



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Transcriptional mapping and nucleotide sequence of a vaccinia virus gene encoding a polypeptide with extensive homology to DNA ligases

Citation for published version:

Smith, GL, Chan, YS & Kerr, SM 1989, 'Transcriptional mapping and nucleotide sequence of a vaccinia virus gene encoding a polypeptide with extensive homology to DNA ligases', *Nucleic Acids Research*, vol. 17, no. 22, pp. 9051-62. <https://doi.org/10.1093/nar/17.22.9051>

Digital Object Identifier (DOI):

[10.1093/nar/17.22.9051](https://doi.org/10.1093/nar/17.22.9051)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Nucleic Acids Research

Publisher Rights Statement:

Copyright © 2013 Oxford University Press

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Transcriptional mapping and nucleotide sequence of a vaccinia virus gene encoding a polypeptide with extensive homology to DNA ligases

Geoffrey L.Smith*, Y.Sang Chan and Shona M.Kerr

Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

Received August 25, 1989; Accepted September 25, 1989

EMBL accession no. X16512

ABSTRACT

Nucleotide sequencing of the vaccinia virus Sali F DNA fragment identified an open reading frame of 552 amino acids encoding a protein of 63.3 kDa. The deduced amino acid sequence shares 30% identity with *S. pombe* and *S. cerevisiae* DNA ligases with homology strongest near the carboxy terminus and around the lysine residue required for ligase-adenylate formation. Prokaryotic DNA ligases are poorly related to the vaccinia sequence. The initiation codon of the ORF forms part of a late transcriptional initiation sequence TAAATG and is preceded by two overlapping early transcriptional termination signals, TTTTTTAT. Nonetheless, RNA mapping showed that the ligase gene is transcribed early during infection and the 5' end of the mRNA maps to the TAAATG motif. The possible roles of a DNA ligase in vaccinia virus DNA replication and recombination are discussed.

INTRODUCTION

Vaccinia virus is an orthopoxvirus, a group of viruses which transcribe and replicate their large double stranded DNA genomes in the cytoplasm of infected cells (1). This replication site distinguishes poxviruses from many other DNA viruses which replicate in the nucleus and which mostly use endogenous nuclear enzymes for transcription and replication. Cytoplasmic replication requires that the enzymes for these processes are provided by the infecting virus and consequently poxviruses encode many enzymes involved in nucleic acid metabolism. Thus, they provide an excellent system for studying transcription, DNA replication and recombination in a eukaryotic cell. Poxviruses encode and package a multi-subunit DNA-dependent RNA polymerase (2–6) and enzymes to cap, methylate and polyadenylate mRNA (7–10). For DNA metabolism the virus packages a topoisomerase (11–12), nicking-joining enzyme (13–14) and single stranded-specific DNAase (15). Within the infected cell a DNA-dependent DNA polymerase (16–19), hetero-dimeric ribonucleotide reductase (20–22), thymidine kinase (23–25) and DNA ligase (26,27) are expressed. The single strand-specific DNAase, nicking-joining enzyme and DNA ligase have not been proven to be virus encoded.

DNA ligases and the genes encoding them have been identified in *E. coli*, bacteriophages T4 and T7 and yeasts *S. cerevisiae* and *S. pombe* (28–32). In these organisms the DNA ligase is an essential enzyme involved in DNA replication, recombination and repair (reviewed in 33). The genes encoding mammalian DNA ligases have not been identified although there has been considerable biochemical characterisation of the enzymes (reviewed in 34). In this report the structure of a vaccinia virus gene encoding a polypeptide of 63.3 kDa which has high amino acid sequence homology to DNA ligases of yeasts and limited homology to prokaryotic DNA ligases is presented. Transcriptional analyses demonstrate that the gene is transcribed early during infection. The high degree of homology to both

yeast DNA ligases particularly near the carboxy terminus may aid the identification and cloning of mammalian DNA ligase genes via use of synthetic peptides or oligonucleotides.

MATERIALS AND METHODS

Nucleotide sequencing

The 13.4 kb SalI F fragment of vaccinia virus (strain WR) was isolated from cosmid 6, which contains virus DNA derived from a rifampicin resistant mutant (35), and was cloned into pUC13. The SalI fragment was separated from plasmid sequences and self-ligated with T4 DNA ligase. Circular molecules were randomly sheared by sonication, end-repaired with T4 DNA polymerase and Klenow enzyme and fragments of greater than 300 nucleotides cloned into SmaI cut M13mp18. Single stranded DNA was prepared and sequenced using the dideoxynucleotide chain termination method (36), using [³⁵S]-dATP and buffer gradient polyacrylamide gels (37). For further details see (38).

Computer analysis

Nucleotide sequence data were read from autoradiographs by sonic digitiser and assembled into contiguous sequences using programmes DBAUTO and DBUTIL (39,40) on a VAX 8350 computer. The consensus sequence was translated in 6 frames using programmes ORFFILE and DELIB (M. Boursnell, Institute of Animal Health, Houghton, UK.). Open reading frames were compared against SWISSPROT protein database and against our own database of vaccinia amino acid sequences using programme FASTP (41). Alignments of multiple protein sequences were performed using programme MULTALIGN (42).

RNA preparation

Early or late virus mRNA was prepared from human TK⁻143 cells infected with vaccinia virus strain WR at 30 plaque forming units per cell as described (43).

Northern analysis

Denatured RNA samples were electrophoresed through agarose gels, transferred to nitrocellulose membranes and probed with single stranded [³²P]-labelled DNA probes as described (43).

S1 nuclease mapping of RNA

RNA samples (10 µg) were mixed with 90 µg tRNA and a 365 bp ClaI-BclI DNA fragment [³²P]-labelled at the 5' terminal phosphate of the BclI site. Nucleic acid was ethanol precipitated, redissolved in 30 µl hybridization buffer (80% dionized formamide, 0.4 M NaCl, 1 mM EDTA (pH 8.0), 40 mM PIPES), denatured at 72°C for 15 min and hybridized overnight at 30°C. 300 units of S1 nuclease were added in 0.3 ml 0.56 M NaCl, 100 mM sodium acetate (pH 4.6), 9 mM ZnSO₄, 20 µg/ml denatured carrier DNA and incubated for 1 h at 30°C. The reaction was stopped with 4 µl 0.25 M EDTA, nucleic acid extracted with phenol/chloroform and precipitated with ethanol. Samples were electrophoresed on a 12% polyacrylamide gel in TBE buffer.

