with termination at the next TTTTTNT motif downstream. This is located within the SalI fragment after an early gene predicted to encode a 38.5K serine proteinase inhibitor (Smith et al., 1989). There is a low level of late transcription through B12R but this is probably readthrough from late genes upstream in view of the absence of an upstream TAAAT motif.

Homology of the B1R gene product to protein kinases

The predicted translation product of B1R was compared with the protein database SWISSPROT using program FASTP (Lipman & Pearson, 1985). B1R has the strongest similarity to members of the protein kinase family. The closest matches are to the HSV US3 gene product (McGeoch et al., 1985), myosin light chain protein kinase (Takio et al., 1986) and yeast cell division control protein 7 (Patterson et al., 1986) with optimized FASTP scores (using a KTUP of 1) of 112, 107 and 107, respectively. B1R has 27% amino acid identity with HSV US3 over an 85 amino acid region. For comparison the HSV US3 protein kinase has 28% identity over a 149 amino acid region with one of its most closely related protein kinases. These homology scores are low but reflect the overall diversity of protein kinases. Nonetheless within the catalytic domain of the protein kinase family there are distinct region and amino acid motifs, interspersed among more variable sequences, which are highly conserved and indicate whether it is likely that the polypeptide functions as a protein kinase (Hanks et al., 1988; Smith & Smith, 1989). Fig. 5 shows an alignment of B1R with several protein kinases, including those with tyrosine (lines 1 and 2) rather than serine/threonine specificity (lines 3 to 6), which was initially made using the MULTALIGN program (Barton & Sternberg, 1987) and then modified where necessary to align conserved domains. Reference is made to the six major conserved domains described by Smith & Smith (1989) and to the alignments of Hanks et al. (1988).

Region I containing the consensus motif G-X-G-X-X-G is involved in nucleotide binding and is found in many nucleotide-binding proteins as well as in protein kinases (Wierenga & Hol, 1983; Gorbalenya et al., 1988; Hodgman, 1988). It is proposed that the conserved glycine residues cause a fold in the polypeptide chain allowing contact with the ribose and phosphate moieties of the nucleotide. In B1R all three glycine residues are conserved. There are also conserved hydrophobic residues: one upstream and two downstream of the first and third glycine residues, respectively.

Region II contains an invariant lysine which is critical for kinase activity since substitutions of other residues at this position obviate activity (Chen et al., 1987). B1R has two possible sites for this region, the first ends with Lys 41 and the second with Lys 45. Lys 41 is preceded by three hydrophobic residues, whereas there is a glutamate residue two positions upstream of Lys 45. All residues upstream of the lysine in region II of other protein kinases are characteristically uncharged and predominantly strongly hydrophobic. In view of this, Lys 41 is probably the invariant, catalytic lysine. If this is correct, the distance between the last glycine of region I and
Fig. 5. Alignment of amino acid sequences of the catalytic domains of several protein kinases with the deduced amino acid sequence of gene B1R. The numbered boxes are conserved domains referred to in the text. Numbers in parentheses indicate the number of amino acids which are present amino- or carboxy-terminal to the sequences shown for each protein kinase. Numbers underneath the amino acid sequences refer to the number of amino acids from the amino terminus of B1R. Residues conserved in all kinases shown are indicated with an asterisk. Underlined residues indicate serine/threonine or tyrosine kinase specificity. Protein kinase genes are: LSK-T, murine lymphocyte tyrosine protein kinase (Marth et al., 1985); src, Rous sarcoma virus (Schwartz et al., 1982); MLCK, rabbit skeletal muscle myosin light chain kinase (Takio et al., 1986); CK-II, casein kinase II from Drosophila melanogaster (Saxena et al., 1987); C-PIM-1, putative protooncogene (Selten et al., 1986); HSV-US3, herpes simplex virus type 1 US3 gene protein kinase (McGeoch et al., 1985).
the catalytic lysine is 13 residues in B1R, one less than the usual 14 to 23 residues. However this spacing has not been shown to be critical (Hanks et al., 1988) and the putative phosphotransferases encoded by members of alpha-, beta- and gammaherpesviruses also have less than 14 residues between these motifs (Chee et al., 1989; Smith & Smith, 1989).

