

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/21109807>

# Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near the right inverted terminal repeat

Article in *Journal of General Virology* · July 1991

DOI: 10.1099/0022-1317-72-6-1349 · Source: PubMed

CITATIONS

170

READS

327

3 authors:



**Geoffrey L. Smith**

University of Oxford

461 PUBLICATIONS 31,476 CITATIONS

[SEE PROFILE](#)



**Yuk-Sang Chan**

University of Cambridge

56 PUBLICATIONS 1,837 CITATIONS

[SEE PROFILE](#)



**Susan T. Howard**

University of Texas Health Science Center at Tyler

27 PUBLICATIONS 2,315 CITATIONS

[SEE PROFILE](#)

## Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near the right inverted terminal repeat

Geoffrey L. Smith,<sup>1</sup>\* Y. Sang Chan<sup>2</sup> and Susan T. Howard<sup>1</sup>

<sup>1</sup> Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE and <sup>2</sup> Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, U.K.

The nucleotide sequence of 42 090 bp of vaccinia virus strain WR is presented. The sequence includes the *Sal*I L, F, G and I fragments and starts near the centre of the *Hind*III A fragment and extends rightwards towards the genomic terminus, finishing approximately 0.5 kb internal of the inverted terminal repeat (ITR). Translation of this region has identified 65 open reading frames (ORFs) of greater than 65 amino acids in length. Fifty-one of these which do not extensively overlap other larger ORFs have been subjected to further analysis; the other 14 are termed minor ORFs. In the rightmost 28.7 kb, the genes are, with one exception, transcribed towards the genomic terminus, similar to the arrangement of genes at the left end of the virus genome. Internal of this region the genes are expressed off either DNA strand but still predominantly rightwards. ORFs are tightly packed with few intergenic non-coding regions of greater than 250 bp. Protein sequence comparisons have established a remarkably high number of homologies with entries in existing protein databases. Of these, DNA ligase,

thymidylate kinase, two serine-threonine protein kinases, two serine proteinase inhibitors (serpins), two interleukin-1 receptor homologues and a discontinuous ORF related to tumour necrosis factor receptor have been reported. Other homologies include lectins, profilin, 3 $\beta$ -hydroxy steroid dehydrogenase, superoxide dismutase, guanylate kinase, ankyrin and complement factor H. In addition, there are a number of polypeptides with predicted properties of membrane-associated, secretory or glyco-proteins. Twelve gene families are described here and elsewhere. There is considerable similarity between genes from the right and left end of the virus genome that may have arisen by terminal transposition events. Several differences from the corresponding region of vaccinia virus strain Copenhagen sequence are noted. Near the right terminus the sequences diverge completely, and internal of this there are multiple examples of deletion of short sequences (eight to 10 nucleotides) that lie within penta- or hexanucleotide direct repeats.

### Introduction

Vaccinia virus, the live vaccine used to eradicate smallpox, is the prototypic orthopoxvirus (Moss, 1990). After the construction of restriction maps (Wittek *et al.*, 1977; Mackett & Archard, 1979) and the cloning of individual restriction fragments (Wittek *et al.*, 1980; Belle Isle *et al.*, 1981), the 186 kb genome of the laboratory strain of vaccinia virus (Western Reserve, WR) has been studied in detail. The structure of the terminal hairpins has been determined (Baroudy *et al.*, 1982) and the complete sequence of several of the viral *Hind*III restriction fragments and many individual genes have been reported from WR and other strains of virus (Ahn *et al.*, 1990; Amegadzie *et al.*, 1991; Baldick & Moss, 1987; Bournsnel *et al.*, 1988; Bertholet *et al.*, 1985; Broyles & Moss, 1986; Earl *et al.*, 1986; Gillard *et al.*, 1986; Gordon *et al.*, 1988; Hirt *et al.*, 1986; Howard &

Smith, 1989; Howard *et al.*, 1991; Jin *et al.*, 1989; Kotwal & Moss, 1988*a, b*, 1989; Niles *et al.*, 1986; Plucienniczak *et al.*, 1985; Rodriguez *et al.*, 1986; Rodriguez & Esteban, 1987; Rosel *et al.*, 1986; Rosel & Moss, 1985; Roseman & Hruby, 1987; Roseman & Slabaugh, 1990; Schmitt & Stunnenberg, 1989; Slabaugh & Roseman, 1989; Slabaugh *et al.*, 1988; Shida, 1986; Smith *et al.*, 1989*a, b, c*; Smith & Chan, 1991; Tamin *et al.*, 1988; Tengelsen *et al.*, 1988; Traktman *et al.*, 1989; Tsoa *et al.*, 1986; Ueda *et al.*, 1990; Van Meir & Wittek, 1988; Venkatesan *et al.*, 1981, 1982; Weinrich & Hruby, 1986; Weir & Moss, 1983, 1984). Recently the complete sequence of vaccinia virus strain Copenhagen was determined (Goebel *et al.*, 1990). Collectively these studies have demonstrated that the vaccinia virus genes are closely packed with little DNA between open reading frames (ORFs). Protein-coding regions in DNA are not interrupted by introns and there are no reports of

splicing. In the centre of the genome the coding regions are distributed on both DNA strands (Niles *et al.*, 1986; Rosel *et al.*, 1986; Plucienniczak *et al.*, 1985) but near the left end the genes are transcribed towards the genomic terminus (Kotwal & Moss, 1988*a*; Bournnell *et al.*, 1988). Similarly, the genes previously reported from the right end of the genome are transcribed towards the terminus (Kotwal & Moss, 1989; Smith *et al.*, 1989*a, b, c*, Howard & Smith, 1989; Traktman *et al.*, 1989; Rempel *et al.*, 1990; Ueda *et al.*, 1990; Goebel *et al.*, 1990).

At the outset of this work relatively little was known about the genes located near the right terminus in contrast to the detailed structural and genetic information on the left and central regions of the virus genome. In this report the sequence of 42 kb of the vaccinia virus genome (strain WR) from the centre of the *Hind*III A fragment to close to the right inverted terminal repeat (ITR) is presented together with the positions of the deduced ORFs and a description of protein homologies. The most surprising feature of this sequence is the relatively large number of predicted vaccinia virus proteins with homology to known protein sequences. The types of homology discovered may provide further insights into how the virus interacts with the host cell (e.g. profilin, ankyrin, lectins and a candidate sodium ion transport protein), interferes with the host immune response [serine protease inhibitors (serpins), complement factors, interleukin receptors, tumour necrosis factor receptor (TNFR)] and modifies the biochemical activity of the host cell [DNA ligase, thymidylate kinase (TpmK), guanylate kinase, serine-threonine protein kinases, 3 $\beta$ -hydroxy steroid dehydrogenase (3 $\beta$ -HSD) and superoxide dismutase]. Another notable finding has been the discovery of more families of related vaccinia virus genes, indicating gene duplication and diversification during the evolution of poxviruses. Lastly, a comparison of the WR and Copenhagen sequences (Goebel *et al.*, 1990) over the region reported here reveals several notable differences.

## Methods

**DNA sequencing.** The sequence of the *Sal*I, L, F, G and I restriction fragment of the WR strain of vaccinia virus was determined as previously described (Smith *et al.*, 1989*a*). Briefly, the individual fragments were randomly sheared by sonication and cloned into M13 vectors. Single-stranded DNA was sequenced using dideoxynucleoside triphosphates, [<sup>35</sup>S]dATP and the Klenow enzyme. The sequences at the junctions between these restriction fragments were determined using oligonucleotides as sequencing primers on restriction fragments spanning the *Sal*I sites.

**Computing.** Random DNA sequences were read using a sonic digitiser and assembled into contiguous sequence using computer programs DBUTIL and DBAUTO (Staden, 1982). ORFs were

identified using program ORFFILE, and files for individual protein sequences were created using DELIB (both kindly provided by M. E. G. Bournnell, Institute for Animal Health, Houghton, U.K.). Protein sequences were compared against the SWISSPROT protein database (version 14) using program FASTA (Pearson & Lipman, 1988). Multiple protein alignments were computed with program MULTALIGN (Barton & Sternberg, 1987). Sequence similarities were evaluated using the ALIGN program (Dayhoff *et al.*, 1983).

## Results

The region of the vaccinia virus genome (strain WR) sequenced in this study is indicated in Fig. 1. In Fig. 2 the nucleotide sequence (42090 nucleotides) is shown together with the predicted amino acid sequences of the major ORFs, which are defined as those starting with a methionine, that are at least 65 amino acids in length, and do not substantially overlap a larger ORF. These criteria were adopted to be consistent with that used for the complete sequence of the Copenhagen strain of vaccinia virus (Goebel *et al.*, 1990). In consequence, some of the former names of genes reported in this region are changed, and minor ORFs (defined as above, but which do overlap a larger ORF) are omitted from detailed analysis. These could subsequently be included as necessary. Using these criteria 51 major and 14 minor ORFs are identified (Fig. 2 and Table 1). The nomenclature adopted for vaccinia virus genes derives from the *Hind*III restriction fragment in which they lie, and the position and orientation of the ORF within that fragment (Rosel *et al.*, 1986). However, since the complete sequence of the *Hind*III A fragment is not available for this strain of virus, *Sal*I nomenclature has been adopted for the sequence to the left of the *Hind*III A-B junction. It is anticipated that these genes will be renamed with the final *Hind*III designations in due course. In Fig. 2 the positions of potential early and late transcriptional signals are marked (Yuen & Moss, 1987; Rosel *et al.*, 1986; Davison & Moss, 1989). Regions of the protein sequences that are potential hydrophobic transmembrane or signal sequences, or which are possible sites for *N*-linked glycosylation are also indicated. In Fig.

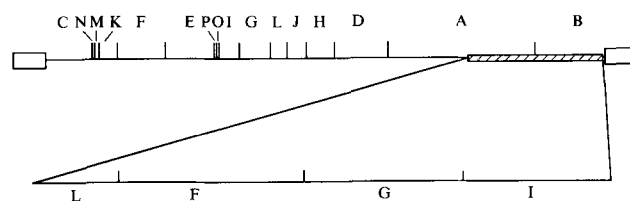


Fig. 1. *Hind*III restriction map of the 186 kb vaccinia virus genome (strain WR), top. The 10 kb ITRs are shown as open boxes. The region sequenced in this study is hatched and expanded to show the position of the *Sal*I L, F, G and I restriction fragments.

3 the arrangement of the ORFs and the principal homologies are summarized. The map positions, sizes and principal homologies of the ORFs are indicated in Table 1.

From Table 1 it is evident that there are only four examples of long (> 250 nucleotides) non-coding regions of which three are greater than 500 nucleotides. These are located after SalF16R (510 nucleotides), B3R (529 nucleotides) and B18R (744 nucleotides). Codon usage analyses and protein similarity searches suggest that these 'non-coding' regions may once have been part of adjacent vaccinia genes. SalF16R encodes a protein of 12·0K with homology to TNFR and nerve growth factor receptor (NGFR) (Howard *et al.*, 1991), proteins containing multiple cysteine-rich domains. SalF16R contains only one such domain, but downstream sequences show the presence of two more repeats either in a different reading frame or separated by a translational termination codon (Howard *et al.*, 1991). Vaccinia virus Copenhagen has the same arrangement as WR. A predecessor of vaccinia virus presumably encoded a larger SalF16R so making the downstream non-coding region much smaller. This is supported by the presence in another poxvirus, Shope fibroma virus (SFV), of an ORF of similar length to the total of SalF16R and the downstream fragments (Upton *et al.*, 1987).