Preparation of prime cut probe

Single stranded DNA from M13 clone 3F, which contains vaccinia DNA from entirely within the SalF 15R sequence, was annealed with sequencing primer and incubated with Klenow enzyme, 10 µCi [³²P]-labelled dCTP and 0.2 mM dGTP, dATP and dCTP for 15 min at 25°C, then 0.2 mM dCTP was added and incubation continued for 15 min. The product was cut with EcoRI and SalI and the digests electrophoresed on a 6% polyacrylamide gel in TBE. The labelled, single stranded DNA fragment was detected by brief autoradiography and eluted from crushed acrylamide slice in 500 mM ammonium

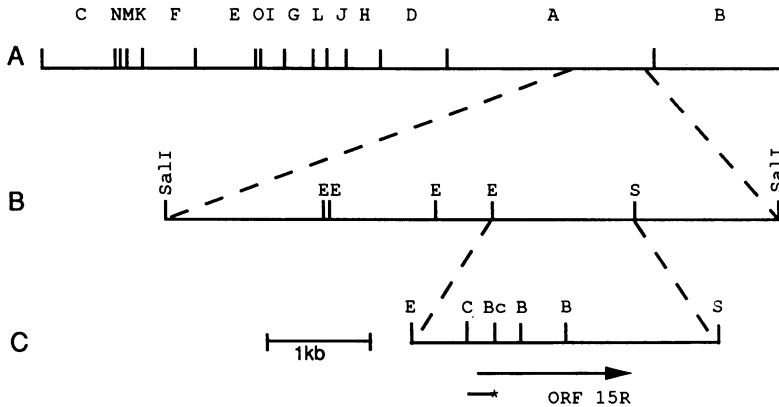


Figure 1 Location and direction of transcription of DNA ligase gene within the vaccinia virus genome. A. Vaccinia virus HindII restriction map. B. The 13.4 kb SalI F restriction fragment is expanded and the positions of EcoRI (E) SmaI(S) restriction sites are indicated. C. A 3300 bp EcoRI-SmaI fragment is expanded showing the position of the DNA ligase gene (arrow), the position of [³²P]-labelled probe used for S1 nuclease mapping and the positions of EcoRI (E), ClaI (C), BclI (Bc), BglII (B) and SmaI (S) restriction sites.

acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS and 10 μ g/ml tRNA overnight at 37°C.

Primer extension analyses

An oligonucleotide 5' CCGGGAAGCAATAGCTTAATG 3' was 5'-labelled with γ -[³²P]-ATP using polynucleotide kinase and hybridized to denatured RNA samples as for S1 nuclease analysis. Hybridized nucleic acid was recovered by ethanol precipitation and incubated with avian myeloblastosis virus reverse transcriptase as described (44).

RESULTS

Figure 1 shows the position of the SalI F fragment within the vaccinia virus genome. This fragment was sequenced using the random shotgun method (38) and a contiguous DNA sequence of 13,411 nucleotides with an average character density of 6.1 was obtained. The complete sequence will be published elsewhere. Translation of this sequence detected more than 20 ORFs of greater than 66 amino acids and which fitted the codon usage of established vaccinia genes. The position and direction of transcription of the largest ORF (SalF 15R) is indicated in Figure 1 and the nucleotide sequence and deduced amino acid sequence are shown in Figure 2. This ORF maps between 7625 and 9280 nucleotides from the left end of the SalI F fragment and encodes a polypeptide of 552 amino acids with a predicted molecular weight of 63.3 kD.

The ATG codon at the beginning of SalF 15R is part of a conserved motif (TAAATG) which is found at the transcriptional initiation site of most late genes (45,46). Immediately upstream is the sequence TTTTAT which represents two overlapping, early, transcriptional termination signals (47). The next early transcriptional termination signal is 2730 nucleotides downstream, 1060 nucleotides beyond the end of SalF 15R. Interestingly, the sequence TTTTNT is found just upstream of the TAAATG motif within two strongly expressed late poxvirus promoters. The cowpox virus gene encoding an abundant