The next two regions, III and IV, are believed to be involved in nucleotide binding and phosphate transfer and are the regions with the greatest frequency of highly conserved residues. The histidine, aspartate and asparagine residues in region III and the aspartate and glycine residues in region IV are conserved in all sequences shown in Fig. 5. With the exception of histidine these residues are also invariant in all known protein kinases. In addition, both aspartates and the asparagine are conserved in bacterial phosphotransferases which confer antibiotic resistance (Brenner, 1987). These aspartate residues may bind the phosphate groups of ATP by Mg$^2+$ salt bridges. Within these conserved regions the other residues of B1R conform to defined sequence patterns of protein kinases (Hanks et al., 1988; Smith & Smith, 1989). For instance, there are two hydrophobic residues before the conserved aspartate and after the conserved glycine of region IV.

In regions V and VI, particularly the former, the sequence of B1R diverges from the general pattern. Region V is considered as a key catalytic domain indicator and has the consensus [A,G,S,P] [P,A,I,L] E, where letters in brackets signify acceptable alternatives (Chee et al., 1989; Smith & Smith, 1989). B1R conforms to this pattern at the first two positions, but at the third position the sequence contains aspartate rather than the invariant glutamate of other protein kinases. Within region VI the most notable difference between B1R and other sequences is the presence of a charged residue (glutamate) at the third position instead of a residue with an aromatic R group (phenylalanine, tryptophan or tyrosine), most commonly tryptophan. However the two most highly conserved residues of region VI, aspartate at position 1 and glycine at position 6, are present in B1R. Carboxy-terminal to region VI, other protein kinases contain an invariant arginine which occurs 11 residues after a strongly hydrophobic residue. B1R contains two potential arginine residues for this position but both contain a threonine 11 residues upstream.

In summary, all the conserved regions which characterize the catalytic domain of the protein kinase family are conserved in B1R, and with one exception each invariant residue is also maintained. Biochemical characterization of the B1R gene product is required to determine whether it has protein kinase activity and whether the differences between B1R and other protein kinases are reflected in different enzymic properties. Based upon the overall similarity B1R should be classified as a putative protein kinase. The probable substrate specificity of B1R, serine/threonine or tyrosine, is considered under Discussion.

Homology between B12R and B1R

A search of the SWISSPROT database using FASTP with the deduced amino acid sequence of B12R showed it had weak similarity to several protein kinases. In addition, a search of our own database of vaccinia virus amino acid sequences, including those derived from the nucleotide sequence of the SalI G fragment, showed a surprisingly high homology to B1R. At the amino acid level the two vaccinia virus genes have 36% identity, which increases to 48% when conservative changes are included. There is also 51% nucleotide sequence identity. Fig. 6 shows the alignment of B1R and B12R with identical residues and conservative changes indicated. The underlined sequences of B1R represent the conserved regions boxed in Fig. 5 and are discussed above. The alignment demonstrates that B12R differs from protein kinases at several important positions and for this reason the sequence is omitted from Fig. 5. The most notable difference is the absence of a recognizable ATP-binding site corresponding to conserved region I. Although there are glycine residues in the same vicinity as those of B1R and others are found 50 to 60 residues downstream, these do not show a definite match to the G-X-G-X-X-G consensus pattern of nucleotide-binding proteins. Other important differences between B12R and protein kinases include the lack of the invariant aspartate and glycine residues in regions III and IV respectively, and considerable divergence from the normal pattern in region V. The sequence of region VI exhibits the best conservation, with 20 out of 24 residues identical
Fig. 6. Alignment of amino acid sequences of B1R and B12R and the partial amino acid sequence of a CPV homologue of B12R deduced from the nucleotide sequence data of Pickup et al. (1986). Pairs of dots indicate amino acid identity, single dots indicate conservative changes as defined by Patel & Pickup (1989).
between B1R and B12R. Overall the divergence of B12R from protein kinases at several key
conserved domains suggests that it is unlikely to have protein kinase activity. The similarity to
B1R suggests it may have arisen by a gene duplication event and has subsequently diverged in
sequence and function.