Downstream of B3R there is an ORF in the -1 reading frame that lacks an initiating methionine. Codon usage analysis show that this region closely matches the codon usage of other vaccinia virus genes (data not shown), indicating that either it was originally part of B3R and lost the last 99 amino acids due to a frameshift mutation, or it was a separate gene that has lost its initiating methionine. The 744 nucleotides between B18R and B19R contain two ORFs without initiation codons, which also have codon usage similar to other vaccinia virus genes. Immediately after the translational termination codon of B18R there is an ORF of 161 amino acids that lacks an ATG until codon 109 and which has 29 tyrosine residues. This may have been part of B18R prior to the introduction of the termination codon. The second ORF represents a possible former 99 amino-terminal extension of B19R without an ATG codon.

#### Transcriptional arrangement

Fig. 2, Fig. 3 and Table 1 show that, with the exception of B16L, all 32 genes to the right of TmpK (SalF11R), are transcribed towards the end of the genome. This arrangement contrasts with the genes to the left of TmpK where five out of 19 genes are transcribed leftwards. Within the 42 kb analysed there are no situations in

which two adjacent genes are transcribed leftwards. In those positions where two adjacent genes are transcribed in opposite directions the intergenic region is predicted to function as either a bidirectional promoter (genes transcribed away from each other), or as a bidirectional transcriptional termination site if the genes are expressed towards each other early during infection. In the latter case there are usually multiple early transcriptional termination sites, TTTTNT (Yuen & Moss, 1987), on both DNA strands separating the genes. For example, in the intergenic gap between SalF2R and SalF3L, and SalF6R and SalF7L there are three such signals on one strand and two on the other. Similarly, between SalL7R and SalL8L there are four potential early transcriptional termination signals on one strand and one on the other. There are only two positions where early termination signals are not present following a leftward transcribed gene (SalL2L and B16L), but both genes may be transcribed late because there is a TAAAT motif at or near the 5' end (Rosel *et al.*, 1986; Davison & Moss, 1989) and in the case of SalL2L there is an early termination signal within the protein coding region. The opposing genes (SalL1R and B15R) may be expressed early.

Direct transcriptional mapping data are available for nine of the ORFs from this region and these data are summarized in Fig. 3. Genes known to be transcribed early during infection are DNA ligase (renamed SalF13R) (Smith *et al.*, 1989b), TmpK (renamed SalF11R) (Smith *et al.*, 1989c), serpins B13R and B22R (renamed) (Smith *et al.*, 1989a), candidate protein kinases B1R and B12R (Howard & Smith, 1989; Traktman *et al.*, 1989; Rempel *et al.*, 1990), B2R (Rempel *et al.*, 1990) and B18R (Ueda *et al.*, 1990). The only late gene described from this region is the haemagglutinin (HA) (SalG1R) (Shida, 1986). Although potential transcriptional control signals have been underlined in Fig. 2 and provide some indication of the likely transcriptional regulation of individual genes, direct experimental data are needed. The dangers in deducing the possible temporal regulation of transcription from nucleotide sequence data are illustrated with the DNA ligase gene. This contains TAAATG at the beginning of the ORF and no T<sub>5</sub>TNT near the 3' end, and would therefore have been classified as a likely late gene, yet this gene is transcribed early (Smith *et al.*, 1989b). Direct transcriptional data are also needed to indicate the regions of ORFs translated into protein. For instance, ORF SalF11R (TmpK) has the early RNA start site downstream of the computer-predicted beginning of the ORF (Smith *et al.*, 1989c).

#### Amino acid homologies

Table 1 summarizes the nucleotide positions of the ORFs

**Sall1R**→  
 S T A D I F N M A S I L N T L R F L E K T S F Y N C N D S I T K E K I K I K H K 40  
 GTCGACGGCGACATTTTAAACATGGCATCTATTTAAATACACTTAGGTTTTGGAAAAACATCATTTTATAATTGTAAACGATTCAATAACAAAGAAAGATTAAAGATTAAACATAA 120

G M S F V F Y K P K H S T V V K Y L S G G G I Y H D D L V V L G K V T I N N L K 80  
 GGGAAATGTCATTTGTATTTTATAAGCCAAAGCATCTACCGTTGTAAATACTTGTCTGGAGGAGGTATATATCATGATGATTGGTTGTATTTGGGAAGGTAAACAAATAATCTAAA 240

M M L F Y M D L S Y H G V T S S G A I Y K L G S S I D R L S L N R T I V T K V N 120  
 GATGATGCTATTTTACATGGATTATCATATCATGGAGTGACAAGTAGTGAGCAATTTACAATTTGGGATCGTCTATCGATAGACTTTCTCTAAATAGGACTATTGTTACAAAAGTTAA 360

N Y D D T F F D D D D \*  
 TAATTATGATACATTTTGGACGATGATGATCGTATGACACAATTTGTTTTTACTTTCTAATATAGCGTTAGATTCTTTTCATGTGCGAATATTGATTACTAAAAT 480  
 \* S S V N K S S S S Q D S N C L K T K K S E L I A N L N K K M H S Y Q N V L I 270

ATCGATGTTTAACTTTTGTCTATGACGCTCTTACAGCGGTATCGGTACATATACGTAATTCACCTTCACAAAATACGGAGTCTTCGATAAATAAGCCAATCGATTATTGGATCTAGC 600  
 D I N L K Q E I V D K D A T D T C I R L E G E C F V S D E I I I A L R N N S R A 232

TGTCGTATCATATTCAACATGTTTAAATATACCTTTCGTTTCCCTTTACAGGCATCGATCGTAGCATATTTTCCGCGTCTGAGATGGAAATGTTAAACTACAAAAATGCGTAATGTT 720  
 T Q I M N L M N L I D K R K G K V P M S R L M N E A D S I S I N F S C F H T I N 192

AGCCGCTCAATATGTCAGTGTCTATAAGTTTGGCATAGTAGAATAGACGTTGTTTAAATGCCCTCCAAAGTTTAAAGAAATCTATTAGAGTATTCATTTTGTAGTTTATCACC 840  
 A R G L I P V H R Y T Q C L I I S T N L H R G F N L F E I L T N C K S L K D G 152

TACATCATCAAAAAAAGTAAAGTGTGCTGATTTTATGATTTTGTGCGACAGCAATACATTTTCTATGTTACTTTTATGTCGTATCAGATTATATCTAGAGATTCTGACTACT 960  
 V D D F I L L F H A S K K H N Q A V A I C K E I N S K T R I L N Y E L S E Q S S 112

AACGAAATTAATATGATTGGCCAAATGTATCCATCATAATCTGGGTTATAAAGGGGTAAACAAGAATATATGTTTATATTTTAACTAGTGTAGAAAACAGAGATAGTAAATAGAT 1080  
 V F N I H N P W I Y G D Y D P N Y V P T F L F I H K Y K K V L T S F L S L L Y I 72

AGTTTTTCCAGATCCAGATCTCCCGTTAAACCATTTCAAACGGCATTTTAAATAAATTTTCTCTGAAAATGTTTTCCTTGGAAACAATTCATAATTATATTACAGTTACTAAAT 1200  
 T K G S G S G G T L V M R F P M K L L N E R S F Q K E Q F C N M 32

← **Sall2L** **Sall3R**→  
 M M T P E N D E E 9  
 AATTGTATAATAAATCAAAATATGAAAACATAAGGTCGTTAGTAGGGAGGAGAACAAAGAGGCACATCGTGACAATAAATACATTATTATCATGATGACACCAAGAAACGACGAAGAG 1320

Q T S V F S A T V Y G D K I O G K N K R K R V I G L C I R I S M V I S L L S M I 49  
 CAGACATCTGTGTTCTCGCTACTGTTTACGGAGACAAAATCAAGGAAAGAAATAACGCAACCGGTGATGGTCTATGTATTAGAATATCTATGTTTATTTCACTACTATCTATGATT 1440

T M S A F L I V R L N Q C M S A N E A A I T D A A V A V A A A S S T H R K V A S 89  
 ACCATGTCGCGGTTCTCATAGTGCCTAAATCAATGCATGCTGCTAACGAGGCTGCTATTACTGACGCCGCTGTTGCCGTTGCTGCTGCATCTACTCATAGAAAGGTTGCGTCT 1560

S T T Q Y D H K E S C N G L Y Y Q G S C Y I L H S D Y Q L F S D A K A N C T A E 129  
 AGCACTACACAATATGATCACAAGAAAGCTGTAATGGTTTATATTACAGGATCTTGTGATTATTAATACAAAGAAAGAACTACCGATTATCTCGGATGCTAAAGCAAAATGCACTGCCGAA 1680

S S T L P N K S D V L I T W L I D Y V E D T W G S D G N P I T K T T S D Y Q D S 169  
 TCATCAACACTACCAATAAATCCGATGCTTGTATTACCTGGCTCATTGATTATGTTGAGGATACATGGGGATCTGATGTTAATCCAATTACAAAAACTACATCCGATTATCAAGATTCT 1800

D V S Q E V R K Y F C V K T M N \* **Sall4R**→  
 M K S L N R O T V S R F K K L S 185  
 GATGTATCACAAGAAGTTAGAAAGTAATTTTGTGTTAAACAATGAACATAATTTTATTTTGTACATTAATAATGAAATCGCTTAATAGACAACTGTAAGTAGGTTTAAAGATTGT 1920

V P A A I M M I L S T I I S G I G T F L H Y K E E L M P S A C A N G W I Q Y D K 56  
 CGGTGCCGCGCGCTATATATACAAAGTCTGGAAGAACTGGTAAACAGATGTAAAGTACTCAATCTGACTATGTGTTAAAAATTTCTACAAGTGACACAAAAATGAATTAAT 2040

H C Y L D T N I K M S T D N A V Y O C R K L R A R L P R P D T R H L R V L F S I 96  
 AACATTGTTATTTAGATACTAACATTAATGTCTACAGATAATCGGTTTATCAGTGTCTGTAATTTACGAGCCAGATGCTTAGACCGGATACAGACATCTGAGAGTATTGTTAGTA 2160

F Y K D Y W V S L K K T N D K W L D I N N D K D I D I S K L T N F K Q L N S T T 136  
 TTTTATTAAGATTATTGGTAAAGTTTAAAGAAAGCAATGATAAATGGTTAGATATTAATAATGATAAAGATATAGATATTAGTAAATTAACAAATTTTAAACAACATAACAGTACGA 2280

D A E A C Y I Y K S G K L V K T V C K S T Q S V L C V K K F Y K \* 168  
 CGGATGCTGAAGCGTGTATATATACAAAGTCTGGAAGAACTGGTAAACAGATGTAAAGTACTCAATCTGACTATGTGTTAAAAATTTCTACAAGTGACACAAAAATGAATTAAT 2400

**Sall5R**→  
 M D A A F V I T P M G V L T I T D T L Y D D L D I S I M D F I G 32  
 AATAAGTCGTTAAGCTACGCCCATGGACGCCGCTTGTATTATCTCAATGCTGTTGACTATACACAGATACATGATGATGATCTCGATATCTCAATCATGGACTTTATAGGA 2520

P Y I I G N I K T V Q I D V R D I K Y S D M Q K C Y F S Y K G K I V P Q D S N D 72  
 CCATACATTATAGGTAAACATAAAATGTCCAAATAGATGTACGGGATATAAAATATCCGACATGCAAAAATGCTACTTTAGCTATAAGGGTAAAAATAGTTCTCTCAGGATTCTAATGAT 2640

L A R F N I Y S I C A A Y R S K N T I I I A C D Y D I M L D I E D K H Q P F Y L 112  
 TTGGCTAGATTCAACATTTATAGCATTTGTGCCGATACAGATCAAAAAATACCATCATATAGCATGCGCATGATATCATGTTAGATATAGAAGATAAACATCAGCCATTTTATCTA 2760