!!! M T S L R E F R K L C C D I Y H A S G Y 20
TAATAATGATATCCGGAAACGATATTTCTGTAATAATTAAGAAATTTAGAAATTTAGCTGTGATATATATCAGCATCAGGATAT 120
K E K S K L I R D F I T D R D K Y L I I K L L L P G L D R I Y N M N D K Q I 60
MAAGAAAAATCTAAATTAATAGACATTAACACAGATAGGATGATAAATTTTGATCATTAAGCTATTTGCTTCCCGGATTTAGACGATAGAATTTATAACATGAACGATAAACAATTT 240
I K L Y S I I F K Q S Q E D M L Q D L G Y G Y I G D T I R T F F K E N T E I R P 100
TAATAAATATATAGTATAATTTAAACAATCTCAGGAAGATATGCTACAGATTTAGGATACGGATATATAGGACACACTATTAGGACTTTCTTCAAGAGAACAACAGAAATCCCGTCCA 360
R D K S I L T L E D V D S F L T T L S S V T K E S H Q I K L L T D I A S V C T C 140
CCGAGATAAAGCATTTTAACTTTAGAACGCTGGATAGTTTCTTAACACGTTACGTATCATCCGTAACATAAGAAATCGCATCAAAATAAAATTTATTGACTGATATCGCATCCGTTTGTACATGT 480
N D L K C V V M L I D K D L K I K A G P R Y V L N A I S P N A Y D V F R K S N N 180
TAATGATTTAAAAATGCTGTCATGCTTATTGATTAAGATCTAAAAAATTAAGCGGGCTCTCGGTACGCTTAACCGCTATTAGTCTTAATGCCATGATGATGTGTTTGAAGAAATCTAATAAC 600
L K E I I E N A S K Q N L D S I S I S V M T P I N P M L A E S C D S V N K A F K 220
TTTGAAGAGATAATAGAAATGCACTTAACAAAATCTAGACCTATATCTATTCTGTTAGACTCCAATTAATCCCATGTTAGCGGAATCGTGTGATCTGTGCAATAAGGCGTTTAAA 720
K F P S G M F A E V K Y D G E R V Q V H K N N N E F A F S R N M K P V L S H K 260
AAAATTTCCATCAGGAATGTTCCGGAATGTAATACGATCGTGAAGAGATCAAGTCTCATAAAAATTAATACAGGTTTCCCTTCTTTAGTAGAACAATGAACCCAGTACTCTCTCATAAA 840
V D Y L K E Y I P K A F K K A T S I V L D S E I V L V D E H N V P L P F G S L G 300
GTGGATTAATCTCAAGAAATACATACCGAAACCATTTAAAAAGCTACGTCTATCGTATTTGATTTCTGAAATTTGTTCTGTAGACCAACATTAATGTCACCGTCCCGTTTGAAGTTTAGGA 960
I H K K K E Y K N S N M C L F V F D C L Y F D G F D M T D I P L Y E R R S F L K 340
ATACACAAAAGAAAGAAATATAAAAATCTCAACATGTGTTGTTGCTGTTGACGTGTTGACTTTTGATTCGATTCGATATGACGACATTCGATTTGACGAACGAGATCTTTTCTCAAA 1080
D V M V E I P N R I V F S E L T N I S N E S Q L T D V L D D A L T R K L E G L V 380
GATGTTATGGTTGAATATCCCAATAGAAATAGTATTTCTCAGAGTTGACGAATATTAAGTAAACGAGTCTCAGTTAACTGACGTATTTGGATGATGCACTTAACGAGAAATTAGAAGATTTGGTC 1200
L K D I N G V Y E P G K R R W L K I K R D Y L N E G S M A D S A D L V L G A Y 420
TTAAAGAAATTAATGGAGTATACGAACCGGAAAGAGAAGATGGTTAAAAAATAAAGCAGAGACTATTTGAAACGAGGGTTCCATGGCAGATTTCTCCGATTTAGTAGTACTAGTGTCTTAC 1320
Y G K G A K G I M A V F L M G C Y D D E S G K W K T V T K C S G H D D N T L R 460
TATGTGTAAGAGCAAGGGTGGTATCATCGCAGTCTTTCTAAATGGGTTGTTTACGACATGAATCCCGTAAATGGAAGACGGTTACCAAGTGTTCAGACACGATGATATACGTTAAGG 1440
V L Q D Q L T M V K I N K D P K K I P E W L V V N K I Y I P D F V V E D P K Q S 500
GTTTGTGCAAGACCAATTAACGATGGTTAAATAAACAAGATCCCAAAAAATAATCCAGAGTGGTTGGTAGTTTAATAAAAAATCTATATCCCGATTTTGTAGTAGAGTATCCGAACAATCT 1560
Q I W E I S G A E F T S S K S H T A N G I S I R F P R F T R I R E D K T W K E S 540
CAGATATGGGAAATTTTCAGGACGAGGTTTACATCTTTCCCAAGTCCCATACCGCAAAATGGAAATATCCCATTAGATTTCTCTAGATTTACTAGGATTTAGAGAATAAAACGTGGAAAGATCT 1680
T H L N D L V N L T K S * * 552
ACTCATCTAAACGATTTAGTAAATCTTTAGCTAAATCTTAATAGTTTACATACAACTGAAAAATTAATAAATAACACTATTATTAGTTTGGTGTGCTCCCATGGA 1776

160 kDa protein found in type A inclusion bodies contains TTTTTTTAT separated from TAAATG by AA (48,49) and the vaccinia virus gene encoding an 11 kD core basic protein has the sequence TTTTTTCT six nucleotides upstream of the TAAATG motif (50). These sequences suggested that the SalF 15R gene might be transcribed late during infection.

Transcriptional analyses

Virus mRNA produced during the early or late phase of infection was analysed first by Northern blotting with a single stranded DNA probe derived from entirely within the coding sequence of SalF 15R. The probe was complementary to the nucleotide sequence shown in Figure 2 and the result is shown in Figure 3A. Early during infection a major transcript of approximately 3000 nucleotides and a minor transcript of approximately 4000 nucleotides are produced (lane 1). These correspond to the sizes of transcripts initiating close to the beginning of SalF 15R and terminating at the T₅NT motifs 2730 or 3790 nucleotides downstream, allowing for addition of poly A tails of up to 200 nucleotides. Surprisingly, there is no T₅NT motif near the end of SalF 15R so that the mRNAs include SalF 16R and SalF 17R coding sequences and are therefore potentially bi- or tricistronic. Late during infection there is a very small amount of the 3000 nucleotide transcript and also a faint smear of RNA between 2–4,000 nucleotides (lane 2). The discrete size of the 3 kb transcript suggests that it is residual early RNA since late vaccinia mRNAs are heterogeneous in size (1).

The S1 nuclease technique was used to precisely locate the 5' ends of the mRNAs (51) (Figure 3B). A 365 nucleotide ClaI-BclI fragment labelled with ³²P at the 5' end of the BclI site was used as probe (see Figure 1). The DNA fragments protected from S1 nuclease by hybridization to early (lane 2), late (lane 3) or mock-infected (lane 4) RNA are shown in Figure 3B. The sequencing ladder (lanes TCGA) is from an unrelated M13 clone extended from the universal primer with Klenow enzyme and isotopically labelled with [³⁵S]-dATP. Early RNA protected a major cluster of DNA fragments 113–122 nucleotides long with the strongest signal at 118 and 119. There are other minor clusters of protected fragments 38–39, 74–77 and 89–91 nucleotides long visible on longer exposure. Collectively all these minor bands represent only a very small fraction of total protected DNA. The 5'-labelled nucleotide of the probe is 119 nucleotides downstream of the A nucleotide in the translational initiation codon and, therefore, the major 5' end of the early mRNA maps to within, or just downstream of, the TAAATG motif. The positions of the RNA start sites are marked in Figure 2. The S1 nuclease technique can underestimate the length of RNA by a few nucleotides so it is likely that the majority of the early transcripts initiate within the TAAATG motif. Late RNA contains a small amount of transcripts initiating at the same position but it is uncertain if this reflects true late transcription or a small residual amount of early RNA that was also noted in Northern analysis.