**Cowpox virus (CPV) contains a homologue of B12R**

The next ORF downstream of B12R initiates translation approximately 100 nt after the
termination of the B12R ORF. This gene, designated B13R, encodes a 38-45K protein with
homology to the family of serine proteinase inhibitors (serpins) (Kotwal & Moss, 1989; Smith
et al., 1989) and shows 92% amino acid identity to a 38K protein of CPV (Pickup et al., 1986). The
high homology between CPV and vaccinia virus serpin genes (95% at the nucleotide level)
prompted us to search the sequence upstream of the CPV serpin gene for a homologue of B12R.
Translation of the 300 nt upstream of the CPV serpin (Pickup et al., 1986) revealed an
incomplete ORF encoding 66 amino acids. This shows perfect identity with the carboxy
terminus of B12R except for one conservative change, lysine to arginine at position 278 of B12R
(Fig. 6). Within these coding regions there are only four nucleotide differences between the CPV
and vaccinia virus sequences, resulting in 98% identity. However the intergenic region between
B12R and B13R (92 nucleotides) shows lower conservation with the CPV sequence, with only
77% identity.

**DISCUSSION**

Within the *SalI* G fragment of vaccinia virus there are two related genes, B1R and B12R,
which share 36% amino acid and 51% nucleotide sequence identity. They provide another
example of a family of vaccinia virus genes. Other examples are the three serpin genes (Boursnell
et al., 1988; Kotwal & Moss, 1989; Smith et al., 1989), a gene from the *HindIII* K region with
homology to the gene in *HindIII* F that encodes a major 37K envelope antigen (Boursnell et al.,
1988; Hirt et al., 1986) and a pair of related genes from the ITR and *HindIII* C region which
encode, respectively, a 42K and a 37K protein which share 60% amino acid identity
(Venkatesan et al., 1982; Kotwal & Moss, 1988). We have also identified a gene within the *SalI*
fragment with homology to *HindIII* K ORF 7 (Boursnell et al., 1988) and to Shope fibroma virus
ORF T3A (Upton et al., 1987; G. L. Smith, unpublished data). Gene families have also been
found in other poxvirus genera. Shope fibroma virus, a leporipoxvirus, contains three similar
genes within each ITR (Upton et al., 1987), and fowlpox virus, an avipoxvirus, contains a family
of genes near the left end of the genome whose products are related to hepatic lectins (Tomley et
al., 1988). Other gene families have been found in African swine fever virus (Blasco et al., 1988)
and in human cytomegalovirus (Weston & Barrell, 1986). These large DNA viruses appear to
have increased their genome complexity by gene duplication events and subsequent divergence
of duplicated members.

Genes B1R and B12R both have homology to the catalytic domain of protein kinases,
although the degree of relatedness of B12R is low. B1R has sequences corresponding to all of the
major conserved regions except for region V (Fig. 5) where instead of a glutamate residue which
is found in all other protein kinases, there is an aspartate residue. In view of this difference we
initially considered other residues as candidates for region V. However the proline-isoleucine-
glutamate triplet we have chosen is probably correct for the following reasons. First, although
mutagenesis studies have shown that the substitution of a positively charged lysine for the
conserved glutamate reduces enzyme activity of src (Bryant & Parsons, 1984), it is possible that
the conservative change to aspartate is an acceptable alteration. Second, the two possible
alternatives for region V, asparagine-glycine-glutamate and threonine-leucine-glutamate
located downstream of region IV, contain the conserved glutamate but do not completely fit the
consensus for the first two amino acids. The asparagine-glycine-glutamate triplet is also only
eight residues from region IV, much closer than the usual minimum of 19 residues. Thirdly, the
threonine residue which starts the third possible match to region V is very likely to be part of the
sequence indicating the substrate specificity of the enzyme.
Protein kinases phosphorylate either serine/threonine or tyrosine residues. Each class of enzyme contains distinct characteristics within the catalytic domain which indicate probable specificity, in addition to the common conserved domains described above (Hanks et al., 1988; Smith & Smith, 1989). These characteristic residues lie between regions IV and V and within region III (Fig. 5). At position 7 of region III all serine/threonine protein kinases contain an invariant lysine whereas tyrosine protein kinases contain alanine or arginine residues. B1R contains lysine and therefore is probably a serine/threonine protein kinase. This is supported by comparison of the six residues preceding region V. Threonine/serine-type kinases have glycine-serine/threonine at the first two positions, whereas tyrosine kinases contain proline-isoleucine/valine-lysine/arginine-tryptophan-threonine/methionine in positions 2 to 6. B1R contains glycine-threonine and fits the threonine/serine protein kinase subfamily.

