JOURNAL OF VIROLOGY, July 1991, p. 3583–3589
0022-538X/91/073583-07$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 65, No. 7

Characterization of a Molluscum Contagiosum Virus Homolog of the Vaccinia Virus p37K Major Envelope Antigen

NEIL W. BLAKE, COLIN D. PORTER,† AND LEONARD C. ARCHARD*

Department of Biochemistry, Charing Cross and Westminster Medical School, St. Dunstans Road, London W6 8RP, United Kingdom

Received 30 November 1990/Accepted 12 April 1991

We present the first nucleotide sequence data for molluscum contagiosum virus (MCV), an unclassified poxvirus. A 2,276-bp *XhoI* fragment from a near left-terminal fragment of MCV subtype I (MCVI) and a 1,920-bp *XhoI* fragment from the corresponding locus of MCV subtype II (MCVII) were sequenced and analyzed for open reading frames (ORFs). A large, complete ORF of 1,167 bp was present in both fragments. The putative polypeptide has a calculated molecular mass of 43 kDa (p43K protein) and was shown to have a high degree of homology to the vaccinia virus p37K major envelope antigen (40% amino acid identity and 22% conservative changes). The nucleotide content of the MCV fragments sequenced was 66% G or C. The codon usage within the gene for p43K reflected this high G+C content, with position 3 of codons being predominantly G or C (82 and 87% for MCVI and MCVII, respectively). The MCV p43K-encoding gene has motifs immediately upstream which are similar to those required for vaccinia virus late gene expression. The location and direction of transcription of the MCV p43K-encoding gene were equivalent to those of the vaccinia virus p37K gene, revealing similarity in genetic organization between MCV and vaccinia virus. Another, incomplete ORF was identified downstream of the p43K-encoding gene in both MCVI and MCVII. The sequence immediately upstream of this ORF overlapped the termination codon of the p43K-encoding gene and contained a motif which had homology to the derived consensus sequence for vaccinia virus early gene promoters.

Molluscum contagiosum virus (MCV) is an unclassified poxvirus of humans responsible for benign skin tumors. The skin lesion consists of a localized mass of hypertrophied and hyperplastic epidermis extending down into the underlying dermis and projecting above the surface as a papule. Individual lesions can persist for several weeks or longer, and the host immune system is implicated in their resolution (32). Initially, Parr et al. (24) described three subtypes of MCV on the basis of different restriction enzyme cleavage patterns of viral DNA. Other researchers reported two subtypes and showed heterogeneity in the DNA restriction patterns of the subtype: (5, 29–31, 38). More recently, three distinct subtypes have been identified (28a). These MCV subtypes differ in frequency of occurrence but induce clinically similar lesions (30, 31, 37). The MCV genome is similar to that of orthopoxviruses with regard to size, terminal cross-linking, and the presence of inverted terminal repeats (29), although there is no conservation of restriction sites within the internal region of the genome of different types as is seen among orthopoxviruses (19). However, fragments from each MCV subtype are colinear and cross-hybridize (29).

Vaccinia virus is the prototype poxvirus and has been studied in detail. Infection with vaccinia virus produces two types of infectious progeny virus: that extruded from the cell prior to cell lysis and that released during lysis. These extracellular and intracellular virus particles differ from each other with respect to the viral membrane. The extracellular vaccinia virus particles have an additional membrane which originates from the host Golgi apparatus and contains an abundant viral protein (5 to 7% of virion protein) with a molecular mass of 37 kDa (p37K) which is the major antigen of the extracellular virus (13, 14). Hirt et al. (15) have mapped and sequenced the p37K-encoding gene and characterized it as a late gene transcribed after the onset of DNA replication. These extracellular particles of vaccinia virus bearing envelopes containing the p37K protein have enhanced cellular absorption in infection compared with intracellular particles (28).

Expression of vaccinia virus genes is regulated temporally. Genes designated early or late are transcribed prior to DNA replication or after the onset of DNA replication, respectively (22). Elements responsible for regulation of this temporal transcription are present in the immediate 5′ upstream regions. Early promoters have been shown to consist of an A+T-rich critical sequence approximately 11 bases upstream from the site of initiation of transcription (6). The sequence TAAAT is highly conserved among vaccinia virus late genes, and transcription has been shown to initiate from within or close to this sequence (15, 34, 35, 43). These promoter elements appear to be highly conserved among different poxvirus genera (3, 10, 18, 42).

Detailed analysis of MCV has been limited by its inability to propagate in vitro, and little is known about the viral membrane and the proteins involved, although Oda et al. (23) identified 40 polypeptides from purified virions. During nucleotide sequencing studies, we identified an open reading frame (ORF) from a near left-terminal region of the MCV genome. Despite a lack of cross-hybridization between MCV and vaccinia virus DNAs, the encoded polypeptide was found to have extensive homology to vaccinia virus protein p37K.

MATERIALS AND METHODS

MCV purification and DNA extraction. Lesions were obtained from a variety of sites. Virus was purified, and the DNA was extracted as described by Porter and Archard (29).

* Corresponding author.
† Present address: Department of Immunology, Institute of Child Health, London WC1N 1EH, United Kingdom.

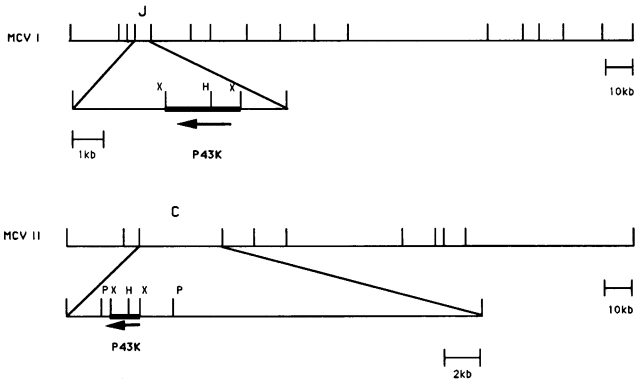

FIG. 1. *Bam*HI restriction map for MCVI and MCVII *Bam*HI fragments J of MCVI and C of MCVII are expanded. The *Xho*I fragments sequenced are indicated by bold lines. The arrows indicate the position and direction of transcription of the gene for p43K. X, *Xho*I; H, *Hin*dIII; P, *Pst*I.