The presence of early RNA initiating at, or close to, a potential late transcription initiation site raised the possibility that these RNAs contain 5' poly A sequences of up to 35 nucleotides, a feature of late vaccinia mRNAs (52,53). This was addressed by primer

Figure 2 Nucleotide sequence of a 1776 nucleotide region of the 13.4 kb SalI F fragment. The deduced sequence of a 552 amino acid open reading frame is shown. The positions of potential transcription control signals are underlined. The major and minor 5' ends of early mRNAs mapped by S1 nuclease protection are indicated by open triangles or arrow heads, respectively. Exclamation marks indicate the positions of the 5' ends of early mRNAs mapped by primer extension. Nucleotides complementary to the oligonucleotide used for primer extension are marked with an asterisk. Numbers indicate either the nucleotide number, starting 60 nucleotides upstream of the open reading frame, or amino acid number starting from the first methionine.

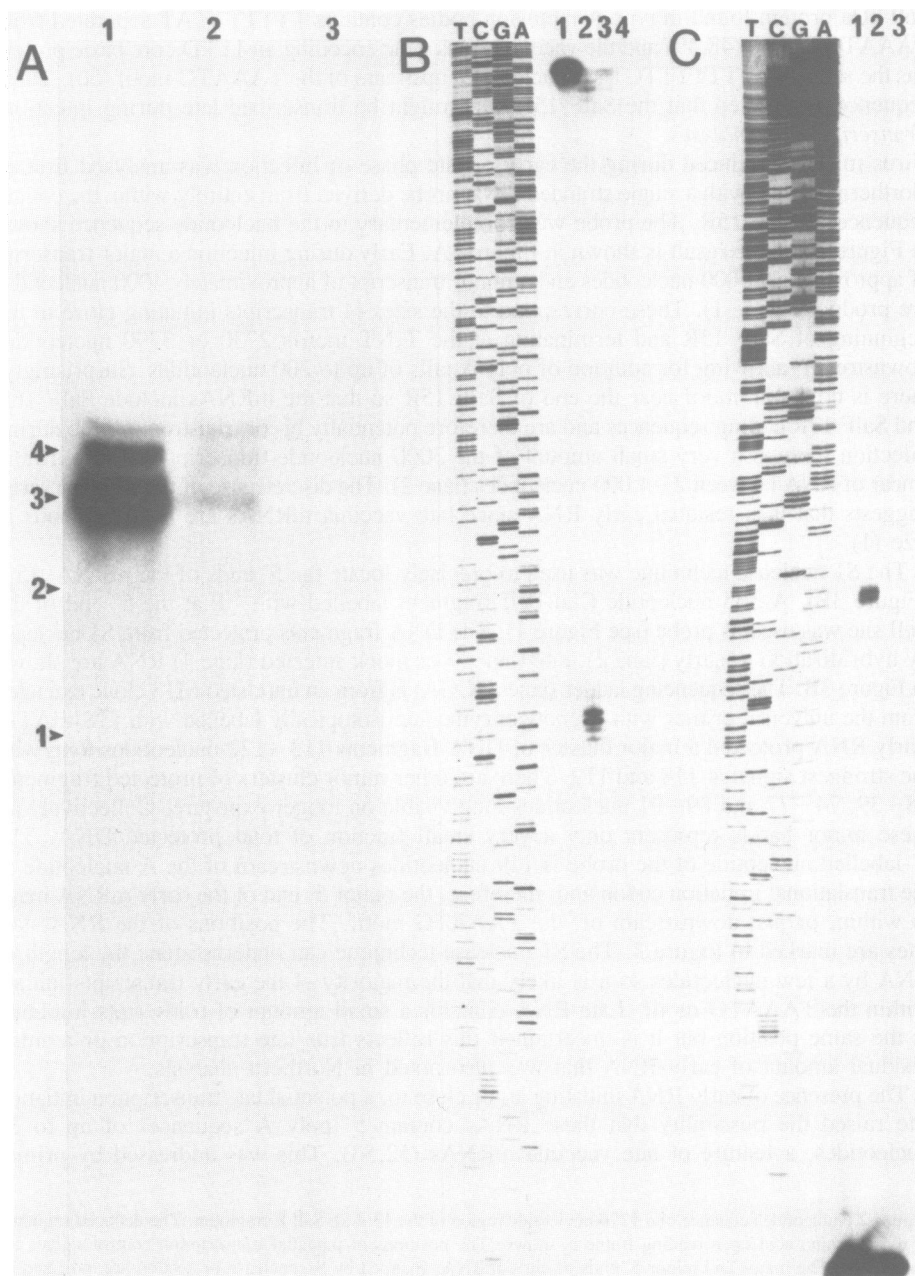


Figure 3 Transcriptional analyses. A. Northern analysis of virus mRNA. RNA extracted 6 hrs after infection in the presence of cycloheximide (early, lane 1) or 10 hrs after infection without drug (late, lane 2) or from uninfected cells (lane 3) was resolved by electrophoresis through denaturing agarose gels, transferred by blotting

extension analysis using an oligonucleotide complementary to nucleotides 119–139 (inclusive) downstream of the A of the translation initiation codon. The position of this oligonucleotide is indicated in Figure 2 and the results of this analysis are shown in Figure 3C. With early RNA (lane 2), primer extended products of 141–146 nucleotides are produced. These map at, or just upstream of, the TAAATG motif and the S1 nuclease start site. The tight clustering of the primer extended products at a position approximately 5 nucleotides upstream of the 5' ends as determined by S1 nuclease protection, indicated that the early mRNAs do not contain 5' poly A sequences of heterogeneous length. The slight difference in length of the primer extended products versus the S1 nuclease protected fragments may be attributable to S1 digestion of a few terminal nucleotides. Alternatively, the true 5' end of the mRNA may contain a few (4 or 5) A residues between the cap structure and the AUG. No extended product was produced by reverse transcription of late RNA (lane 3). This experiment was performed with a different preparation of late RNA than used in Northern and S1 experiments and the absence of any extended product supports the conclusion that a small amount of early virus mRNA remained in the late RNA preparation used for Northern and S1 analyses.