F P S I D V F N A T I I E A Y N L Y T A G D Y H L I I N P S D N L K M K L L F N 152  
 TTCCCATCTATTGATGTTTAAACGCTACATCATAGAAGCGTATAACCTGTATACAGCTGGAGATTATCATCTAATCATCAATCCTTCAGATAATCTGAAAATGAAATTGTTGTTAAT 2880

S S F C I S D G N G W I I I D G K C N S N F L S \* 176  
 TCTTCATCTGATATCAGACGGCAATGGATGGATCATAAATGATGGGAAATGCAATAGTAAATTTTATCATAAAGTTGTAAGTAAATAAATAAATAAATAAATTATTGAAGTACGATGAT 3000

**Sall6R**→  
 M M L V P L I T V T V V A G T I L V C Y I L Y I C R K K I R T V Y 33  
 ACGTATATTGAGCAATCAGAAATGATGCTGGTACCTCTTATCAGCGTGACCGTAGTTCGGGGAACAATATTAGTATGTTATATATTATATTATTGTTAGGAAAAAGATACGTAAGTCTAT 3120

N D N K I I M T K L K K I K S S N S S K S S K S T D S E S D W E D H C S A M E Q 73  
 AATGACATAAAATTTATCATGACAAATTAATAAAGATAGAGAGTTCTAATTCAGCAAAATCTAGTAAATCAACTGATAGCGAATCAGACTGGGAGGATCACTGTAGTGTATGGAACAA 3240

N N D V D N I S R N E I L D D D S F A G S L I W D N E S N V M A P S T E H I Y D 113  
 AACAAATGACGTAGATAATTTCTAGGAATGAGATATTGGACGATGATAGCTTCGCTGGTAGTTTAAATATGGGATAACCAATCAATGTCATGGCGCTAGCAGACAACACATTTACGAT 3360

S V A G S T L L I N N D R N E Q T I Y Q N T T V G I N E T E T V E V L N E D T K 153  
 AGTGTTCGGAAGCAGCTGCTAATAAATAATGATCGTAATGAACAGACTATTATTACAGAACACTACAGTGAATTAATGAGACGGAGACTGTGGAAGTACTTAAGTGAAGATACCAAA 3480

Q N P N Y S S N P F V N Y N K T S I C S K S N P F I T E L N N K F S E N N P F R 193  
 CAGAATCTAATATTATCAATCTTTCGTAATTAATAAACCAGTATTGTAGCAAGTCAATCCGTTTATTACAGAACTTAACAATAAATTTAGTGAGATAATCCGTTTATGA 3600

R A H S D D Y L N K Q E Q D H E H D D I E S S V V S L V \* 221  
 CGAGCACATAGCGATGATTATCTTAATAAGCAAGAAGATCATGAACACGATGATAGAAATCATCGGTCGTATCATTGGTGTGATTAGTTTCC TTTTTATAAAATGAAGTAATATT 3720  
**SalL7R→**  
 M E I F P V F G I S K I S N F I A N N D C R Y Y I D T E H Q 30  
 TAGTATTATTGCTGCCGTACCGTTGTACAAATGGAGATATCCCTGTATTCCGCATTTCTAAATTAGCAATTTTATTGCTAAATAGACTGTAGATATTATATAGATACAGAACATCAA 3840  
 K I I S D E I N R O M D E T V L L T N I L S V E V V N D N E M Y H L I P H R L S 70  
 AAAATTATATCTGATGAGATCAATAGACAGATGGATGAACGGTACTTCTTACCAACATCTTAAGCGTAGAAGTGTAAATGACAATGAGATGTACCATCTTATCCCCATAGCATTCG 3960  
 T I I L C I S S V G G C V I S I D N D I N D K N I L T F P I D H A V I I S P L S 110  
 ACTATTACTCTGTATTAGTTCTGCGGAGGATGTGTATCTCTATAGATAATGACATCAATGACAAAAATTTCTAACATTTCCTCATGTATGCTGTAAATCATATCCCCACTGAGT 4080  
 K C V V V S K G P T T I L V V K A D I P S K R L V T S F T N D I L Y V N N L S L 150  
 AAATGTGTCGTAGTTAGCAAGGTCCTACAAACATATTGGTTGTTAAAGCGGATATACCTAGCAACGATTGGTAACATCGTTTACAAACGACATACTATATGTAACAACTGTGTCACG 4200  
 I N Y L P L S V F I I R R V T D Y L D R H I C D Q I F A N N K W Y S I I T I D D 190  
 ATTAATTATTGCGGTGTGCTGTATTATTAGACGAGTCACCGACTATTTGGATAGACACATATGCGCATCAGATATTTCGGAATAATAAGTGGTATTCATTATAACCATCGACGAT 4320  
 K Q Y P I P S N C I G M S S A K Y I N S S I E Q D T L I H V C N L E H P F D L V 230  
 AAGCAATATCCTATTCCATCAAACTGTATAGGTATGCTCTGCGCAAGTACATAAATCTAGCATCGAGCAAGATACCTTAATACATGTTTGTAACTCGAGCATCCATTCGACTTAGTA 4440  
 Y K K M Q S Y N S V P I K E Q I L Y G R I D N I N M S I S I S V D \* 263  
TACAAAAAATGCAGTCGTACAATCTGTACCTATCAAGGAACAAATATTGTACGGTAGAATTGATAATATAAATAGCATTAGTATTCTGTGGATTAAATAGATTCTTAGTAGTGGG 4560  
 ATCATTAACTCATCTCTAATCTCTAAATACCTCATAAAAACGAAAAAATACTGCTACAAATGAAAAAAGCTATTATCAAACTGTACGGAATGGATTCTCTCTCTTTTTATGA 4680  
 AACTCTGTTGTATATCTACTGATAAACTGGAAGCAAAAACTCTGATAAAAAGAATAAGAATAAGATCAAGGATTATATGGAACACGATTATTATAAAAAAACAATAGTTCTCGTTCCT 4800  
 CTTCACGCTCTACTAGCTCGTGGTATTATACACATGCCTAGTAATAGTCTCTTTGCGTTSACGGAAGCAGACTAGAAATAACAGGCTAAATGTTTCAGACACCATAATAGTTCCCAACC 4920  
 \* Y Y D R Q T S P F C V L F L L S F H E S V M I T G L G 277  
 CAGATAAATACAGAGTACCATCAACACATTCCTTTAACTCAATCCCAACCAAAACCGTTAAATGTATCCGCCCAATTGATAGTAGATAATGAGGTGTACAGCGCATGATAATTTAC 5040  
S L L L L T G D V C E K L S L G L G L V T L I Y G A L Q Y Y I I L H V A C S L K C 250  
 ACAGTAACCAAAATGAAAATCTTTAGTAATTATAAGAAATATAGATGGTAACGTATCATCACAATCCAATAATATGCCGGAGAGTAAACATTGACGGATAAAACAAAAATGCTCCGC 5160  
 L L W F S F V K T I I L F I S P L T M M L L G I I H R L T F M S P Y F L F A G C 210  
 ATAACCTATCATCGGCAATAACACAACCAATCTTGTAAAGATTCTTAAATAGTAGAATAACACGGATATCGATGTATAAGTATCTCGAGAAATAAAGAATAAAGTAATGCCCG 5280  
L E I M A I V C G F Q L I G L N T S F V V S I S T Y T I E L F L T I G T 170  
 TAAAGATAAATCAACATTCATGTTGGTAATCATTAACCAATTAGTATGACGTTGAACATTTTACAGTAGATTTTATCCAGTGTATCTCGCATGTATACGTACCTGGTAAGATAT 5400  
F I F M L M T Q Y D N F W N T H R Q V L K V T S K I G T N D E C T Y T G P L I D 130  
 CTTTATATTCTAATCAATGAGACATCACTATCCGATAACGAATGAAGTCTAGCACTAGTATGCCATTTACTTAATATGGTCTGCTTGAAGTTTATTATAAGTTAAAAATCATGGT 5520  
 K Y E I I L S V D S D S L S H L R A S T H W K S L I T T K S T K N Y T L I D H N 90  
**SalL9R→**  
 M M V L L H A V Y S I V F V D V I I I K V Q R 23  
 TGTCCAATTTCCATCTAATATACCTTTGTCGGATTATCTATAGTACAGGAATATGATGGTATTATTAATGCTGTATCTATAGTCTTTGTAGATGTTATAATCAAAAAGTACAGA 5640  
 D L K W R I Y K T P N D I T C P I I I T N N C A T Y E I T K T S T I I M F T I G T 50  
 Y I N D I L T L D I F Y L F K M I P L L F I L F Y F A N G I E W H K F E T S E E 63  
 GGTATATCAACGATATTCTAATCTCTGACATTTTATTATTATTAATGATACCTTTGTTATTATTATTATTCTATTGCTAACGGTATCGAATGGCATAAGTTTGAACAGAGTGAAG 5760  
Y I L S I R V R S M 10  
**←SalL8L**  
 I I S T Y L L D D V L Y T G V N G A V Y T F S N N K L N K T G L T N N N Y I T T 103  
 AAATAATTTCTACTTACTTATTAGACGAGTATTATACCGGGTGTAAATGGGGCGGTATACACATTTTCAAATAATAAACTAAACAAAACTGGTTTAACTAATAATAATTATATAACAA 5880  
 S I K V E D A D K D T L V C G T N N G N P K C W K I D G S D D P K H R G R G Y A 143  
 CATCTATAAAAGTAGAGGATGCGGATAAGGATACATTAGTATGCGGAACCAATAACGGAATCCCAATGTTGGAATAATAGACGGTTCAGACGACCCAAAACATAGAGGTAGAGGATACG 6000  
 P Y Q N S K V T I I S H N G C V L S D I N I S K E G I K R W R R F D G P C G Y D 183  
 CTCCTATCAAAATAGCAAAAGTAACGATAATCAGTCACACGGATGTGTACTATCTGACATAAATATCAAAAGAAGGAATTAACGATGAGAGAGATTGACGGACCATGTGGTTATG 6120  
 L Y T A D N V I P K D G L R G A F V D K D G T Y D K V Y I L F T D T I G S K R I 223  
 ATTTATACACGGCGGATAACGTAATTCAAAAGATGGTTACGAGGAGCATTCTGTCGATAAAGATGGTACTTATGACAAAGTTTACATTCTTTCTACTGATACTATCGGCTCAAAGAGAA 6240  
 V K I P Y I A Q M C L N D E G G P S S I S S H R W S T F L K V E L E C D I D G R 263  
 TTGTCAAAATTCGGTATATAGCACAAATGTGCCTAACGACGAAGGTGGTCCATCATCTTGTCTAGTCATAGATGGTCGACGTTTCTCAAAGTCAAATTAGATGTGATATCGACGGAA 6360  
**SalF1R→**  
 M N T I K Q S F 8  
 S Y R Q I I H S R T I K T D N D T I L Y V F F D S P H Y V P I L \* 295  
 GAAGTATAGACAAATATTTCTTCTAGAACTATAAAACAGATAATGATACGATACTATATGTATTCTTCGATAGTCCGCATTATGTACCTATTCTATGAATACCATTAAACAATCTTT 6480  
 S T S K L E G Y T K Q L P S P A P G I C L P A G G A K V V P H T T F E V I E K Y N V 48  
TTCTACGTCAAAATTTGGAAGGATATACAAAGCAATGCGGCTCCAGCTCTCGGTATATGCTCTACGAGTGGAAAGTGTTCACATACCAGTTTGAAGTCATAGAAAAATATAATGT 6600  
 L D D I I K P L S N Q P I F E G P S G V K W F D I K E K E N E H R E Y R I Y F I 88  
 ACTAGATGATATTATAAGCCCTTTATCTAACCAACCTATCTTCAAGGACCGTCTGGTGTAAATGGTTCGATATAAAGGAGAAGGAAATGAACATCGGGAATATAGAATATATTCTCAT 6720  
 K E N S I Y S F D T K S K Q T R S S Q V D A R L F S V M V T S K P L F I A D I G 128  
 AAAAGAAAATCTATATATCGTTCGATACAAATCTAACAACTCGTAGCTCGCAAGTCGATGCGCGACTATTTTCAGTAATGGTAACCTCGAAACCGTTATTATAGCAGATATAGG 6840  
 I G V G M P Q M K K I L K M \*  
**SalF2R→**  
 M N K H K T D Y A G Y A C C V I C 17  
 GATAGGAGTAGGAATGCCCAAAATGAAAAAA TACTTAAATGTAATCTTAATCGAGTACACACACGACAATGAACAAACATAAGACAGATTATGCTGGTTATGCTGCTCGTAATAT 6960  
G L I V G I I F T A T L L K V V E R K L V H T P S I D K T I K D A Y I R E D C P 57  
 CGCGTCTAATTTGTTGAATATTTTACACGGACACTATTAAGTTGTAGAAGTAAATAGTTCTACACACCATCAATAGATAAAACGATAAAAGATGCATATATTAGAGAAGATTGTC 7080  
 T D W I S Y N N K C I H L S T D R K T W E E G R N A C K A L N P N S D L I K I E 97  
 CTACTGACTGGATAAGCTATAATAATAATGTATCCATTATCTACTGATCGAAAAACCTGGGAGGAAGACGTAATGCATGCAAAAGCTCTAAATCCAAATTCGGATCTAATTAAGATAG 7200  
 T P N E L S F L R S I R R G Y W V G E S E I L N O T T P Y N F I A K N A T K N G 137