Characterization of MCV subtypes. Viral DNA was cleaved with restriction endonuclease *Bam*HI (GIBCO-BRL) by standard methods (20). The viral isolate was then typed as either MCV subtype I (MCVI) or II (MCVII) on the basis of the cleavage pattern, as previously reported (29).

Cloning and sequencing. Cloning was done as described by Maniatis et al. (20). For MCVI, the *Bam*HI 6.4-kb J fragment was ligated into *Bam*HI-cut pBR322 (GIBCO-BRL), generating recombinant pB/J. From this, a 2.3-kb *Xho*I fragment was subcloned into *Sal*I-cut M13mp19 (Pharmacia) for sequencing. For MCVII, the *Pst*I 4.7-kb I fragment (3a) was ligated into *Pst*I-cut pUC19 (GIBCO-BRL), generating recombinant pUP/I, from which a 1.9-kb *Xho*I fragment was subcloned into *Sal*I-cut M13mp19 (Pharmacia) for sequencing. Templates for sequencing were prepared by creating unidirectional deletions with exonuclease III (GIBCO-BRL) as described by Henikoff (12). A sequence was generated by the chain termination method of Sanger et al. (36) by using the T7 DNA polymerase Sequenase (USB) (40). Problem templates were sequenced by using the DNA polymerase from *Thermus aquaticus* (*Taq* DNA polymerase) and the TaqTrack sequencing system (Promega Corp.).

Sequence analysis. Sequences were read with a sonic digitizer (Graf/Bar; Science Accessories Corp.). Data were assembled and analyzed by using the programs of Staden (39) and Devereux et al. (8) on a VAX/VMS V4.7. Analysis was also done with the DNA Strider version 1.0 program available for Apple Macintosh computers (21).

Nucleotide sequence accession number. The nucleotide sequence data reported here will appear in the EMBL, GenBank, and DDBJ nucleotide sequence data bases under accession numbers M63486 (MCVI) and M63487 (MCVII).

RESULTS

Cloning and sequencing. The *Bam*HI 6.4-kb J fragment of the MCVI genome was cloned into pBR322, from which a 2,276-bp *Xho*I fragment of MCVI DNA was subcloned into M13mp19 (*Sal*I site) in both orientations. Figure 1 shows the positions of the *Bam*HI J and *Xho*I fragments that were sequenced. Templates for sequencing were generated by creating a series of targeted deletions with exonuclease III (12) and sequenced by the dideoxy-chain termination method (36). Over 95% of the sequence was determined from both strands or from several different M13 subclones. The sequence of the 2,276-bp fragment is presented in Fig. 2A. The G+C content was 66%, in contrast to the 34% of vaccinia virus and in general agreement with earlier comparative denaturation studies (24).

The sequence was analyzed for possible ORFs (Fig. 3a), of which there were a large number because of the scarcity of stop codons resulting from the low A+T content. The most significant ORF (nucleotides 601 to 1767) was subjected to further analysis. It is 1,167 bp long and capable of encoding a polypeptide of 388 amino acids (Fig. 2A). There were several potential initiation codons within this region; that at position 601 was favored because its context (within the sequence TAAAATG) was very similar to the conserved sequence of TAAATG found at the start of vaccinia virus late genes (see below).

To ascertain whether MCVII encodes an equivalent polypeptide, the corresponding genetic locus of MCVII DNA was determined by studying cross-hybridization, which has been shown to be extensive outside the terminal repeats (28a). A 4.7-kb *Pst*I fragment was cloned into pUC19, from which a 1,920-bp *Xho*I fragment was subcloned into M13mp19 (Fig. 1) and sequenced as described above (Fig. 2B). The most significant ORF was in a similar position (Fig. 3b) and was 94% homologous at the nucleotide level to that of MCVI. The ORF of MCVII was 1,167 bp long, corresponding to a putative polypeptide of 388 amino acids (Fig. 2B). The 1.9-kb *Xho*I fragment of MCVII had a similarly high G+C content of 67%. The MCVII sequence data confirmed that the presumed start codon for the MCVI ORF was correct: there was no in-frame ATG codon upstream from the corresponding MCVII initiation codon. Codon usage within these ORFs reflected the high G+C content of the MCV genome. There was a strong bias for G or C in codon position 3: 82 and 87% of all codons for MCVI and MCVII, respectively.

Analysis of the p43 gene product. The ORF from MCVI codes for a putative polypeptide of 388 amino acids with a calculated molecular mass of 42.8 kDa, and that from MCVII encodes a putative polypeptide of 388 amino acids with a molecular mass of 42.6 kDa. Sequence comparison of the polypeptides revealed them to be very similar, with 95% identical amino acids and a further 2% conservative changes (Fig. 4a). We propose to call this potential protein p43K in accordance with its calculated molecular weight. A search of the NBRF protein data base with the p43K amino acid sequence revealed a high degree of homology with vaccinia virus major envelope antigen p37K (15). The molecular mass of MCVI protein p43K is slightly higher than the calculated size of 41.7 kDa for the vaccinia virus major envelope antigen (size, 37 kDa by gel electrophoresis). Comparison of MCVI p43K with the vaccinia virus envelope antigen demonstrated 40% amino acid identity with a further 22% conservative changes (Fig. 5). The greatest homology was at the C-terminal regions of the proteins (35 of 58 amino acids identical with a further 10 similar), although the MCV p43K carboxy terminus extends 13 amino acids beyond that of the vaccinia virus protein. Two significant internal gaps were inserted to obtain the best alignment: one of eight amino acids in MCVI p43K and one of seven amino acids in the vaccinia virus envelope antigen. Comparison of MCVII p43K showed the same features as MCVI (data not shown). Protein p37K is an integral part of the envelope of extracellular vaccinia virus and is reported to be associated with the Golgi apparatus within the cell (14). Two regions have been suggested to act as possible membrane-spanning regions,