Analysis of the open reading frame

Comparison of the amino acid sequence of SalF 15R with our database of vaccinia virus proteins using programme FASTP (41) found no strong matches. However, a search of the protein database SWISSPROT revealed extensive homology to DNA ligase of *Saccharomyces cerevisiae* (31). An optimised FASTP score of 527 was obtained (KTUP of 1) and the two proteins had 30% amino acid identity over a 412 amino acid region. A similar degree of homology exists between SalF 15R and *Saccharomyces pombe* DNA ligase (32) although fission yeast *S. pombe* and the budding yeast *S. cerevisiae* are evolutionarily divergent. Only weak homology was detected with bacteriophage T4 and T7 and *E. coli* DNA ligases. An alignment of the amino acid sequences of DNA ligases from yeasts and vaccinia virus is shown in Figure 4. This alignment shows that the amino-terminal region of the vaccinia protein is divergent from both yeast sequences and there are regions which are absent in vaccinia but present in both yeasts. The latter point is reflected in the predicted sizes of the proteins, with vaccinia DNA ligase (63.3 kD) being considerably smaller than DNA ligases of *S. pombe* (86.2 kD) and *S. cerevisiae* (84.8 kD). The yeast DNA ligases are also least conserved in the amino terminal region. In contrast, in the carboxy-terminal region the three sequences are almost colinear and have extensive amino acid identity and conservative changes. The presumed catalytic lysine at the ATP binding site (marked with asterisk) is conserved in all these sequences and as

to a nitrocellulose membrane and probed with a [³²P]-labelled, prime-cut, single stranded DNA fragment complementary in sequence to an internal region of the DNA ligase gene shown in Figure 2. An autoradiograph is shown. The positions of double stranded DNA molecular weight markers are indicated in kilobases. B. S1 nuclease protection mapping of 5' ends of mRNA. A 365 bp ClaI-BclI fragment, [³²P]-labelled at the 5' end of the BclI site, was hybridized with early (lane 2) or late (lane 3) vaccinia virus mRNA or mock infected RNA (lane 4). After digestion with S1 nuclease, DNA fragments were resolved on a 6% buffer gradient TBE sequencing gel. Lanes TCGA represent a sequencing ladder from an M13 clone containing vaccinia virus DNA and lane 1 contains undigested DNA probe. The length of the protected fragments were determined by counting the number of nucleotides in the parallel sequencing ladder starting from the 5' end of the universal sequencing primer. C. Primer extension. A 21 base oligonucleotide was 5' labelled with γ -[³²P]-ATP and polynucleotide kinase and hybridized to early (lane 2) or late (lane 3) virus mRNA. Extension from the oligonucleotide was performed with reverse transcriptase and products were resolved on a 6% buffer gradient sequencing gel. An M13 clone containing vaccinia virus DNA sequenced from universal primer was used as the size ladder (lanes TCGA). Lane 1, [³²P]-labelled oligonucleotide.


```

sc MRRLLTGCLSSSARPLKSLDPL MSSLSPSSACPKQKQATLARFTTSMKNKPTGTPSPKSSKHMLEEDNDVSVSGEEDATKKLKQTAVHTHTVAAP
sv RTVFSQIPRFKQVQKQVIRMTSQSDISFNFFSSAKSEHVEVSVSSSDSNKVDGRSTSSKAVESVKLITDSEKHNHDDGTQGVNERENNVISEAKKQ
vv MTSLRREFRRKCDIYHNSISYENSLIRDFITDIDDD

sc SSMGSNFFSSIPSSAPSSGVADSPQSSQVGEVEALSSNNNDHYSSNIPYSEVCFVNKEPAISSRLERTRICSPFIKIKRKOSKNVIPTTYLFINI
sv KTLGSSSSSSDAVSSNNDSGASTPIPEKPELFLSNARNDLKLGHTAFAMVKAFTKATKSKRIEIDIMTGTGILRDLSTLACLVSLINIL
vv LIILGLLGLIDRLIMNDKQIKLYLSIEKOSQEDMLQIL

sc SPDEAGLEGLGLGENLLMKTSPIGKGSMSCKLKYDIDGLDGEIAMGARNVQPTMFKKPKPLTGEVFKIRAIRATGOGKDSQLKKKIKRMLTACKGI
sv PDVSGLELIGESIMAKKSGTGTLQQLKSLRKVGDLGLVAGTSSQNPQTFMKPAALTIPFLPDSKIKKIAMSGSQNSQNRKQVYIKRLLSISCEGA
vv LGDILRTFENETELSPRDSILTLEDSFTLISVVAESHQKILDTIASGTM

sc EKAFILRSLESHDGLAEKTVIISIKALLHDHENRSDPKPVMVDVLESAQQIKRIRCCQVRYEIVNSCLEHGINLNDKYCTLRPGELKFMPLAK
sv EPKYILRAGELHGLQAEKTVIMANATAYHADKNGEKLQQDRIEGEQLRQVCCOLPSYDILVPHLEHGGTILRETCKLTPGTGTFMPLAK
vv DGLVMLDKIDIKAGRPYVINSFNAYDVERKSNKILENASKQNSDISISVMTEINMLDE

sc PTKAINEVLDRHGEETSSKNYDGEROVHLLNOGTMRIYSRNGENMTERYPEINITDFIQDLDTTKNLILLQSAVAWDKDGKLPQVOSTFRKRDV
sv QPTKISEVLNLDQAFTKQKNYDGEROVHETDGGVYFSSRNMSYNSVYISVSWKKKPKPLISLIDQAVGVDWDLPLPKONLATFRKRDV
vv SCDSVKAFKFSPSGMAEYKVDGEROVHNNHNAFFSRNMPKVLSHKVLKYLKPEKAPKPLISLISLVLDVDEHVLPLPGSGIHKH

sc ELNDVKVKKLEAFDMLCYNDERLINKSMERPEVITKTKVVEGEFOYATQITTNLDELQKELFESVNHSGEGLMMMLGCESHSYEFSKGRNWLML
sv KNGIKVYRALEAFDMLMLLCPLEETIPERKSHIYSMEQPSGTDFAKHSDSGSIEESTIEFESVSKDSEGIMLVMLGDSHSEYFSKGRNHLV
vv ISNMLEAFDMLYDFDMDTIDIPERKSHIYSMEQPSGTDFAKHSDSGSIEESTIEFESVSKDSEGIMLVMLGDSHSEYFSKGRNHLV

sc KDYLEGVGQSDIDILGAYYGKRTTYGGFGLCYNDDEFEPTCKRIGTFSEMDQLHDLSTTTTGPKATFVFDSSAEPDVVFEPTTL
sv KDYLEGVGQSDIDILGAYYGKRTSVYGAFFGLCYDPTDTYVCSIQGLTFSEEHETFTVQDKDIVSKKDFYAHSDVAHQPDVVFEFPL
vv KDYLENGGMSDIDVLVLAGAYYGKAKGIMAFIRECYDDESKWKNVCSHNDITRIVQDLNVRNADPKPIPELVNKKIYIPDEVFD

sc FEVLTDILSLPIYKAGSATFDKQSLRFPFRFIREDKGVGVEDSSDQVHNYEQSHMN
sv WEVLAADILSPYKAGIYQVEDKQSLRFPFRFIREDKVEDATISGEVSEFYRYQVAYSQKEKGSFAAEY
vv KQSGOIEWISGAFTSKSHTANGISURFPFRFIREDKVETKSHINDIMTKS