AGACTCCAAACGAGTAAAGTTTAAAAAGCAATTAGACGCGGATATTGGGTAGGAGATCCGAAATATTAAACACGACACCCCATATAATTTATAGCTAAGAAATGCCACGAAGAATG 159

T K K R K Y I C S T T N T P K L H S C Y T I \* 159  
GACTAAAAACGGAATATATTGTAGCACACGAATATCCCAACTGCATTGCGTTACACTATATAACAATTACACTACATTTTATCATACCCTACTTCGGTTAGATGTTTTAG 7440

AAAAAATAAATATCGCGTACCGCTTCTGTGTTTATAAAAAAATAACAATTAAACAATTACAAA TTTTCTT TTAATTTTACGTGGTTGACCATCTCTGGTGGTAAAAATAATCTCTTAG 7560  
\* C N D F K K K L I K R P Q G N K T T F Y D R L 219

TGTTGGAATGGAATGCTGTTTAAATGTTTCCACACTCATCGTATATTTGACGTATGTAGTCACATCGTTTACGCAATAGICAGACTGTAGTTCTATCATGCTTCCTACATTAGAAGAGG 7680  
T P I S H Q K I N G C E D Y I K V Y T T V D N V C Y D S Q L E I M S G V N S P P 196

AACAGTTTAAAGTCTCTGGTTTAACTATTACCGTTAGTTTTCATGAAATCCTTGTGTTTATCCACTTCCACTTAAATTAAGTCCACTATACATTCTCTGTTAAATTTTACTAG 7800  
V T K F D R P K L R N G N T K M F D K T K D V E C K L Y I D V I C E E T L K V L 156

ATCGTCATGGGTCATAGAATTTATAGGTTCCGTAGTCCATGGATCCAACTAGCAAACTTCGCGTATACGGTATCGCGATTAGTGTATACACCAACTGTATGAAAATTAAGAAAAACAGTT 7920  
D D H T M S N I P E T T W P D L S A F K A Y V T D R N T Y V G V T H F N L F C N 116

TAATAGATCAACAGAAATATTAACTCTCGGTTGTACAGATGCACCATATTTATGGATTTGGATTACACAGTGTGTTGTCTGAGGGGTCGTCTAGCGTTGCTTACATAAACTTC 8040  
L L D V S I N L G G N S V S A G Y K H I K S E C T T Q R L P E D L T A E V Y V E 76

TATCCCATATCTTATTGTGAGAAATCGCATACCGATTATCATCATACACTGTTTGAAGTAAATGGTATACACATCAAAATAACAAATACTAACGAGTAACTTCGCAATATTG 8160  
I G M Y E K N D S D C V S K D D Y V T Q E S F P I C M L I V F V L S Y M 36

TTATCGTAATTGGAATAATAGTGTTCGAGTGAGTTGGATTATGTGAGTATGGATTGTATATTTTATTTTATTTTGAATAAGAATAAAATGCTAATGTCAAGTTTATTTCCAATAGAT 8280  
SalF4R→  
M A E W H K I I E D I S K N N K F E D A A I V D Y K T T K N 30  
GTCTTATTAATAAACATATATAA TAAATAACAATGGCTGAATGGCTAATAAATATTCGAGGATATCTCAAAAAATAAAGTTCGAGGATCGCGCCATCGTTGATTACAAGACTACAAGAA 8400

V L A A I P N R T F A K I N P G E I I P L I T A T N R N I L K P L I G Q K Y C I V Y 70  
TGTTCTAGCTGCTATTCTCAACAGAACATTGCGCAAGATTAACTCGGGTGAATTTATCTCTCTACTAATCGTAATATCTAAAACCTCTTATGTGTCAGAAATTTGTATGTGATA 8520

T N S L M D E N T Y A M E L L T G Y A P V S P I V I A R T H T A L I F L M G K P 110  
TACTAACTCTTAATGGATGAGAACACGTATGCTAGGATTGCTTACTGGTACGCCCCGTATCTCCGATCGTTATAGCGAGAACTCATACCGCACTTATATTTTGTATGGTAAAGC 8640

T T S R R D V Y R T C R D H A T R V R A T G N \* SalF5R→ 133  
M M M M 4  
AACAACATCCAGACGTGACGTGTATAGAACGTGTAGAGATCACGCTACCCGTGTACGCGCAACTGGTAATTAATAAAAAA GTAATATTCATATGTAGTGCAATTT TAAATA ATGATGA 8760

K W I I S I L T M S I M P V L A Y S S S I F R F H S E D V E L C Y G H L Y F D R 44  
TGAATGGATAATATCCATATTGACGATGTCAATAATCCGGTATTGGCATACAGCTCATCGATTTTATAGATTTCATTCAGAGGATGTGGAATATGTTATGGGCATTGTATTTTGATA 8880

I Y N V V N I K Y N P H I P Y R Y N F I N R T L T V D E L D D N V F F T H G Y F 84  
GGATCTATAATGTAGTAAATATAAAATATAATCCGATATTCATATAGATATAAATTTTATTAATCGCACGTTAACCGTAGATGAACCTAGACGATAATGCTTTTTTACACATGGTTATT 9000

L K H K Y G S L N P S L I V S L S G N L K Y N D I O C S V N V S C L I K N L A T 124  
TTTTAAAAACAAATATGGTTCATTAACTAGTTGATTGTCTCAATTATCAGGAACCTAAAAATAATGATATACAATGCTCAGTAAATGTATCGTGCTCATTAAAAATTTGGCAA 9120

S T S T I L T S K H K T Y S L H R S T C I T I I G Y D S I I W Y K D I N D K Y N 164  
CGAGTACATCTACTATATTAACATCTAAACATAAGACTTATTCTCTACATCGGTCACGTGATTACTATAATAGGATACGATTCTATTATATGTATAGGATATAAATGACAAGTATA 9240

SalF6R→  
M L L E M D 6  
G I Y D E T A I C M L I A S T L I V T I Y V F K K I K M N S \* 194  
ATGGCATCTATGATTTTACTGCAATATGTATGCTAATAGCGTCTACATTGATAGTGACCATATACGTGTTTAAAAAAAATAAAATGAACCTCTTAATTATGCTATGCTATTAGAAATGGAT 9360

K I K I T V D S K I G N V T I S Y N L E K I T I D V T P K K K K E K D V L L A 46  
AAAAACAAATACCGTTGATTCAAAATTTGTAATGTTGTATACCATATCGTAACTTGGAAAAGATAAATGTTGATGTACACCTAAAAAGAAAAAGAAAGGATGTATTATTAGCG 9480

Q S V A V E E A K D V K V E E K N I D I E D D D M D V E S A \* 78  
CAATCAGTTGCTGTGGAAGGCAAAAGATGTCAAGGTAGA AAAAAAATATTATCGATATTGAAGATGACGATGATATGATGTAGAAAGCGCATAATACGATCT AAAAAAATAAGTA 9600

TATAAATATCTTTTATTTACTGTACTCTTACTGTGTAGTGGTGATACCCCTACTCGATTATTTTTTAAAAAATACTTATTCTGATTCTTCTAGCCATTCCGTTGCTGTTCCGAATGCCA 9720  
\* E S E E L W K R T R E F A V 346

CATCGACGTTAAAGATAGGGGAGTGTGAAATCTAGTTCTGCATTGTTGGTAGCACCCTCAAATGTAGTGTGGATATCTTCAACGTATAGTTGTTGAGTAGTGATGTTTCTAAATA 9840  
D V N F I P S Y N F D L E A N N T R V E F T T T N S I K L T Y N N L L S P K R F L 332

GAATCTCTTCATCATCTTTCACGCGTACATTTTATGATCCATCTTGAATCTCTAGATCCTGTTCTTCTTCCCAATGGTTTCATCAATAGAAGTTAAACATATGCTATCGAACACG 9960  
I R K M D N N K C A Y M K L M W R P I R S G Q E I G L P K M L L N A F M D Y C S S 292

ATGGAGAGTAATCGTAGCAAAAGTAAGCATTTTCTTTAATCTCAGATCCCGGATACGGATA:ATTTTGCAGCCAAACAGTGCATCCATGCAACATTTCTACATATACCCGCGTATGCA 10080  
P S Y D Y C F Y A N G K I E S G P Y Q I Y K A A L V H M W A V N G V Y V R S H V 252

CCGCGTCATCATGACTGTACGATACATAATGTTACCGTGTGCTTACATTGCTCGTAAAGACTTTGCTCAATTTGCTCTCTCTCCGTAATTCAGTGGGCTCTTAGGCAACAAGTAT 10200  
A D D D V T R Y M I N G H Q K C Q E Y F V K T L K D G E G Y I G T P R L C C T Y 212

ACAATTTGCTCCATTCATGATTACGGAATTTGGCTTTTCATAACCACTGCTCGGCCATACGTTTACTTTTTCGCTATACATGCTCTGGTGATATATCATAAGGGTATGCTCATGGC 10320  
L K A G M N I V S A N N K M V L Q E A M R K S K A G P C Y V H G S S I D Y L T H E H G 172

CGATGAATGGATACCGGTGTTTATTGGTCTATTGCTTCCATCTACTAGTATAGATCAAACTCTGATTCCTAGGTCCACACAGCTGCCAATATAGTCTGTGTCATAATAGTTTA 10440  
I F P D G H K N P G I A E M S T Y I L Y K I G L D V C A A L I T Q T G C Y Y N V 132

CTTTCATGATTTTATTATCGGTGATTATTTCCAAATACATCCACTAGACGACCGGTATGAATAATCAGATTACCCCATCTAGCGCTTCTCTACCTTATCAAAGTCGTTTATATCACATT 10560  
K M I E N D T Y K G F V D V L A A T H I I L N V G L D L A E R V K D F D N I D C Q 92

GTATATAGTTTATAACCTTAACTTTCGAGGTTATGGTTGTGGATCTTCTACAATATCTAGACTCTGATTCTTGAACATCATCTGCCTAATTAACAGTTTACTATATACCTGCCTA 10680  
I Y N I V K V K S T I P Q P D E V I D I V R I E Q V D A S I L L K V I Y R G L 52