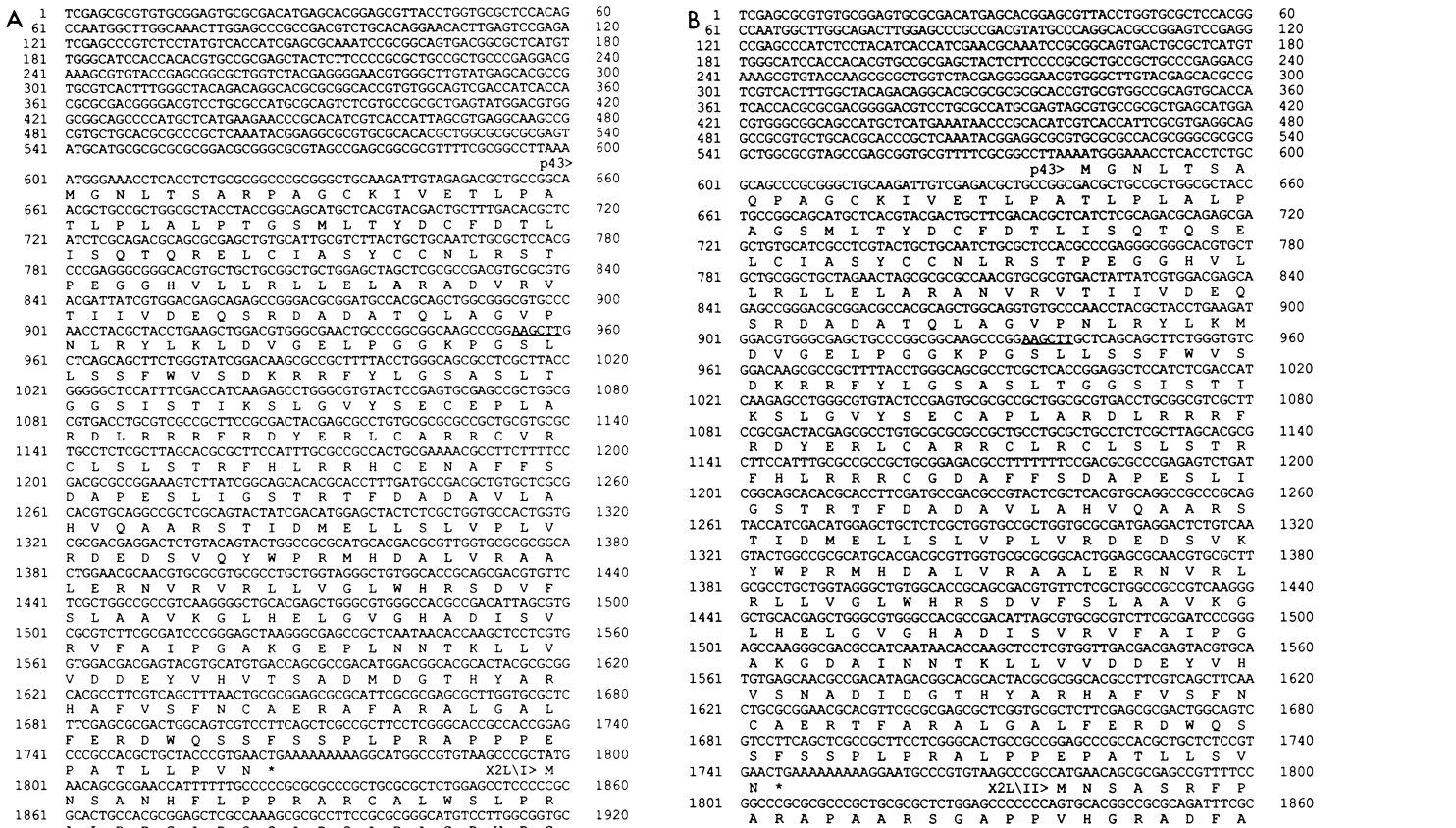

FIG. 2. Nucleotide sequences of the 2,276-bp *Xho*I fragment of MCVI (A) and the 1,920-bp *Xho*I fragment of MCVII (B). The amino acid sequences encoded by the gene for p43K and the partial X2L ORFs of both subtypes are shown in single-letter code under the nucleotide sequence. The internal *Hin*dIII site is underlined. The asterisk indicates the termination codon of the p43K genes.

amino acids 130 to 157 and 175 to 192 (15). Analysis of the hydrophobicity of p43K was done with the algorithm of Kyte and Doolittle (17). Two hydrophobic regions were identified in both MCVI p43K and MCVII p43K, at amino acids 17 to 35 and 133 to 153, respectively (Fig. 6), of which the latter is equivalent to hydrophobic amino acids 130 to 157, which form the putative transmembrane domain of vaccinia virus p37K.

Regulatory elements. The sequences immediately upstream of the MCVI and MCVII ORFs were analyzed for potential regulatory motifs. The 100 bases upstream of the translation initiation site are 72% G+C rich. The upstream sequences of the two genes are identical over the 37 bases preceding the initiation codon, with the exception of position 20 (C in MCVI and T in MCVII). Further upstream, the sequences of the two types diverge, although the high G+C content is maintained. Within the G+C-rich region, there are two short sequences notable for being A+T rich (Fig. 7a). For both MCVI and MCVII, there was the sequence TAAAATG, which includes the initiation codon, and 13 bp upstream of the start codon there was a stretch of four thymidine residues. The initiation codon of the p37K-encoding gene, a late gene in vaccinia virus, occurs within the sequence TAAATG and further upstream, starting at positions −12 and −69 bp, are stretches of five and four thymidine residues, respectively.