```

Figure 4 Alignment of the amino acid sequences of DNA ligases from vaccinia virus (vv), *S. pombe* (sp) and *S. cerevisiae* (sc) made using programme MULTALIGN (42). The position of the conserved catalytic lysine from the ATP binding site is marked by an asterisk. Amino acids which are identical in vaccinia virus and either or both yeast sequences are boxed.

well as in T4 and T7 DNA ligases. In the *E. coli* enzyme, which uses NAD rather than ATP as cofactor, this site is less conserved. The most highly conserved region is very close to the carboxy terminus and is rich in basic amino acids. Over a 16 amino acid region the vaccinia protein shares identity with *S. pombe* at 15 positions with *S. cerevisiae* at 14 positions with a conservative isoleucine to valine change at one of the two divergent amino acids. This region is also well conserved in T4 with 6 identical residues and several conservative changes. The high conservation of this region suggests it plays some critical role in DNA ligase function, and its basic composition is consistent with an interaction with the DNA substrate. Comparisons of the vaccinia virus protein with prokaryotic DNA ligases using DIAGON (date not shown) showed very little similarity with only short conserved domains. This is not surprising since the prokaryotic DNA ligases are themselves divergent with little homology.

DISCUSSION

The nucleotide sequence of a vaccinia virus gene which maps in the SalI F fragment and which potentially encodes a polypeptide of 63.3 kD is presented. The deduced amino acid sequence of the open reading frame has extensive similarity to DNA ligases of *S. pombe* and *S. cerevisiae* but poor similarity to prokaryotic DNA ligases outside a few highly conserved domains. This provides direct evidence that vaccinia virus encodes a DNA ligase and supports early data (26) showing a 13-fold increase in DNA ligase activity in the cytoplasm of vaccinia virus-infected cells. Spadari (1976) concluded that the increase in DNA ligase activity was probably not virus-encoded since the enzyme had similar biochemical characteristics to cellular DNA ligase I, but may be attributable to enhanced leakage of the nuclear enzyme into the cytoplasm of virus infected cells. The high structural relatedness of the predicted vaccinia virus protein to known DNA ligases is alone probably

sufficient to conclude that the protein is a DNA ligase, but this conclusion is supported by the observed increase in ligase activity in virus-infected cells and our own biochemical data demonstrating that the encoded protein possesses DNA ligase activity (Kerr and Smith, submitted for publication).

The amino acid sequence of this vaccinia enzyme is the first reported primary structure of a 'mammalian' DNA ligase. It is also the only example of a eukaryotic virus encoding a DNA ligase, although other large DNA viruses which replicate in the cytoplasm, such as African Swine Fever Virus, probably encode this enzyme. Although much is known of mammalian DNA ligases the genes encoding these enzymes have not been mapped. The search for these genes has received extra importance with the demonstration that the genetically inheritable disease, Bloom's syndrome, is associated with a defect in DNA ligase I resulting in enzyme thermolability and decreased activity (54,55,34). This search may be aided by the nucleotide and deduced amino acid sequence of the vaccinia DNA ligase. The highly conserved amino acid domains in the carboxy-terminal region may be used to design synthetic oligopeptides which could be used to raise antisera to screen expression vector libraries containing mammalian cDNAs. Alternatively, mixtures of synthetic oligonucleotides representing the sequences from these domains could be used in polymerase chain reactions to amplify and subsequently clone the ligase gene(s).

Transcriptional analyses demonstrate that the vaccinia DNA ligase gene is transcribed early during infection. This finding was contradictory to what may have been predicted from inspection of the nucleotide sequence proximal to the translational start site. At this position there is a sequence TAAATG characteristic of transcriptional start sites of late vaccinia genes and is immediately preceded by two overlapping early transcriptional termination signals (TTTTTTTAT). Perversely, the sequence T₇NT is also present just upstream of the TAAATG motif of two of the strongest late poxvirus promoters known (46,48,49). The function of these early transcriptional termination signals at this position is unclear, but it is unlikely to be the prevention of early transcripts running through the ligase promoter from upstream and thereby interfering with transcription initiating from the ligase promoter, since transcription termination induced by this signal occurs approximately 50 nucleotides downstream. Surprisingly, there are no early transcription termination signals near the 3' end of the DNA ligase gene. Instead the early mRNAs terminate after the first or second gene downstream (SalF 16R and SalF 17R) which are transcribed in the same direction. Consequently, the DNA ligase mRNAs are potentially polycistronic.