SalF7R→  
M A V C I I D H D N I R G 13  
GAAATCCGCGACCAACAGTAACCCGCTACACGGCCATTGCTGCCACTCATAATATCAGACTACTTATCTATTTTAC TAAATAATGGCTGTTGTATAATAGACCAGGATAATATCAGAG 10800  
F G A G G T V A Y V A M 12

SalF7L  
V I Y F E P V H G K D K V L G S V I G L K S G T Y S L I I H R Y G D I S Q G C D 53

GAGTTATTACTTTGAACCACTCCATGGAAAAGATAAAGTTTATAGGATCAGTTATTGGATTAAATCCGGAACGTATAGTTTGATAATTCATCGTTACGGAGATATTAGTCAAGGATGTG 10920

S I G S P E I F I G N I F V N R Y G V A Y V Y L D T D V N I S T I I G K A L S I 93  
ATTCCATAGGCAGTCCAGAAATATTATCGGTAAACATCTTTGTAACAGATATGGTGTAGCATATGTTTATTAGATACAGATGTAATATATCTACAATATTGGAAAGCGGTATCTA 11040

**SalF9R→**  
M A F D I S V N A S 10  
S K N D Q R L A C G V I G I S Y I N E K I I H F L T I N E N G V \* 125  
TTTCAAAAATGATCAGAGATTAGCGTGTGGAGTTATTGGTATTCTTACATAAATGAAAAGATAATACATTTTCTTACAATTAAACGAGAATGGCGTTTGATATATCAGTTAATCGCTCT 11160

K T I N A L V Y F S T Q Q N K L V I R N E V N D T H Y T V E F D R D K V V D T F 50  
AAAACAATAAATGCATTAGTTTACTTTCTACTCAGCAAAATAAATAGTTCATACGTAATGAAGTTAATGATACACACTACACTGCGAATTGTATAGGACAAAGTAGTTGACACGTTT 11280

I S Y N R H N D T I E I R G V L P E E T N I G C A V N T P V S M T Y L Y N K Y S 90  
ATTTTCATATAATAGACATAATGACACCATAGAGATAAGAGGGGTGCTTCCAGAGGAACTAATATTGGTTGCGCGGTTAATACGCCGGTTAGTATGACTTACTTGTATATAAGTATAGT 11400

F K L I L A E Y I R H R N T I S G N I Y S A L M T L D D L A I K Q Y G D I D L L 130  
TTTAAATGATTTTAGCAGAAATATAGACACAGAAATCTATATCCGCAATATTATTCCG/ ATTGATGACACTAGATGATTGGCTATTAAACAGTATGGAGACATTGATCTATTAT 11520

F N E K L K V D S D S G L F D F V N F V K D M I C C D S R I V V A L S S L V S K 170  
TTTAAATGAGAACTTAAAGTAGACTCCGATTCCGGACTATTGACTTTTGCAACTTTGTAAGGATATGATATGTTGTGATTCTAGAAATAGTAGTCTCTATCTAGTCTAGTATCTAAA 11640

H W E L T N K K Y R C M A L A E H I S D S I P I S E L S R L R Y N L C K Y L R G 210  
CATTGGGAATTGACAAATAAAAGATATAGGTGTATGCAATTAGCCGAACATATATCTGATAGTATTCGAATCTGAGCTATCTAGACTACGATACAAATCTATGTAAGTATCTACGCGGA 11760

H T E S I E D K F D Y F E D D D S S T C S A V T D R E T D V \* 240  
CACACTGAGAGCATAGAGGATAAATTTGATTATTTTGAAGACGATGATTCGTCTACATGTTCTGCGCTAACCCGACAGGAAACGGATGTATAATTTTTTTAT AGCGTGAAGGATATG AT 11880

AAAAATATAATTTGTTGATTATCCCATTCCAATCACCTTATATGATTCTGTAACACAATGAAGGAGTCTCATAGATGTATAGAGGTGAGATACTGGTTTGATAAAGCTTTTATCCAC 12000  
\* G M G I V K Y S E T V C H L L R M S T Y L D S V P K I F Q K N W M 252

ATGAGTATGTTTACTTTATGGTTAGACCCGATACCTTTAAACAATCACTGAAAAATGGAGTTAGGTTATGACCTCTCAGAAATCAGTTGCCGTTCTGGAACATTAAATGTATTTTTATG 12120  
L I N S K I T L G A Y K L L D S F I P T L Y Q G R L I L Q R E P V N F T N K I I 219

ATATACCTCAACGCATTTATGTGGGCATACAACAGTCATTACTAATGGAGTATCCAAGLGTTTTGTGTCTAGTATTTAAACAGAGAAGAGATTCAACAGACTGTTTATGAAGCTCG 12240  
Y E L A N I L H A Y A L L D N S I S Y E L L K L Q R T N L L L L S K L L S I N F E F 179

AATGCCGCTCATTTGTCGTTTATTTGATGATGTCGAATTTCCCAATATCATCACCGATGAGTAGCTCATCTTGTATCGGGATCCAAGTTTCTAAAGATGTCATTAAACCCCTCGATC 12360  
A A E N D S I N I I D F E G L I M V S S Y S M K N D P D L N E L S T M L G E I M 139

ATGAATGGATTATCATCATCGTTTATTTATGTTGGACATGAGCTTAGTCCGTTTGTCCACATCTATAGACGACGATTCTGAATATTTTATATATCCCTCTCTTTAACTCCAGGAAGCTTG 12480  
F P N I M M T K I N S M L K T R K D V D I S S S K Q I I K Y I G R K L E L F K D 99

TCAGGATGGTCTACTTTAATATGTTCTCGTCTAAGAGATGAAAACTTTTGGATGGTTGCACGCGACTTTTCTCTAAAGGATGACGTTGCCCAAGATCTCTCTTAAATGAATCCATCTTA 12600  
P H D V K I H E R R L S S F R Q I T A R S K E R F S S T A W S G R K F S T A D M K D 59

**SalF11R→**  
m t y i f i f s f i i 11  
TCCTTGGACAAGATGGACAGTCTATTTTCTTAGATGGTTTAAATATTTTGTACCCTATGATCTATAAAGGTAGACCTAATCGTCTCGGATGACCTATATTTTATTTTTCAGTTTATTA 12720  
K S L I S L R N E K S P K I N K N G M 19

**←SalF10L**  
r i n c k k y v r f t k m s r g a l i v f e g l d k s g k t t q c M N I M E S I 51  
TACGCATAAATGTAAAAATATGTTAGGTTTACAAAATGTCTCGTGGGGCATTAACTCGTTTGTGAAGGATTGGACAAATCTGGAAGAACACACAATGTATGAACATCATGGAATCTA 12840

P A N T I K Y L N F P Q R S T V T G K M I D D Y L T R K K K T Y N D H I V N L L F 91  
TACCGGCAAAACAGGATAAAATATCTTAACCTTCTCAGAGATCCACTGCTCAGGAAAGATGATAGATGACTATCTAAGCTGTAAGAAACCTATAATGATCATATAGTTAATCTATTAT 12960

C A N R W E F A S F I Q E Q L E O G I T L I V D R Y A F S G V A Y A A A K G A S 131  
TTTGTGCAATAGATGGGAGTTTGCATCTTTTATACAAGAACAACTAGAACAGGGAATACTTTAATAGTTGATAGATACGCGTTCTCTGGAGTAGCGTATGCCGCGCTAAAGCGCGT 13080

M T L S K S Y E S G L P K P D L V I F L E S G S K E I N R N V G E E I Y E D V T 171  
CAATGACTCTCAGTAAGAGTTATGAATCTGGATTGCTTAAACCCGACTTAGTTATATTCTTGAATCTGGTAGCAAGAAATTAATAGAAACGTGGCGAGGAAATTTATGAAGATGTTA 13200

F Q Q K V L Q E Y K K M I E E G D I H W Q I I S S E F E F E E D V K K E L I K N I V 211  
CATTTCAACAAAAGGTATTACAGAATATAAAAAATGATTGAAGAAGGAGATATTCTATGGCAAATATTCTTCTGAATTCGAGGAAGATGTAAAGAACGAGTTGATTGAAGATATAG 13320

**SalF12R→**  
M D E A Y Y S 7  
I E A I H T V T G P V G Q L W M \* 227  
TTATAGAGGCTATACACACGGTTACTGGACCACTGGGGCACTGTGGATGTAATAGTGAAATTACATTTTTTATAAATAGATGTTAGTACAGTGTATTAATGATGAAGCATATTACTC 13440

G N L E S V L G Y V S D M H T E L A S I S Q L V I A K I E T I D N D I L N K D I 47  
TGGCACTTGGAAATCAGTACTCGGATAGCTGTCCGATATGCATACCGAATCGCATCAATATCTCAATTAGTTATTGCAAGATAGAAACTATAGATAATGATATATTAACAAGGACAT 13560

V N F I M C R S N L D N P F I S F L D T V Y T I I D Q E N Y Q T E L I N S L D D 87  
TGTAATTTTATCATGTGTAGATCAAACTTGGATAATCCATTATCTCTTCTAGATAGTATATATATAGATCAAGAGAACTATCAGACCGAATGATTGAATTCATTAGACGA 13680

N E I I D C I V N K F M S F Y K C D N L E N I V D A I I T L K Y I M N N P D F K T 127  
CAATGAAATATCGATTGTATAGTTAATAGTTTATGAGCTTTTATAGGATAGCTAGAAAATATAGTAGATGCTATCATTACTCTAAAATATATATGAATATCCAGATTTTAAAC 13800

T Y A E V L G S R I A D I D I K Q V I R E N I L Q L S N D I R E R Y L \* 172  
TACGTATGCCGAAGTACTCGTTCCAGAAATAGCCGATATAGATATTAACAAGTGATACGTGAGAATATACTACAAATGTCTAATGATATCCGGAACGATATTTGTGAAAAATATTTAA 13920

**SalF13R→**  
M T S L R E F R K L C C D I Y H A S G Y K E K S K L I R D F I T D 33  
AAAAATATCTTTTTTATTAATGACGTCGCTCCGGAATTTAGAAAATATGCTGTGATATATATCACCATCAGGATATAAGAAAAATCTAAATTAATAGAGACTTTTAAACAGAT 14040

R D D K Y L I I K L L L P G L D D R I Y N M N D K Q I I K L Y S I I F K Q S Q E 73  
AGGGATGATAAATATTGATCATTAAGCTATTGCTTCCCGATTAGCAGTATGATTAATACATGAACGATAAACAATATAAATTTATATAGTATAAATTTAAACAATCTCAGGAA 14160

D M L Q D L G Y I G L D T I R T T F K E N T E I R P R D K S I L T L E D V D S 113  
GATATGCTACAAGATTTAGGATACGGATATATAGGAGACACTATTAGGACTTTCTTCAAGAGAGAACACAGAAATCCGTCACGAGATAAAAGCATTTTAACCTTTAGAAAGCATGGATAGT 14280

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 N D L K C V V M L I D K D 153  
TTCTTAACATCGTTATCATCGTAACTAAAGAAATCGCATCAAAATAAATATTGACTGATATCGCATCCGTTTGTACATGTAATGATTAAATGTGTAGTCATGCTTATGATAAAGAT 14400

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 L K E I I E N A S K Q N L 193



CTAAAAATTAAAGCGGGTCTCGTACGTACTTAAACGCTATTAGTCCTAATGCCTATGATGTGTTAGAAAACTAATAACTTGAAGAGATAATAGAAAATGCATCTAAACAAAATCTA 14520

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 K F P S G M F A E V K Y D 233  
GACTCTATACTATTCTGTTATGACTCCAATTAAATCCCATGTTAGCGGAATCGTGTGATTCTGTCAATAAGCGCTTTAAAAAATTTCCATCAGGAATGTTTCCGGAAGTCAAATACGAT 14640