Sequences downstream of the MCV genes were also compared. Within the first 70 bp downstream of the termination codon, there was 75% homology between the sequences of MCVI and MCVII. The G+C content was maintained; MCVI was 67% G+C rich in 200 bp, whereas MCVII was 69% G+C rich in 173 bp following the termination codon. There is no reported conserved transcription termination sequence for vaccinia virus late genes, in contrast to early genes, where the sequence TTTTTNT is important for efficient termination (33, 44). Although the termination codon used by MCV was different from that used by the vaccinia virus p37K-encoding gene (TGA and TAA, respectively), the sequences immediately following the termination codon showed some similarity. The sequence seen for both MCVI and MCVII was AA AAAAAAGG, which bears a strong resemblance to the

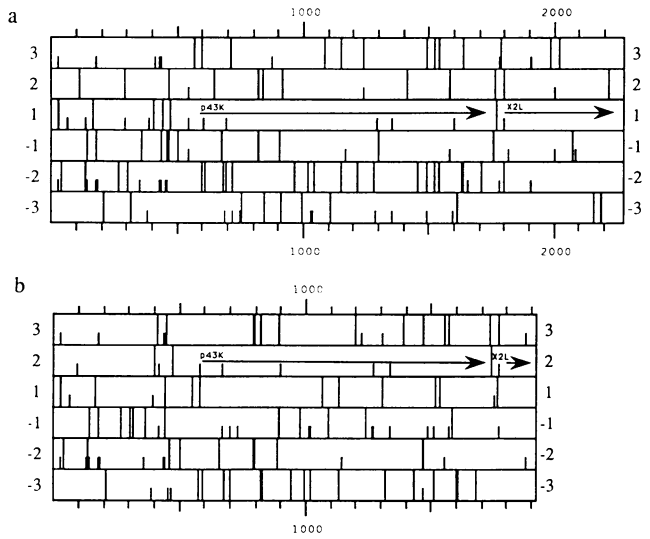

FIG. 3. Six-phase ORF analysis of the 2,276-bp *Xho*I fragment of MCVI (a) and the 1,920-bp *Xho*I fragment of MCVII (b). Short bars represent possible initiation codons, and tall bars represent stop codons. The ORFs of the genes for p43K are indicated by arrows (MCVI position 601 to 1,767 bp and MCVII position 581 to 1,747 bp), along with partial ORFs X2L\I and X2L\II (nucleotides 1798 to 2276 and 1778 to 1920).

p37K-encoding gene sequence of AAAAAAGAAAA (Fig. 7b). The sequence TTTTTNT was not found near the termination codon of either MCVI or MCVII, although the sequence TTTTTT occurred 48 bp downstream from the termination codon in the MCVI gene but not in the MCVII gene.

```
(a) MCV I p43    MGNLTSARPAGCKIVETLPATLPLALPTGSMLTYDCFDTLISQTQRELCI 50
    MCV IIp43    MGNLTSAQPAGCKIVETLPATLPLALPAGSMLTYDCFDTLISQTQSELCI 50
    MCV I p43    ASYCCNLRSTPEGGHVLLRLLELARADVRVTIIVDEQSRDADATQLAGVP 100
    MCV IIp43    ASYCCNLRSTPEGGHVLLRLLELARANVRVTIIVDEQSRDADATQLAGVP 100
    MCV I p43    NLRYLKLDVGELPGGKPGSLLSSFWVSDKRRFYLGSASLTGGSISTIKSL 150
    MCV IIp43    NLRYLKMDVGELPGGKPGSLLSSFWVSDKRRFYLGSASLTGGSISTIKSL 150
    MCV I p43    GVYSECEPLARDLRRRFRDYERLCARRCVRCLSLSTRFHLRRHCENAFFS 200
    MCV IIp43    GVYSECAPLARDLRRRFRDYERLCARRCLRCLSLSTRFHLRRRCGDAFFS 200
    MCV I p43    DAPESLIGSTRTFDADAVLAHVQAARSTIDMELLSLVPLVRDEDSVQYWP 250
    MCV IIp43    DAPESLIGSTRTFDADAVLAHVQAARSTIDMELLSLVPLVRDEDSVKYWP 250
    MCV I p43    RMHDALVRAALERNVRVRLLVGLWHRSDVFSLAAVKGLHELGVGHADISV 300
    MCV IIp43    RMHDALVRAALERNVRLRLLVGLWHRSDVFSLAAVKGLHELGVGHADISV 300
    MCV I p43    RVFAIPGAKGEPLNNTKLLVVDDEYVHVTSADMDGTHYARHAFVSFNCAE 350
    MCV IIp43    RVFAIPGAKGDAINNTKLLVVDDEYVHVSNADIDGTHYARHAFVSFNCAE 350
    MCV I p43    RAFARALGALFERDWQSSFSSPLPRAPPPEPATLLPVN 388
    MCV IIp43    RTFARALGALFERDWQSSFSSPLPRALPPEPATLLSVN 388

(b) X2L\I      1 MNSANHFLPPRARCALWSLPRALPRGARQSAPSARACPWRCAKQLRTREH 50
    X2L\II     1 MNSASRF.PARAPAARSGAPPVHGRADFAAGEARRGHVLGGAQSNRGL.. 47
```