Mapping of the 5' end of transcripts from the DNA ligase gene by S1 nuclease protection showed a major cluster of 5' ends at, or just downstream of, the TAAATG motif, with the strongest signals at the T and G nucleotides of the ATG codon. There are also three other very minor start sites downstream within the first 80 nucleotides of the open reading frame. The abundance of these minor start sites was even lower when analysed by primer extension and are not discernible on the autoradiograph shown but only after prolonged exposure. The significance of these minor, smaller transcripts is unclear but translation from them would give rise to a protein lacking the amino terminal 54 amino acids. Primer extension confirmed that the 5' end of >95% of the mRNAs mapped to near the TAAATG motif but placed the major 5' end five nucleotides upstream of the position determined by S1 nuclease protection. This is possibly due to nibbling of the terminal nucleotides by S1 nuclease. Alternatively, it could be due to extra A nucleotides on the 5' end of mRNA between the cap structure and AUG codon and might explain how the 5' terminal

nucleotide, predicted to be predominantly T from the primer extension result, could be A (or G) as shown from analysis of the cap structure of total early virus mRNA (56). The TAAATG motif is therefore not exclusively a start site for late vaccinia mRNAs. A similar result was obtained with gene 5 in the HindIII D fragment which contains this motif at this site of initiation of transcription early during infection (57).

DNA ligase can now be added to the list of enzymes known to be encoded by vaccinia virus. Other virus encoded enzymes involved in DNA metabolism are DNA polymerase, topoisomerase, ribonucleotide reductase and thymidine kinase. In addition a single strand-specific nuclease and nicking-joining enzyme are packaged in the virion and likely to be virus-encoded. Probably this cytoplasmically replicating virus contains a complete DNA replication system that may function independently of host enzymes. In view of this the existence of a virus-coded DNA ligase might be predicted. However, vaccinia contains two other enzymes with DNA strand sealing activity (topoisomerase and nicking-joining enzyme) and models for virus DNA replication have been proposed which do not require a conventional DNA ligase (58,59). In one model the linear double stranded DNA genome with covalently closed hairpin ends is nicked on one strand near one, or both, terminal hairpins to provide a 3' OH from which polymerisation may initiate. Elongation proceeds around the terminal hairpin, down the linear genome and around the opposite hairpin to produce concatemeric DNA molecules by a strand displacement mechanism. This model does not require, but may use, lagging strand synthesis. An alternative model requires the use of short primers, lagging strand synthesis and a DNA ligase (60). Resolution of concatemeric junctions into closed hairpins can be accomplished by nicking-joining enzyme (so named to distinguish it from topoisomerase I, nicking-closing) (14). While the existence of a virus DNA ligase does not establish either mechanism it is consistent with lagging strand synthesis for which some evidence has been presented (61,62). An alternative, or additional, function of the virus DNA ligase could be in DNA recombination. In bacteriophage T4 replication, the processes of DNA replication and recombination are dependent upon each other and DNA ligase is essential for both activities (63). Similarly, vaccinia virus DNA replication and recombination are linked and no conditional lethal mutants have been described which are defective in only one of these processes. No conditional lethal mutant has been assigned to the DNA ligase gene. An exciting possible explanation is that a temperature sensitive DNA ligase may result in a recombination defective phenotype and therefore marker rescue at the nonpermissive temperature would not occur. The isolation of such a mutant or the construction of a virus expressing DNA ligase conditionally from an inducible promoter (64, Rodriguez and Smith, in preparation) will be valuable tools in determining if DNA ligase is essential and in elucidating its role in DNA replication and recombination.

ACKNOWLEDGEMENTS

We thank Helen Newman and Henrietta Mountain for typing and Paco Rodriguez for critical reading of the manuscript. This work was supported by MRC grant G8763796CA and GLS is a Lister Institute-Jenner Research Fellow.

*To whom correspondence should be addressed at Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