G E R V Q V H K N N N E F A F F S R N M K P V L S H K V D Y L K E Y I P K A F K 273  
GGTGAAAGATACAAGTTCATAAAAAATAACGAGTTTGCTTCTTTAGTAGAAACATGAACCAAGTACTCTCATAAAGTGGATTATCTCAAAGAAATACATACCGAAAGCATTTAAA 14760

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 I H K K K E Y K N S N M C 313  
AAAGCTACGCTATCGTATTGGATTCTGAATTTGTTCTTGTAGACGAACATAATGTACCGCTCCCGTTTGAAGTTTAGGAATACACAAAAGAAAGAAATATAAAACTCTAACATGTGT 14880

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 D V M V E I P N R I V F S 353  
TTGTTCTGTTTACTGTTTGTACTTTGATGGATTGCGATATGACGGACATTCATTTGTACGAACGAAGATCTTTTCTCAAAGATGTTATGGTTGAATACCCAATAGAATAGTATTCTCA 15000

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 L K D I N G V Y E P G K R 393  
GAGTTGACGAATATTAGTAACGAGTCTCACTAATGATGATGACCTAACGAGAAAATAGAAAGTTGGTCTTAAAGATATTAAAGATATACGAACCGGGAAGAGA 15120

R W L K I K R D Y L N E G S M A D S A D L V V L G A Y Y G K G A K G G I M A V F 433  
AGATGGTTAAAAATAAGCGAGACTATTGACGAGGGTCCATGGCAGATTCTGCCGATTAGTAGTACTAGGTGCTTACTATGGTAAAGGAGCAAGGGTGGTATCATGCGAGTCTTT 15240

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 V L Q D Q L T M V K I N K 473  
CTAATGGTGTGTACGACGATGAATCCGGTAAATGGAAGACGGTTACCAAGTGTTCAGGACACGATGATAATACGTTAAGGGTTTGAAGACCAATTAACGATGGTTAAAAATAACAAG 15360

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 Q I W E I S G A E F T S S 513  
GATCCCAAAAAATCCAGAGTGGTGTGATGTTAATAAACTATATTCCCGATTTTGTAGTAGAGGATCCGAAACATCTCAGATATGGGAAATTTCCAGAGCAGAGTTTACATCTTCC 15480

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 T H L N D L T V N L T K S \* 552  
AAGTCCCATACCGCAATGGAATATCCATTAGATTTCCTAGATTACTAGGATTAGAGAAACGTTGGAAGAAATCTACTCATCTAAACGATTAGTAAAGTTGATTAAGTTGATTAATCTTAA 15600

**SalF14R→**  
M D G V I V Y C L N A L V K H G E E I N H I K 23  
TAGTTACATACAACTGAAAATTAATAACACTATTAGTTGGTGGTCCCATGGATGGTGTATTGTATAGTCTGTCTAAACGCGTTAGTAAACATAGCGGAGGAAATAAATCATATAAA 15720

N D F M I K P C C E R V C E K V K N V H I G G Q S K N N T V I A D L P Y M D N A 63  
AAATGATTTCATGATTAAACCATGTTGTGAAAGAGTTTGTGAAAAGTCAAGAACGTTACATTTGGCGGACATCTAAAAACATACAGTATTGCGAGATTGCCATATATGGATAATGC 15840

V S D V C N S L Y K K N V S R I S R F A N L I K I D D D D K T P T G V Y N Y F K 103  
GGTATCCGATGTATGCAATTCACGTGTATAAAAGAATGTATCAAGAATATCCAGATTGTCTAATTGTATAAGATAGATGACGATGACAAGACTCTACTGGTGTATATAATTATTTTAA 15960

P K D V I P V I I S I G K D K D V C E L L I S S D I S C A C V E L N S Y H V A I 143  
ACCTAAAGATGTTATTCTGTATTATCATATCTATAGGAAGGATAAAGATGTCTGTGAACATTTAATCTCATCAGACATATCGTGTGCATGCGTGGAGTTAAATTCATATCAGTAGCCAT 16080

L P M D V S F F T K G N A S L I I L L F D F S I D A A P L L R S V T D N N V I I 183  
TCTTCCCATGGATGTTTCTTTTACCAAGGAAATGCATCATTGATTATCTCTGTTTGTATTCTCTATCGATGACGACCTCTCTTAAAGAGTGAACCGATAATAATGTTATTAT 16200

S R H Q R L H D E L P S S N W F K F Y I S I K S D Y C S I L Y M V V D G S V M H 223  
ATCTAGACACACGCTCTACATGACGAGCTCCGAGTTCGAATTGGTTCAAGTTTACATAAGTATAAAGTCCGACTATTGTTCTATATTATATATGTTGTTGTGATGATGCA 16320

A I A D N R T H A I I S K N I L D N T T I N D E C R C C Y F E P O I R I L D R D 263  
TGGATAGCTGATAATAGAAGTACCGCAATTATTAGCAAAAATATATAGACAATCTACGATTAAACGATGAGTGTAGATGCTGTATTGTTGAACACAGATTAGGATCTCTGATAGAGA 16440

E M L N G S S C D M N R H C I M M N L P D V G K F G S S M L G K Y E P D M I K I 303  
TGAGATGCTCAATGGATCATCGTGTATAGAACAGACATTGATTATGATGAATTTACCTGATTAGGCAAAATTTGGATCTAGTATGTTGGGGAAATATGAACCTGACATGATTAAGAT 16560

A L S V A G N L I R N R D Y I P G R R G Y S Y Y V Y G I A S R 334  
TGCTCTTTCGGTGGCTGTTAATTTAATAAGAAATCGAGACTACATCCCGGGAGCGAGGATATAGCTACTACGTTTACGGTATAGCCTCTAGATAA TTTTTTAAGCAGCAATAAAAA 16680

**SalF15R→**  
M D I K I D I S I S G D K F T V T T R R E N E E R 25  
ACATAATTTTAAACCAATCTATTTCATCACTATTGTTGTGATCACCATGGACATAAAGATAGATATTAGTATTCTGTTGATAAATTTACGGTGACTACTAGGAGGAAAAATGAAGAAAG 16800

K K Y L P L Q K E K T T D V I K P D Y L E Y D D L L D R D E M F T I L E E Y F M 65  
AAAAAATATCTACCTTCCAAAAAGAAAAAATCTACTGATGTTATCAACCTGATTATCTTGAGTACGATGACTTGTTAGATAGAGATGAGATGTTTACTATTCTAGAGAAATATTTTAT 18920

Y R G L L G L R I K Y G R L C F T F A E N E I A K K F D N D A E E Q F G T I E E A L K Q K L R 105  
GTACAGAGGCTATTAGGCCTCAGAAATAAATATGGACGACTTTTAAACGAAATTAATAAAATTCGACAATGATGCGGAAGAACAAATTCGGTACTATAGAAGAACTCAAGCAAGAACTTAG 17040

L N S E E G A D N F I D Y I K V Q K Q D I V K L T V Y D C I S M I G L C A C V V 145  
ATTAATTTCTGAAGAGGAGCAGATAACTTTATAGATTATATAAAGTACAAAAACAGGATATCGTCAAACTTACTGTATACGATTGCATATCTATGATAGGATTGTGTGCATGCGTGGT 17160

D V W R N E K L F S R W K Y C L R A I K L F I N D H M L D K I K S I L Q N R L V 185  
AGATGTTTGAGAAATGAGAAATGTTTCTAGATGGAATATTGTTTACGAGCTATTAACTGTTTATTAATGATCAGATGCTTGATAAGATAAAATCTATACTGCGAATAGACTAGT 17280

Y V E M S \* 190  
ATATGTGGAATGTATAGAAAGTTAAAGTTAATGAGAGCAAAATATATAAGGTTGATTCCATATTGTTAT TTTTCTGTAATAGTTAGAAAAATACATTCGATGGTCTATCTACC 17400

AGATTATTATGTTTATAAGGTAC TTTTCTCATATAAACTAGAGTATGAGTAAGATAGTGTGTTTCAAAACATATAAACTCAAAATTTGATGGATGAGATATACAGCTATTAATTTTCA 17520

**SalF16R→**  
M I I L K D G Y K E F A 12  
AAATATATTTTAACTCTGATAACTTTAAACATGGATTTTGTGTTGTTTAAACGTTTAAAAAGATTTTGTATTGTAGTATATGATAATATAAAGATGGATATAAAGAATTTGCT 17640

D C M Y Y F L H Y Y I G Y G R Y T Y S A T N G S C D K G E Y L D K R H N Q C C N 52  
GACTGCATGTACTATTTTACATTACTACATTGGCTACGGCAGATATACCTACTCCGCCAACCGGGTCTGTGACAAGGGAGAAATATCTTGATAAGAGGCAATATCAATGTTGTAAT 17760

R C P P G E F A K V R C N G N D N T K C E R C P P H T Y T T I P I I L M D V I N 92  
CGGTGCCACTGAGAAATTTGCCAAGTTAGATGTAATGGTAACGATAACACAAAATGTGAACGCTGCCACCTCATACATATACCACAATCCCAATATTCTTAATGGATGTCATCAAT 17880

V E N A Q P D H L I R \* 103  
GTAGAAAATGCCAACCGGATCATTGTATAAGGTAAGGTGTACCGGAACAGAACAGTAAATGTTCTGTCTTCTGGTGGTATTGTGCTACTGATTCTTCACAGACTGAAGATTGTT 18000

GAAATTGTGTACCAAAAGGAGATGCTCCATGCGGATACCTTGGTGAATAGATGAACAGGAAATCCTATTGTAAATCGTGTGTGTTGGTGAATATTGCGACTACCTACGTAATTATA 18120

GACTTGATCCATTCTCTCCATGCAAACTATCTAAATGTAATTAATTATGATTTTGTATGATAATGTTACCATACTATATCGCTACTTGGTTAGTGATTATTTCAGTATGAAGACCTATT 18240