FIG. 4. (a) Comparison of the amino acid sequences of the p43K proteins of MCVI and MCVII with the GAP program (8). The upper sequence is MCVI p43K, and the lower sequence is MCVII p43K. (b) GAP analysis of ORFs X2L\I and X2L\II. Lines represent identical amino acids and colons indicate conservative changes.

```
MCV I p43    .MGNLTSARPAGCKIVETLPATLPLALPTGSMLTYDCFDTLISQTQRELC 49
VACC p37     MWPFASVPAGAKCRLVETLPEN..MDFRSDHLTTFECFNEIITLAKKYIY 48
MCV I p43    IASYCCNLRSTPEGGHVLLRLLELARADVRVTIIVDEQSRDADATQLAGV 99
VACC p37     IASFCCNPLSTTRGALIFDKLKEASEKGIKIIVLLDERGKRNLGELQSHC 98
MCV I p43    PNLRYLKLDVGELPGGKPGSLLSSFWVSDKRRFYLGSASLTGGSISTIKS 149
VACC p37     PDINFITVNIDK..KNNVGLLLGCFWVSDDERCYVGNASFTGGSIHTIKT 146
MCV I p43    LGVYSECEPLARDLRRRFRDYE.....RLCARRCVRCLSLSTRFHLR 191
VACC p37     LGVYSDYPPLATDLRRRFDTFKAFNSAKNSWLNLCSAACCLPVSTAYHIK 196
MCV I p43    RHCENAFFSDAPESLIGSTRTFDADAVLAHVQAARSTIDMELLSLVPLVR 241
VACC p37     NPIGGVFFTDSPEHLLGYSRDLDTDVVIDKLKSAKTSIDIEHLAIVPTTR 246
MCV I p43    DEDSVQYWPRMHDALVRAALERNVRVRLLVGLWHRSDVFSLAAVKGLHEL 291
VACC p37     VDGNSYYWPDIYNSIIEAAINRGVKIRLLVGNWDKNDVYSMATARSLDAL 296
MCV I p43    GVGHADISVRVFAIPGAKGEPLNNTKLLVVDDEYVHVTSADMDGTHYARH 341
VACC p37     CV.QNDLSVKVFTIQ.....NNTKLLIVDDEYVHITSANFDGTHYQNH 338
MCV I p43    AFVSFNCAERAFARALGALFERDWQSSFSSPLPRAPPPEPATLLPVN 388
VACC p37     GFVSFNSIDKQLVSEAKKIFERDWVSSHSKSLKI..... 372
```

FIG. 5. Comparison of the amino acid sequences of the MCVI p43K and vaccinia virus (VACC) p37K proteins with GAP (8). The upper sequence is p43K, and the lower sequence is p37K. Lines represent identical amino acids, and colons indicate conservative changes.

Analysis of other ORFs. Analysis of other possible ORFs revealed the presence of a partial ORF immediately downstream from the gene for p43K in both subtypes (Fig. 3, ORFs X2L\I and X2L\II). More sequence data were obtained for MCVI than MCVII, and the corresponding partial ORF was larger: ORF X2L\I, nucleotides 1798 to 2276 (479 bp) and ORF X2L\II, nucleotides 1778 to 1920 (143 bp) (Fig. 2). The putative polypeptides encoded by these incomplete ORFs are 159 (MCVI) and 47 (MCVII) amino acids. Comparisons of the two polypeptides showed little homology in the short regions characterized (Fig. 4b). It remains to be determined whether they are in fact complete ORFs of similar lengths and encode related proteins. A search of the NBRF data base revealed no significant matches with any

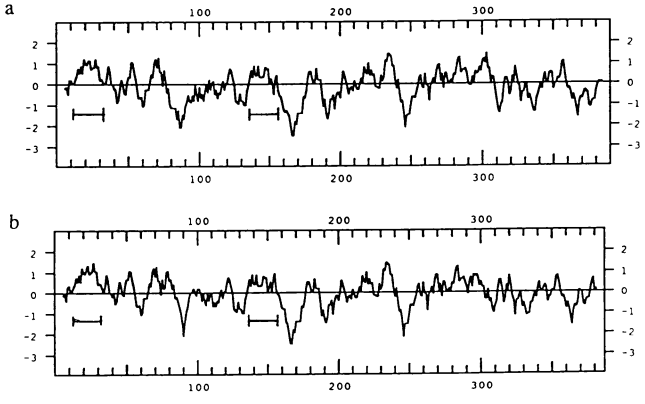

FIG. 6. Hydrophobicity plots, based on the algorithm of Kyte and Doolittle (17), of the MCVI p43K (a) and MCVII p43K (b) proteins. The *x* axis refers to amino acid numbers, and the *y* axis refers to hydrophobicity values. Proposed hydrophobic regions are amino acids 17 to 35 and 133 to 153 (bars).

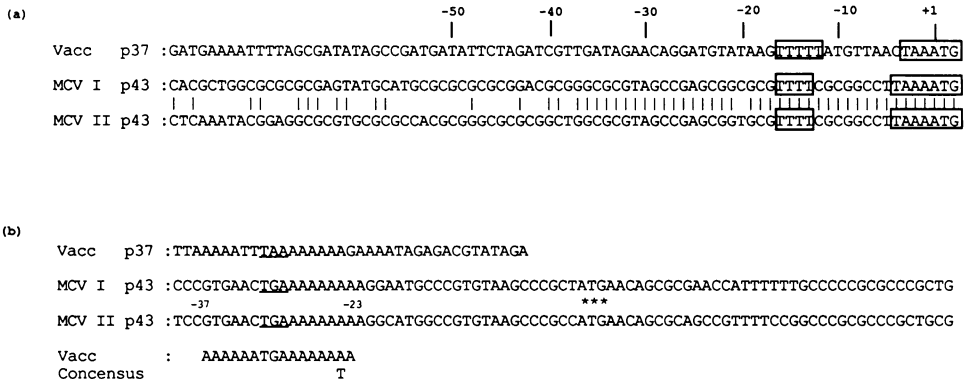

FIG. 7. (a) Comparison of the 5′ upstream regions of the MCVI and MCVII genes for p43K and the vaccinia virus (Vacc) gene for p37K. Bars represent identical nucleotides between the MCVI and MCVII sequences. Boxed regions highlight conserved elements between the MCV and vaccinia virus sequences. (b) Comparison of sequences around the termination codons of the MCVI and MCVII genes for p43K and the vaccinia virus gene for p37K. The termination codons of these genes are underlined, and the initiation codon of ORFs X2L\I and X2L\II is highlighted by asterisks. The consensus sequence for the critical region of vaccinia virus early promoters (6) is aligned with the possible early promoter of the MCV X2L ORFs.

previously identified peptide sequences for these putative translation products.