REFERENCES

1. Moss, B. (1985) In Fields, B.N. (ed), *Virology*. Raven Press, New York. pp 658–703.
2. Kates, J.R. & McAuslan, B. (1967) *Proc. Natl. Acad. Sci. U.S.A.* **58**, 134–141.
3. Munyon, W.E., Paoletti, E. & Grace, J.T. (1967) *Proc. Natl. Acad. Sci. U.S.A.* **58**, 2280–2288.
4. Broyles, S.S. & Moss, B. (1986) *Proc. Natl. Acad. Sci. USA* **83** 3141–3145.
5. Jones, E.V., Puckett, C. & Moss, B. (1987) *J. Virol* **61**, 1765–1771.
6. Patel, D.D. & Pickup, D.J. (1989) *J. Virol.* **63**, 1076–1086.
7. Martin, S.A., Paoletti, E. & Moss, B. (1975) *J. Biol. Chem.* **250**, 9322–9329.
8. Moss, B. Rosenblum, E.N. & Gershowitz, A. (1975) *J. Biol. Chem.* **250**, 143–156.
9. Morgan, J.R., Cohen, L.K. & Roberts, B.E. (1984) *J. Virol.* **52**, 206–214.
10. Niles, E.G., Condit, R.C., Caro, P., Davidson, K., Matusick, L. & Seto, J. (1986) *Virology*, **153**, 96–112.
11. Schuman, S. & Moss, B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7478–7482.
12. Shaffer, R. & Traktman, P. (1987) *J. Biol. Chem.* **262**, 9309–9315.
13. Lakritz, N., Fogelson, P.D., Reddy, M., Baum, S., Hurwitz, J. & Bauer, W.R. (1985) *J. Virol.* **53**, 935–943.
14. Reddy, M.K. & Bauer, W.R. (1989) *J. Biol. Chem.* **264**, 443–449.
15. Pogo, B.G.T. & Dales, S. (1969) *Proc. Natl. Acad. Sci. USA* **63**, 820–827.
16. Challberg, M.D. & Englund, P.T. (1979) *J. Biol. Chem.* **254**, 7812–7819.
17. Traktman, P., Sridar, P., Condit, R.C. & Roberts, B.E. (1984) *J. Virol.* **49**, 125–131.
18. Jones, E.V. & Moss, B. (1984) *J. Virol.* **49**, 72–77.
19. Earl, P.L., Jones, E.V. & Moss, B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3659–3663.
20. Tengelsen, L.A., Slabaugh, M.B., Bibler, J.K. & Hurby, D.E. (1988) *Virology*. **164**, 121–131.
21. Schmitt, J.F.C. & Stunnenberg, H.G. (1988) *J. Virol.* **62**, 1889–1897.
22. Slabaugh, M., Roseman, N., Davis, R. & Matthews, C. (1988) *J. Virol.* **62**, 519–527.
23. Weir, J.P., Bajszar, G. & Moss, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1210–1214.
24. Hruby, D.E. & Ball, L.A. (1982) *J. Virol.* **43**, 403–409.
25. Weir, J.P. & Moss, B. (1983) *J. Virol.* **46**, 530–537.
26. Sambrook, J. & Shatkin, A.J. (1969) *J. Virol.* **4**, 719–726.
27. Spadari, S. (1976) *Nucleic Acids Res.* **3**, 2155–2167.
28. Ishino, Y., Shinagawa, H., Makino, K., Tsunasawa, S., Sakiyama, F. & Nakata, A. (1986) *Mol. Gen. Genet.* **204**, 1–7.
29. Armstrong, J., Brown, R.S. & Tsugita, A. (1983) *Nucleic Acids Res.* **11**, 7145–7156.
30. Dunn, J.J. & Studier, F.W. (1981) *J. Mol. Biol.* **148**, 303–330.
31. Barker, D.G., White, J.H.M. & Johnson, L.H. (1985) *Nucleic Acids Res.* **13**, 8323–8337.
32. Barker, D.G., White, J.H.M., Johnston, L.H. (1987) *Eur. J. Biochem.* **162**, 659–667.
33. Engler, M.J. & Richardson, C.C. (1982) In Boyer, P.D. (Ed.) *The Enzymes*. Academic Press Inc. Vol. XV, pp 3–29.
34. Lindahl, T., Willis, A.E., Lasko, D.D. & Tomkinson, A. (1989) In Lambert, M. and Laval, J. (Eds.) *DNA Repair Mechanisms and their Biological Implications in Mammalian Cells*. Plenum Press, In press.
35. Baldick, C.J. & Moss, B. (1987) *Virology* **156**, 138–145.
36. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
37. Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3693–3695.
38. Bankier, A.T., Western, K.M. & Barrell, B.G. (1987) in Wu R. (Ed.) *Methods in Enzymology* **155**, 51–93. Academic Press, London.
39. Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694.
40. Staden, R. (1982) *Nucleic Acids Res.* **10**, 4731–4751.
41. Lipman, D.J. & Pearson, W.R. (1985) *Science* **227**, 1435–1441.
42. Barton, G.J. & Sternberg, M.J.E. (1987) *J. Mol. Biol.* **198**, 327–337.
43. Smith, G.L., Howard, S.T. & Chan, Y.C. (1989) *J. gen. Virol.* **70**, In press.
44. Ausubel, I. & Frederick, M. (1987) *Current Protocols in Molecular Biology*. Green Publishing Associates & Wiley- Interscience, New York.
45. Rosel, J.L., Earl, P.L., Weir, J.P. & Moss, B. (1986) *J. Virol.* **60**, 436–449.
46. Hanggi, M., Bannwarth, W. & Stunnenberg, H.G. (1986) *EMBO J.* **5**, 1071–1076.
47. Yuen, L. & Moss, B. (1987) *Proc. natl. Acad. Sci. USA* **84**, 6417–6421.
48. Patel, D.D. & Pickup, D.J. (1987) *EMBO J.* **6**, 3787–3794.
49. Funahashi, S., Sato, T. & Shida, H. (1988) *J. gen. Virol.* **69**, 35–47.

50. Bertholet, C., Drillien, R. & Wittek, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2096–2100.
51. Berk, A.J. & Sharp, P.A. (1977) *Cell* **12**, 721–732.
52. Bertholet, C., Van-Meir, E., ten Heggeler-Bordier, B. & Wittek, R. (1987) *Cell* **50**, 153–162.
53. Schwer, B., Visca, P., Vos, J.C. & Stunnenberg, H.G. (1987) *Cell* **50**, 163–169.
54. Willis, A.E., Lindhal, T. (1987) *Nature* **325**, 355–357.
55. Chan, J.Y.H., Becker, F.F., German, J. & Ray, J.H. (1987) *Nature* **325**, 357–359.
56. Boone, R.F. & Moss, B. (1977) *Virology* **79**, 67–80.
57. Lee-Chen, G.-J., Bourgeois, N., Davidson, K., Condit, R.C. & Niles E.G. (1988) *Virology* **163**, 64–79.
58. Moyer, R.W. & Graves, R.L. (1981) *Cell* **27**, 391–401.
59. Baroudy, B.M., Venkatesan, S. & Moss, B. (1982) *Cell* **28**, 315–324.
60. Baroudy, B.M., Venkatesan, S. & Moss, B. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 723–729.
61. Esteban, M., Flores, L. & Holowczak, J. (1977) *Virology* **83**, 467–473.
62. Pogo, B.G.T. & O'Shea, M.T. (1978) *Virology* **84**, 1–8.
63. Mosig, G. (1987) *Ann. Rev. Genet.* **21**, 347–371.
64. Fuerst, T.R., Fernandez, M.P. & Moss, B. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2549–2553.

**This article, submitted on disc, has been automatically
converted into this typeset format by the publisher.**