AATAATTACTTATCTTTGACGATCTGTTTATAATTATAATAAAAACTTATGGCATAGTAACATAAATGCTGACGCGATAAATTCGTAATAATCTGTTTGTCAAAATTTTAT 18360

**SalF17R→**

AAGGAATCTACAGGCATAAAAAATAAATAATTTATAATATACTCTTACAGCGCCCATCATGAATAACAGCAGTGAATGATTGCTGTTATTAATGGATTTAGAAATAGTGGACGAT 18480  
 M N N S S E L I A V I N G F R N S G R F 20  
 C D I S I V I N D E R I N A H K L I L S G A S E Y F S I L F S N N F I D S N E Y 60  
 TTTGTGATATTAGTATAGTTATTAATGATGAAAGGATAAACGCTCATAAACTCATCTATCTGGAGCCTCCGATATTTTCCATTCTGTTTCCAATAATTTATCGATTCTAATGAAT 18600  
 E V N L S H L D Y Q S V N D L I D Y I Y G I P L S L T N D N V K Y I L S T A D F 100  
 ACGAAGTTAATCTAAGTCATTAGATTATCAAAAGCGTTAACGATTGATCGATTACATTATATGGGATACCTTTGAGCCTAACTAACGATAACGTGAAATATATTTCTTCAACCGCTGATT 18720  
 L Q I G S A I T E C E N Y I L K N L C S K N C I D F Y I Y A D K Y N N K K I E S 140  
 TTTTACAAATGGATCTGCTATTACGGAGTGTGAAATACATACTTAAAACTTTTGTCTTAAAACTGTATCGATTTCTACATATACCGTGATAAATATAAACAAGAAAAATAGAAT 18840  
 A S F N T I L Q N I L R L I N D E N F K Y L T E E S M I K I L S D D M L N I K N 180  
 CAGCGTCGTTTAAACACAATATTACAAAATATTTTGAGACTCATCAACGATGAAACTTTAAATACTTAACAGAGGAATTAATGATAAAAAATTTAAGCGATGATATGTTAAATATAAAAA 18960  
 E D F A P L I L I K W L E S T Q Q S C T V E L L R C L R I S L S P Q V I K S L 220  
 ATGAGGATTTTGCACCACTAATTTCTCATTAAATGGTTAGAGAGTACTCAACAATCATGCACCGTGCAGTACTTAGATGCTCAGAATATCATTGCTTTCCCCACAAGTTATAAAATCAC 19080  
 Y S H Q L V S S I Y E C I T F L N N I A F L D E S F P R Y H S I E L I S I G I S 260  
 TTTATAGTCATCAACTGGTTAGTTCAATCTACGAATGTATAACATTCTTAAACAATATAGCATTTCTGGATGAATCATTTCTAGATACCATAGCATCGAGTTGATATCTATCGGTATAA 19200  
 N S H D K I S I N C Y N H K K K N T W E M I S S R R Y R C S F A V A V L D N I I Y 300  
 GTAATTCGATGATAAGATTTCCTATAAACTGCTACAATCATAAAAAAATACATGGGAAATGATATCTTACGTAGATATAGGTGATTTCCGACTGGCCGCTCTGGATAATATTATCT 19320  
 M M G G Y D Q S P Y R S S K V I A Y N T C T N S W I Y D I P E L K Y P R S N C G 340  
 ATATGATGGTGGATATGATCAGTCCCGTATAGAAGTTCAAAGTTATACGTACAAATACATGTACAAATTTCTGGATATATGATATACCAGAGCTAAAATATCTCTGTTCTAATTGTG 19440  
 G L A D D E Y I Y C I G G I R D Q D S S L T S S I D K W K P S K P Y W Q K Y A K 380  
 GGGGACTGGCTGATGACGAATACATTATTGTTATAGCGGCATACGCCATCAGGATTATCGTTGACATCTAGTATTTGATAAATGGAGCCATCAAACCATATTGGCAGAGTATGCTA 19560  
 M R E P K C D M G V A M L N G L I Y V M G G I V K G D T C T D A L E S L S E D G 420  
 AAATGCGGAACCAAAATGTGATATGGGGTTGCGATGTTAAACGGAATTAATATATGTCATGGTGGAAATCGTTAAAGGTGACACGTGTACCGACGCACTAGAGAGTTTATCAGAAGATG 19680  
 W M K H Q R L P I K M S N M S T I V H D G K I Y I S G G Y N N S S V N V I S N 460  
 GATGATGAAGCATACAGCTTCCCAATAAAATGTCCAATATGTGACGATGTTGTCATGATGGCAAGATTATATATCTGGAGGTTACAACAATAGTAGTGTATGTAATGTAATATCGA 19800  
 L V L S Y N P I Y D E W T K L S S L N I P R I N P A L W S A H N K L Y V G G G I 500  
 ATCTAGTCCTTAGCTATATCCGATATATGATGAATGGACCAATATCATCATTAACATTCCTAGAATTAATCCGCTCTATGGTCAGCGCATAAATAATATATGATAGGAGGAGGAA 19920  
 S D D V R T N T S E T Y D K E K D C W T L D N G H V L P R N Y I M Y K C E P I K 540  
 TATCTGATGATGTTCCGAATAACATCTGAACATACAGATAAAGAAAAAGATTGTTGGACATTGGATAATGGTCACGTGTTACCACGCAATTATATAATGTATAAATGCGAACCAGGATA 20040  
 H K Y P L E K T Q Y T N D F L K Y L E S F I G S \* 564  
 AACATAAATATCCATTGGAAAAACACAGTACACGGAATGATTTTCTAAAGTATTTGGAAGTTTATAGGTAGTTGATAGAACAAAATACATAAATTTTGTAAAAAATAAATCACTTTTAT 20160  
**SalGIr→**  
 M T R L P I L L L L L I S L V Y A T P F P Q T S K K I G D D A T L S C N R N N 38  
 ACTAATATGACACGATTACCAATACTTTTGTACTAATATCATTAGTATACGCTACACCTTTTCTCAGACATCTAAAAAATAGGTGATGATGCAACTCTATCATGTAATCGAAATAAT 20280  
 T N D Y V V M S A W Y K E P N S I I L L A A K S D V L Y F D N Y T K D K I S Y D 78  
 ACAATGACTACGTTGTTATGAGTCTTGGTATAAGGAGCCCAATTCATTATTTTAGTCTGCTAAAAGCGACGTCTTGTATTTTGATAAATTATACCAAGGATAAAATATCTTACGAC 20400  
 S P Y D D L V T T I T I K S L T A R D A G T Y V C A F F M T S T T N D T D K V D 118  
 TCTCCATACGATGATCTAGTTACAACATACACAATAAATCATTGACTGCTAGAGATGCCGTACTTATGATGTGCATTCTTTATGACATCAACTACAAATGACACTGATAAAGTAGAT 20520  
 Y E E Y S T E L I V N T D S E S T I D I I L S G S T H S P E T S S K K P D Y I D 158  
 TATGAAGAATCTCCACGAGTTGATTGTAATAACGATAGTGAATACGATACATACATCTCGATCTACACATCACCAGAACTAGTTCTAAGAAACCTGATTATATAGAT 20640  
 N S N C S S V F E I A T P E P I T D N V E D H T D T V T Y T S D S I N T V S A S 198  
 AATTCTAATGCTCGTCGCTGATTTCGAATCGCGACTCCGGAACCAATTACTGATAATGTAGAGATCATACAGACACCGTCACATACACTAGTGATAGCATTACAGTAAGTGCATCA 20760  
 S G E S T T D E T P E P I T D K E D H T V T D T V S Y T T V S T S S G I V T T K 238  
 TCTGGAGAAATCCACAACAGACGAGACTCCGGAACCAATTACTGATAAAGAGATCATACAGTACAGACACTGTCTCATACACTACAGTAAGTACATCATCTGGAATTGTCACACTACTAA 20880  
 S T T D D A D L Y D T Y N D N D T V P P T T V G G S T T S I S N Y K T K D F V E 278  
 TCAACCACCGATGATCGGATCTTTATGATACGTACAATGATAATGATACAGTACCACCACTACTGTAGGCGGTAGTACAACCTCTATTAGCAATTATAAAACCAAGGACTTTGTAGAA 21000  
 I F G I T A L I I L S A V A I F C I T Y Y I Y N K R S R K Y K T E N K V \* 314  
 ATATTGGTATTACCGCATTAATATATATTGTCGCGCGTGGCAATTTCTGTATTACATATTATATATATAATAAACGTTACGTAATAACAAAACAGAGAACAAGATGATGATTTTGTAC 21120  
 TTACATAAATGTCTGGATAGTAAATCTATCATATTGAGCGGACCATCTGGTTTAGGAAAGACAGCCATAGCCAAAAGACTATGGAATATATTGGATTGTGGTGTCCCATACCACT 21240  
**SalG2R→**  
 M E R E G V D Y H Y V N R E A I W K G I A A G N F L E H T E F L G N I 35  
 AGATTTCCTCGTCTATGGAACGAGAAGGTTGATTACCATACGTTAACAGAGAGGCCATCTGGAAGGGAATAGCCGCGGAACTTTCTAGAACATACCTAGATTTTTAGGAAATATT 21360  
 Y G T S K T A V N T A A I N N R I C V M D L N I D G V R S L K N T Y L M P Y S V 75  
 TACGGAATCTTAAAAACAGCTGTGAATACAGCGGCTATTAAATACGTTATTTGTGTGATGATCTAAACATCGACGGTGTAGAAGTCTTAAAAATACGTACCTAATGCTTACTCGGTG 21480  
 Y I R P T S L K M V E T K L R C R N T E A N D E I H R R V I L A K T D M D E A N 115  
 TATATAAGACCTACCTCTCTTAAAAATGGTTGAGACCAAGCTTCGTTGTAGAAACACTGAAGCTAACGATGAGATTTCATCGTCGCGTGATATTGGCAAAAACGGATATGGATGAGGCCAAC 21600  
 E A G L F D T I I I E D D V N L A Y S K L I Q I L Q D R I R M Y F N T N \* 151  
 GAAGCAGTCTATTTCGACACTATTATCATTTGAAGATGATGTGAATTTAGCATATAGTAAGTTAATTCAGATACTACAGACCGTATTAGAATGTATTTTAACACTAATTAGAGACTTAAG 21720  
 ACTTAAAACTTGATAAATAAATAAATACCTGTTTTATATGTGGCTATTTCACGCTAATGTATTAGTTAAATATTAATAAATTTACCACGTAACAACTTAAAAATTTAAATGATATTCA 21840  
**BlR→**  
 M N F Q G L V L T D N C K N Q W V V G P L I G K G G F G S I Y T T 33  
 TTGACAGATAGATCACACATTATGAACCTTCAAGGACTTGTGTTAACTGACAATTCGAAAAATCAATGGGTCGTTGGACCATTAATAGGAAAAGGTGGATTTCGGTAGTATTATACTACT 21960  
 N D N N Y V V K I E P K A N G S L F T E Q A F Y T R V L K P S V I E E W K K S H 73  
 AATGACAATAATTATGATGTAATAAGTACAGCCCAAGCTAACGGATCATTAATTTCGGAACAGGCATTTTATACATAGAGTACTTAAACCATCCGTTTCGAAAGATGGAAAAATCTCAC 22080  
 N I K H V G L I T C K A F G L Y K S I N V E Y R F L V I N R L G A D L D A V I R 113  
 AATATAAAGCAGTAGGCTTATCATCAGTGAAGGCAATTTGGTCTATACAAATCCATTAATGTGAATATCGATTCTTGGTAATTAATAGATTAGTGCAGATCTAGATCGGTGATCAGA 22200

GCCAATAATAATAGATTACCAAAAAGGTCGGTGATGTGATCGGAATCGAAATCTTAATACCATAACAATTTATGCACGAGCAGGATATTCTCACGGAGATATTAAAGCGAGTAAATATA 2320

V L D Q I D K N K A L Y L V D Y G L V S K F M S N G E H V P F I R N P N K A M D G N G 193  
GTCTTGGATCAAATAGATAAGATAAATATATCTAGTGGATTACGGATTGGTTCTAAATTCATGTCTAATGGAGAACATGTTCCATTATAAGAAATCCAAATAAAATGGATAACGGT 22440

T L E F T P I D S H K G Y V V S R R G D L E T L G Y C M I R W L G G I L P W T K 233  
ACTCTAGAATTTACACCTATAGATTGCGATAAAGGATACGTTGTATCTAGACGTGGAGATCTAGAAACACTTGATATTGTATGATTAGATGGTGGGAGGTATCTTGCCATGGACTAAG- 22560

I S E T K N C A L V S A T K Q K Y V N N T A T L L M T S L Q Y A P R E L L Q Y I 273  
ATATCTGAAACAAGAAATTTGTCATTAGTAAGTGCCACAAAAAGAAATATGTTAACAACTACTGCGCATTTGTTAATGACCAGTTTGCAATATGCACCTAGAGAATTGCTGCAATATATT 22680

T M V N S L T Y F E E P N Y D E F R H I L M Q G V Y Y \* 300  
ACCATGGTAAACTCTTTGACATATTTTGGAGAACCAATATGACGAGTTTCGGCACATTAATTAATGCAGGGTGATATATTATTAAGTGTGGTGTGGTTCGATGTAAAA TTTTTGTCGATA 22800