Analysis of the 5′ upstream region of ORFs X2L\I and X2L\II revealed homology between the two subtypes (Fig. 7b). The upstream region overlaps with the termination codon of the p43K-encoding gene in both cases. At −23 to −37 bp upstream of the initiation codon of these ORFs, an A-rich region was found which includes the termination codon and possible termination sequence of the gene for p43K. Inspection of the sequence and comparison with known vaccinia virus early promoters indicated that the sequence may represent an early promoter region and not the previously discussed termination signal. Comparison with a consensus sequence for vaccinia virus early promoters derived from 19 early promoters (6) showed extensive homology between the critical region of vaccinia virus early promoters (AAAAAATGAAAAAAAA) and the sequence from MCV (GTGAACTGAAAAAAAA); i.e., 12 of 16 residues are identical (Fig. 7b). This suggests that the sequence is more likely to be an early promoter in MCV than a late termination signal.

Genetic organization. The gene for p43K was identified from sequence data within the *Bam*HI J fragment of MCVI and the *Bam*HI C fragment of MCVII. These fragments correspond to an equivalent region of the viral genome (29). By positioning of a *Hin*dIII site within the coding region of the gene for p43K, the gene was localized to approximately 30 kb from the left terminus of MCVI and 32 kb from the left terminus of MCVII for the isolates used in this study, corresponding to approximately 25 kb from the junction of the left terminal repeat and left unique sequences (Fig. 1). The sequence data indicate that the direction of transcription of both genes was from right to left, i.e., toward the terminal repeat. With respect to relative position within the MCV genome and its direction of transcription, the p43K-encoding gene was analogous to the vaccinia virus p37K-encoding gene (15). The complete sequence of the vaccinia virus Copenhagen strain has been obtained and reveals an ORF (designated F12L) immediately downstream of the p37K-encoding gene (11). The position and direction of transcription of the X2L ORFs in MCV are equivalent to those of this ORF. However, when the polypeptide sequences of these ORFs (X2L and F12L) were compared no homology was found.

DISCUSSION

Analysis of the nucleotide sequence of a region close to the left terminus of MCV revealed the presence of an ORF, present in both major viral subtypes, which has the potential to encode a polypeptide with a molecular mass of approximately 43 kDa. The predicted polypeptide was found to have a high degree of homology to a previously sequenced vaccinia virus late gene product, protein p37K. The gene for p37K of vaccinia virus is a late gene located within the *Hin*dIII F fragment of the virus genome (15). The gene product, the p37K protein, is the major antigen present on the envelope of extracellular vaccinia virus. It is known to be associated with the Golgi apparatus within infected cells and is proposed to have a role in the release of mature vaccinia virus particles from infected cells (13, 14).

MCV p43K has 42% identity and a further 20% similarity to vaccinia virus p37K at the amino acid level, despite the marked difference in nucleotide composition between the MCV and vaccinia virus DNAs. Vaccinia virus has a G+C content of 34%, whereas MCV has a G+C content of 66%. This difference is reflected in codon usage in the gene for p43K, which has a bias of 82% G or C in the base 3 position of codons.

Hirt et al. (15) proposed that the hydrophobic stretch of amino acids 130 to 157 of vaccinia virus p37K represents a transmembrane region. This region appears to be conserved within MCV p43K, with equivalent amino acids 133 to 153 being hydrophobic. p43K has an additional hydrophobic region much earlier in the sequence, at amino acids 17 to 43, which may also be membrane associated. Vaccinia virus exists in two forms, described as intracellular naked vaccinia virus and extracellular enveloped vaccinia virus (25, 26). The extracellular form has an additional envelope derived from the host Golgi apparatus (13). This additional extracellular envelope has also been reported on other orthopoxviruses closely related to vaccinia virus, including rabbitpox virus (1) and cowpox virus (27). A protein was identified in cowpox virus-infected cells which had a similar molecular

weight and cross-reacted antigenically with vaccinia virus protein p37K (27). p37K is one of several proteins found in the vaccinia virus envelope (25) and is reported to be the major antigen of the extracellular virus (14). The ability of MCV to encode a homolog, protein p43K, suggests that this non-orthopoxvirus also has a modified envelope in the extracellular form. Oda et al. (23) have studied the structural polypeptide components of MCV isolates and identified 40 polypeptides from purified virons. However, no protein was identified by polyacrylamide gel electrophoresis which corresponds to the molecular weight of p43K calculated from the nucleotide sequence. However, it is not known whether protein p43K is posttranslationally modified, and therefore it may have an anomalous size as characterized by polyacrylamide gel electrophoresis.

In addition to conservation of the gene products at the amino acid level, the elements involved in control of expression of the vaccinia virus p37K-encoding and MCV p43K-encoding genes appear to be conserved. The gene for p37K is expressed after the onset of viral DNA replication and has the reported features of late gene promoters (15). The sequence TAAATG, which includes both the transcriptional start site and the initiation codon, or a closely related sequence is found in most of the vaccinia virus late genes sequenced, along with upstream stretches of thymidine or adenine residues (2, 7, 34, 35, 43). The 5′ region of the MCV gene for p43K is very G+C rich (72%) overall, but within this region, the A+T-rich sequence TAAAATG encompassed the initiation codon. At position −13 bp from the initiation codon, there is a string of four thymidine residues. The presence of these two motifs (TAAAATG and TTTT) is more obvious in the G+C-rich MCV genome than in the A+T-rich vaccinia virus genome and implies conservation of control of expression of vaccinia virus p37K and its MCV homolog. The promoter regions of the MCVI and MCVII p43K-encoding genes are highly conserved (96%) until position −33 bp from the initiation codon, before which the two sequences diverge while maintaining the high G+C content. This suggests that the control elements necessary for transcription of this MCV gene are present within the immediate upstream region, analogous to the situation in vaccinia virus genes (2).