Vaccinia virus cores contain a serine/threonine protein kinase (Paoletti & Moss, 1972). The enzyme has an $M_r$ of 62K as determined from its sedimentation rate through sucrose gradients (Kleiman & Moss, 1975a) but it is not known whether this enzyme is of viral or cellular origin. Since the $M_r$ was determined by a non-denaturing method, it is possible that the 62K enzyme is a dimeric or multimeric protein. In most single subunit protein kinases there is an amino-terminal regulatory domain and a carboxy-terminal catalytic domain, whereas oligomeric protein kinases commonly have subunit polypeptides consisting of predominantly catalytic domains with an additional regulatory subunit (Hanks et al., 1988). The 34K B1R gene product is composed almost entirely of a protein kinase catalytic domain. Consequently it is possible that the 62K vaccinia virus enzyme is either a homodimer containing only B1R, or a heterodimer containing B1R with another subunit, perhaps B12R. It is unlikely that B12R as a monomer or homodimer could function as a protein kinase given its divergence from protein kinase sequences at several important catalytic regions. The vaccinia virus protein kinase is tightly associated with the virus core and is not regulated by cyclic mononucleotides (Paoletti & Moss, 1972; Kleiman & Moss, 1973). However, the enzyme is stimulated by basic proteins such as protamine, which acts only poorly as a substrate, but catalyses the phosphorylation of specific heat-resistant virus polypeptides of 37K and 11-7K (Kleiman & Moss, 1975a). These proteins are phosphorylated at serine and threonine residues. Serine phosphorylation of an 11K virion core polypeptide also occurs in infected cells (Kao & Bauer, 1987) but it is uncertain whether either the enzyme or the substrate are the same as those characterized from virions. If the modulation of the enzyme by protamine in vitro reflects regulation in vivo, the regulatory element is likely to reside on a subunit other than B1R. It is intriguing to speculate that B12R could be the regulatory subunit and in this regard it is noteworthy that the carboxy terminus of B12R is rich in basic amino acids. Myosin light chain protein kinase also contains a strongly basic carboxy terminus (Takio et al., 1986) which is required for regulation of the enzyme by calmodulin but not for enzymic activity (Edelman et al., 1985). The transcription of both vaccinia virus genes during the same early phase of infection is consistent with a model proposing an interaction of their gene products. Antisera are being raised to both B1R and B12R gene products to determine whether either or both of these polypeptides constitute the viral protein kinase.

It is not known whether B12R is an essential gene. However deletion mutants of CPV have been isolated which have lost 6 to 7 kb of DNA to the left of the 38K serpin gene (equivalent to vaccinia virus B13R) (Pickup et al., 1986). Unless the CPV homologue of B12R is transposed to another region of the genome in these mutants, or the genome contains a second functional copy, it is considered non-essential. This strongly suggests that the gene B12R in the closely related orthopoxvirus, vaccinia virus, is non-essential. Specific deletion of the B12R gene from vaccinia virus would be a method of demonstrating unequivocally its non-essential status. A vaccinia virus temperature-sensitive mutation with a DNA-negative phenotype has been mapped to the B1R gene (P. Traktman, personal communication). This demonstrates that the B1R protein kinase plays an essential role in vaccinia virus replication. This may be during virus uncoating, or some other early critical event, or during DNA replication itself.
REFERENCES


We thank Mark Chee for providing us with his paper prior to publication, Shona Kerr and Paco Rodriguez for critical reading of the manuscript and Mary Wright and Anita Hancock for typing. This work was supported by MRC grant G8703796CA. S.T.H. is a Commonwealth Scholarship Commission Research Student and G.L.S. is a Lister Institute-Jenner Research Fellow.
Vaccinia virus protein kinase genes


(Received 8 May 1989)