The sequence motif TAAAAT upstream of the gene for p43K differs from the TAAAT motif associated with vaccinia virus late genes by the presence of an additional adenine residue. The motif TAAAAT is found in the same position relative to a designated early gene in vaccinia virus, ORF K3 (4). ORF K3 occurs within the vaccinia virus *Hin*dIII K fragment and encodes a protein with some homology to vaccinia virus protein p37K (25% amino acid identity and 23% conservative changes). p43K of MCV has a similar level of homology to the vaccinia virus K3-encoded protein, with 22% amino acid identity and a further 23 to 25% conservative changes (data not shown). Recently, Kotwal et al. (16) have shown that an ORF (N1L) located at the junction of the *Hin*dIII C and N fragments, encoding a secreted protein, has the sequence TAAAAT, which acts as a late transcriptional start site. Since the homology at the amino acid level of p43K to the K3-encoded protein is much lower (22% identity) than that between p43K and p37K (40% identity), we assume that MCV protein p43K is the homolog of vaccinia virus protein p37K.

A partial ORF (X2L) and a possible early promoter element were identified in the *Xho*I fragments sequence. ORF X2L (X2L\I and X2L\II) immediately follows the gene for p43K in both MCV subtypes. There was no homology between the incomplete polypeptides of 159 and 47 amino acids sequenced from MCVI and MCVII, respectively, and no matches with other proteins were found in a data base search. The upstream region of each of these ORFs overlapped the termination codon of the gene for p43K. Detailed analysis of this 5′ region suggested that the sequence around the termination codon of the gene for p43K is an early promoter. Davison and Moss (6) have shown that an A-rich sequence (consensus of AAAAAATGAAAAAAA/TA) at −13 to −28 bp from the transcriptional start site comprises the critical region of vaccinia virus early promoters. The sequence at −23 to −37 bp from the initiation codon of ORFs X2L\I and X2L\II, which overlaps with the termination codon of the genes for p43K, is very similar to this derived consensus, with 12 of 16 nucleotides identical (Fig. 7b). However, it has still to be established that this is a complete reading frame.

A number of poxviruses are known to be similar in genetic organization. For instance, the cluster of vaccinia virus genes including that for thymidine kinase is present at an equivalent position of the capripoxvirus and Shope fibroma virus genomes (10, 41). Although the thymidine kinase-encoding gene in fowlpox virus has been mapped to a different locus, the region corresponding to the equivalent position of the gene for thymidine kinase in the poxviruses mentioned above shows conservation of the other ORFs (3, 9). The gene for p43K is found in the *Bam*HI J fragment of MCVI, approximately 30 kb from the arbitrarily designated left terminus, and is transcribed from right to left toward the inverted terminal repeat. The situation for MCVII is identical (Fig. 1). The gene for p37K in vaccinia virus is located in the *Hin*dIII F fragment at a position 38 kb from the left end and transcribed toward the terminus. When the sizes of the terminal repeats are taken into account (3 to 5 kb in MCV and 10 kb in vaccinia virus), it is apparent that these genes are found in similar genetic loci, in addition to being in the same orientation. An ORF (F12L) is present downstream of the gene for p37K in vaccinia virus (11). MCV ORFs X2L\I and X2L\II occur in similar positions, although no homology was found between the amino acid translation products of these ORFs. Thus, the head-to-tail arrangement of ORFs within the vaccinia virus genome is also apparent for MCV.

Further analysis of other ORFs and regulatory elements is necessary, but it appears that the genetic organization and control of expression of MCV genes are similar in many respects to those of antigenically unrelated poxviruses.

REFERENCES

1. **Appleyard, G., J. A. Hapel, and E. A. Boulter.** 1971. An antigenic difference between intracellular and extracellular rabbitpox virus. J. Gen. Virol. **13**:9–17.
2. **Bertholet, C., R. Drillien, and R. Wittek.** 1985. One hundred base pairs of 5′ flanking sequence of a vaccinia virus late gene are sufficient to temporally regulate late transcription. Proc. Natl. Acad. Sci. USA **82**:2096–2100.
3. **Binns, M. M., F. M. Tomley, J. I. Campbell, and M. E. Boursnell.** 1988. Comparison of a conserved region in fowlpox virus and vaccinia virus genomes and the translocation of the fowlpox virus thymidine kinase gene. J. Gen. Virol. **69**:1275–1283.

3a. **Blake, N. W., C. D. Porter, and L. C. Archard.** Unpublished data.

4. **Boursnell, M. E., I. J. Foulds, J. I. Campbell, and M. M. Binns.** 1988. 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. J. Gen. Virol. **69**:2995–3003.
5. **Darai, G., H. Reisner, J. Scholz, P. Schnitzler, and H. Lorbacher**

de Ruiz. 1986. Analysis of the genome of *molluscum contagiosum* virus by restriction endonuclease analysis and molecular cloning. J. Med. Virol. **18**:29–39.
6. **Davison, A. J., and B. Moss.** 1989. Structure of vaccinia virus early promoters. J. Mol. Biol. **210**:749–769.
7. **Davison, A. J., and B. Moss.** 1989. Structure of vaccinia virus late promoters. J. Mol. Biol. **210**:771–784.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
9. **Drillien, R., D. Spehner, D. Villeval, and J.-P. Lecocq.** 1987. Similar genetic organization between a region of fowlpox virus DNA and the vaccinia virus *Hin*dIII J fragment despite divergent location of the thymidine kinase gene. Virology **160**:203–209.
10. **Gershon, P. D., and D. N. Black.** 1989. The nucleotide sequence around the capripoxvirus thymidine kinase gene reveals a gene shared specifically with leporipoxvirus. J. Gen. Virol. **70**:525–533.
11. **Goebel, S. J., G. P. Johnson, M. E. Perkus, S. W. Davis, J. P. Winslow, and E. Paoletti.** 1990. The complete DNA sequence of vaccinia virus. Virology **179**:247–266, 517–564 (appendix).
12. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
13. **Hiller, G., H. Eibl, and K. Weber.** 1981. Characterization of intracellular and extracellular vaccinia virus variants: N_1-isonicotinoyl-N_2-3-methyl-4-chlorobenzoylhydrazine interferes with cytoplasmic virus dissemination and release. J. Virol. **39**:903–913.
14. **Hiller, G., and K. Weber.** 1985. Golgi-derived membranes that contain an acylated viral polypeptide are used for vaccinia virus envelopment. J. Virol. **55**:651–659.
15. **Hirt, P., G. Hiller, and R. Wittek.** 1986. Localization and fine structure of a vaccinia virus gene encoding an envelope antigen. J. Virol. **58**:757–764.
16. **Kotwal, G. J., A. W. Hugin, and B. Moss.** 1989. Mapping and insertional mutagenesis of a vaccinia virus gene encoding a 13,800-Da secreted protein. Virology **171**:579–587.
17. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. **157**:105–132.
18. **Macaulay, C., and G. Mcfadden.** 1989. Tumourigenic poxviruses: characterization of an early promoter from Shope fibroma virus. Virology **172**:237–246.
19. **Mackett, M., and L. C. Archard.** 1979. Conservation and variation in *orthopoxvirus* genome structure. J. Gen. Virol. **45**: 683–701.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Marck, C.** 1988. 'DNA Strider'; a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
22. **Moss, B.** 1985. Replication of poxviruses, p. 658–703. *In* B. N. Fields (ed.), Virology. Raven Press, New York.
23. **Oda, H., Y. Ohyama, T. Sameshima, and K. Hirakawa.** 1982. Structural polypeptides of *molluscum contagiosum* virus: their variability within various isolates and location within the viron. J. Med. Virol. **9**:19–25.
24. **Parr, R. P., J. W. Burnett, and C. F. Garon.** 1977. Structural characterisation of the *molluscum contagiosum* virus. Virology **81**:247–256.
25. **Payne, L. G.** 1978. Polypeptide composition of extracellular enveloped vaccinia virus. J. Virol. **27**:28–37.
26. **Payne, L. G.** 1980. Significance of extracellular enveloped virus in the in vitro and in vivo dissemination of vaccinia. J. Gen. Virol. **50**:89–100.
27. **Payne, L. G.** 1986. The existence of an envelope on extracellular cowpox virus and its antigenic relationship to the vaccinia envelope. Arch. Virol. **90**:125–133.
28. **Payne, L. G., and E. Norrby.** 1978. Adsorption and penetration of enveloped and naked vaccinia virus particles. J. Virol. **27**:19–27.
28a. **Porter, C. D.** 1989. Ph.D. thesis. University of London, London, United Kingdom.
29. **Porter, C. D., and L. C. Archard.** 1987. Characterisation and physical mapping of *molluscum contagiosum* virus DNA and location of a sequence capable of encoding a conserved domain of epidermal growth factor. J. Gen. Virol. **68**:673–682.
30. **Porter, C. D., N. W. Blake, L. C. Archard, M. F. Muhlemann, N. Rosedale, and J. J. Cream.** 1989. *Molluscum contagiosum* virus types in genital and non-genital lesions. Br. J. Dermatol. **120**:37–41.
31. **Porter, C. D., M. F. Muhlemann, J. J. Cream, and L. C. Archard.** 1987. *Molluscum contagiosum*: characterization of viral DNA and clinical features. Epidemiol. Infect. **99**:563–567.
32. **Postlethwaite, R.** 1970. *Molluscum contagiosum*: a review. Arch. Environ. Health **21**:432–452.
33. **Rohrmann, G., L. Yuen, and B. Moss.** 1986. Transcription of vaccinia virus early genes by enzymes isolated from vaccinia virions terminates downstream of a regulatory sequence. Cell **46**:1029–1035.
34. **Rosel, J. L., P. L. Earl, J. P. Weir, and B. Moss.** 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 H genome fragment. J. Virol. **60**:436–449.
35. **Rosel, J. L., and B. Moss.** 1985. Transcriptional and translational mapping and nucleotide sequence analysis of a vaccinia virus gene encoding the precursor of the major core polypeptide 4b. J. Virol. **56**:830–838.
36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
37. **Scholz, J., A. Rosen-Wolff, J. Burgert, H. Reisner, M. I. White, G. Darai, and R. Postlethwaite.** 1988. Molecular epidemiology of *molluscum contagiosum*. J. Infect. Dis. **158**:898–900.
38. **Scholz, J., A. Rosen-Wolff, J. Burgert, H. Reisner, M. I. White, G. Darai, and R. Postlethwaite.** 1989. Epidemiology of *molluscum contagiosum* using genetic analysis of the viral DNA. J. Med. Virol. **27**:87–90.
39. **Staden, R.** 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. **14**:217–231.
40. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
41. **Upton, C., and G. McFadden.** 1986. Identification and nucleotide sequence of the thymidine kinase gene of Shope fibroma virus. J. Virol. **60**:920–927.
42. **Upton, C., A. Opgenarth, P. Traktman, and G. McFadden.** 1990. Identification and DNA sequence of the Shope fibroma virus DNA topoisomerase gene. Virology **176**:439–447.
43. **Weir, J. P., and B. Moss.** 1984. Regulation of expression and nucleotide sequence of a late vaccinia virus gene. J. Virol. **51**:662–669.
44. **Yuen, L., and B. Moss.** 1987. Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. Proc. Natl. Acad. Sci. USA **84**:6417–6421.