B2R→  
M A M F Y A H A L G G Y D E N L H A F P G I S 23  
AAAATTAaaaaataacttaattttatatttgatctcggtgtgacaaaccgaatcatggcagtgitttacgcacacgcgtctcggtgggtacgcagagaattcttcagtcccttccctggaaatat 22920

S T V A N D V R K Y S V V S V Y N N K Y D I V K D K Y M W C Y S Q V N K R Y I G 63  
CATCGACTGTTGCCAATGATGTCAGGAAATATTCTGTGTGTGTCAGTTTATAATAACAAGIATGACATTGTTAAAAGACAAATATATGTGGTGTACAGTCAGGTGAACAGAGATATATTG 23040

A L L P M F E C N E Y L Q I G D P I H D Q E G N Q I S I I T Y R H K N Y Y A L S 103  
GAGCACTGCTGCCTATGTTTGAGTGAATGAATATCTACAAAATGGAGATCCGATCCATGATCAAGAAGGAAATCAAATCTCTATCATCACATATCGCCACAAAAACTACTATGCTCTAA 23160

G I G E Y A G S L D L C L E G V G I H H V L E T G N A V Y V G K A V Q O H D Y S T I K E 143  
GCGGATCGGGTAGCAGAGCTAGACATGTTGTTGGAAGGATAGGAGTTCATCATACGTATTGAAACAGGAAACCGCTTATGTAAGAAAGTCAACATGATTATCTACTATCAAAAG 23280

K A K E M N A L S P G P I I D Y H V W I G D C I C Q V T A V D V H G K E I M R M 183  
AGAAGGCCAAAGAAATGAATGCACCTTAGTCCAGGACCTATCATTGATTACACGCTCGGATAGGAGATTGTATCTGTCAAGTTACTGCTGTGGACGTACATGGAAGGAAATATGAGAA 23400

R F K K G A V L P I P N L V K V K L G E N D T E N L S S T I S A A P S R \* 219  
TGAGATCAAAAAGGTCAGGTCCTCCGATCCCAAACTCGTAAAGTTAAACTTGGGAGAAATGATACAGAAATCTTCTCTACTATATCGGGCGGACCACGAGGTAAACCACCTC 23520

B3R→  
M K R L E T I R H M W S V V Y D H F D I V N G K E C C Y V H T 31  
TCTGGAAGACAGCGTGAATAATGTACTCATGAACGTTTGGAACTATACGCCATATGTGGTCTGTGCGTATATGATCATTGATATTGTAATGGTAAAGAAATGCTGTTATGTGCATAC 23640

H L S N Q N L I P S T V K T N L Y M K T M G S C I Q M D S M E A L E Y L S E L K 71  
GCATTGTCTAATCAAATCTTATACCGAGTACTGTAAAAACAAATTTGTACATGAAGCATATGGGATCATGCATTCAAATGGATTCCATGGAAGCTCTAGAGTATCTTAGCGAAGTAA 23760

E S G W S P R P E M Q E F E Y P D G V E D T E S I E R L T G V A G E E F F N R S E L Q 111  
GGAATCAGGTGGATGGAGTCCCAGACCAGAAATGCAGGAATTTGAATATCCAGATGGAGTGGAGACATGAATCAATTGAGAGATTGGTAGAGGATTCTTCAATAGATCAGAACTTCA 23880

A G E S V K F G N S I N V K H T S V S A K Q L R T R I R Q Q L P L Y S H L L P T 151  
GGCTGGTGAATCAGTCAAATTTGGTAATCTATTAAATGTTAAACATACATCTGTTCAGCTAAGCAACTAAGAACACGATATACGGCAGCAGCTTCTTTATACTCTCATCTTTTACCAAC 24000

Q R V D I C S L E L I I I H T K \* 167  
ACAAAGGGTGGATATTGTTCAITGGAGTTGATAATAATACACACAAAGTAATGGATTACGGTGGGTCTATGACTACCTCAGACTGGTAGAGAAATGATATAGAAAAGCATATCAAAAGA 24120

C T T C G T G T G T G C A T T T C T G T G A G A A A A G G A C A T C A A G T A C A C G T G C G A T T C A T C A A G T A T A T A A C C T G G G A T G A G G C T A C C T C G A C A T A C G T G T G C G C T A T C A A A G T G G A A 24240

A G A T G C T G T G T G C T G T G T T T G C A G A T T G G C C A G A A T C A T G G T A T A T G G A T A C T A A T G G T A T C A A G A A G T A T T C C A G A T G A A T G G G T G T C A C A T A T A A A A T T T A A T T A A T G T A A T A G 24360

A G A C A A A A T A A A G G T T G T A A T A T C A T A T A G A C A A T A A C A A T T A A T T A G T A A C T G T T A T C T T T T T T A A C T A A C C A A C T A A C T A T A T A C C T A T T A A T A C A T C G T A A T T A T A G T T 24480

B4R→  
M D F F K K E I 8  
C T T A A C A T C T A T T A A T C A T T A A T T C G C T T C T T T A A T T T T T A T A A A C T A A C A T T G T T A A T T G A A A G G G A A C A C T G T T A C A G A A T A T A A A T T A T A T A T G G A T T T T T T T T A A A A A G G A A A T 24600

L D W S I Y L S L H Y I A R V F S N F S T S H I I Q D Y N L V R T Y E K V D K T 48  
ACTTGACGTGGATATATTTATCTCTTCATTATATAGCACGGGTGTTTCCAATTTTCCACATCCCATATAATACAGGATTATAATCTCGTTTCCGACATACGAGAAAGTGGATAAAAC 25720

I V D F L S R L P N L F H I L E Y G E N I L H I Y S M D D A N T N I I I F F L D 88  
AATAGTIGAT TTTTATCTAGGTTGCCAAATTTATTCATATTTTGAATATGGGAAAATATTCTACATATTTTATCTATGGATGATGCTAATACGAATATTATATA TTTTTTTCTT AGA 25840

R V L N I N K N G S F I H N L R L S S S I N I K E Y V Y Q L V N N D H P D N R I 128  
TAGAGTATTAAATATTAAATAAGAACGGGTCAITTATACACAATCTCAGGTTATCATCATCCATTAAATAAAGAATATGTATATCAATTAGTTAATAATGATCATCCAGATAATAGGAT 24960

R L M L E N G R R T R H F L S Y I S D T V N I Y I C I L I N H G F Y I D A E D S 168  
AAGACTAATGCTTGAANAATGGAGTGAACAAGACAT TTTTGTCTATATATCAGATACAGTTAATATCTATATATGTATTTTAAATAAATCATGGATTTTATATAGATGCCGAAGACAG 25080

Y G C T L L H R C I Y H Y K K S E S E S Y N E L I K I L I N N G S D V D K K D T 208  
TTACGGTGTGATATTACATAGATGTATATATCATCTATAAGAAATCAGAAATCAGAAATCAATCAATGAATTAATTAAAGATATTGTTAAATAATGGATCAGATGAGATAAAAAAGATAC 25200

Y G N T P F I L L C K H D I N N V E L F E I C L E N A N I D S V D E N R Y T P L 248  
GTACGGAACACACCTTTTATCTATTATGTAAACACGATATCAACAACGTGGAATTTGTTGAGATATGTTTAGAGAATGCTAATATAGACTCTGTAGACTTTAATAGATATACACCTCT 25320

H Y V S C R N K Y D F V K L L I S K G A N V N A R N R F G T T P F Y C G I I H G 288  
TCATTATGTCTCATGTGCTAATAAATATGATTTTGTAAAGTTATTAATTTCTAAAGAGGAAATGTTAATGCGGTAAATAGATTGCGAAGTACTCCATTTTATTGTGGAAATATACACGG 25440

I S L I K L Y L E S D T E L E I D N E H I V R H L I I F D A V E S L D Y L L S R 328  
TATCTCGCTTATAAAACTATATTTGGAATCAGACAGAGTTAGAAATAGATAATGAACATATAGTTCGTCACTTTAATAATTTTGTAGTCTGTTGAATCTTTAGATTATCTATTATCCAG 25560

G V I D I N Y R T I Y N E T S I Y D A V S Y N A Y N T L V Y L L N R N G D F E T 368  
AGGAGTTATGTATTAATACTATCGTACTATATAACAACGAACATCTATTACGACGCTGTCAAGTTAATGCGTATAATACGTTGGTCTATCTATTAAACAGAAATGGTGAATTTTGAGAC 25680

I T T S G C T C I S E A V A N N N K I I M E V L L S K R P S L K I M I Q S M I A 408  
GATTACTACTAGTGATGTACATGTATTTTCGAAGCAGTGCAGAACCAACAAAAATAAATGAAGTACTATTGTCTAACGACCATCTTTGAAAATATGATACAGTCTATGATAGC 25800

I T K H K Q H N A D L L K M C I K Y T A C M T D Y D T L I D V Q S L Q Q Y K W Y 448  
AATTACTAAACATAAACAGCATAATGCAGATTTATTGAAAATGTGTAAAAATATCTGCGTGTATGACCGATTATGATACTCTTATAGATGTACAGTGCCTACAGCAATATAAATGGTA 25920

I L K C F D E I D I M K R C Y I A K N K T V F Q L V F C I A K D I N T L M R Y G K H 488  
TATTTTAAATGTTTCGATGAATAGATATCATGAAGAGATGTTATATAAANNAATCTGATTTCCAATAGT TTTTGTATCAAAGACATTAATACATTTAATGAGATACGGTAACA 26040

P S F V K C T S L D V Y G S R V R N I I A S I R Y R Q R L I S L L S K K L D P G 528  
 TCCTTCTTTCGTGAAATGCACTAGTCTCGAGTATACGGAAGTCGTGACGTAAATCATAGCATCTATTAGATATCGTCAGAGATTAATAGTCTATTATCCAAGAAGCTGGATCTCGG 26160  
 D K W S C F P N E I K Y K I L E N F N D N E L S T Y L K I L \* 558  
 AGATAAATGGTTCGTGTTTTCCTAACGAAATAAATAATAAATATTGGAAGAACCTTAAACGATAACGAACTATCCACATATCTAAAAATCTTATAAACACTATTAAAAATAAATCTAAGT 26280  
 B5R→  
 M K T I S V V T L L C V L P A 15  
 AGGATAAAATCACACTACATCATGTTTCTTTTAGTGTCTCGACAGTGTATACATATTTTAACTCA TAAATAAAAAAGAAACGATTTCGGTTGTACGTTGTTATGCGTACTACCTG 26400  
 V V Y S T C T V P T M N N A K L T S T E T S F N D K Q K V T F T C D Q G Y H S S 55  
 CTGTGTTTATTCACATGTACTGTACCCACTATGAATAACGCTAAATTAACGCTACCGAAACATCGTTAATGATAAACAGAAAC TACGTTTACATGTGATCAGGATATCATTTCT 26520  
 D P N A V C E T D K W K Y E N P C K K M C T V S D Y I S E L Y N K P L Y E V N S 95  
 CGGATCCAAATGCTGTCTCGGAAACAGATAAATGGAATACGAAATCCATCCGAAAAAATGTGCACAGTTT TGAATTACATCTCTGAATTATATAATAAACCCGTATACGAAAGTGAATT 26640  
 T M T L S C N G E T K Y F R C E E K N G N T S W N D T V T C P N A E C Q P L Q L 135  
 CCACCATGACACTAAGTTGCAACGCGGAAACAAATATTTTCGTTGCGAAGAAAAAATACCTTCTTGGAAATGATACCTGTTACGTTGCTCTAAATGCGGAATGTCAACCTCTTCAAT 26760  
 E H G S C Q P V K E K Y E F G E Y M T I N C D V G Y E V I G A S Y I S C T A N S 175  
 TAGAACACGGATCGTGTCAACAGTTAAAGAAAAATACTCATTGGGGAATATATGACTATCAACTGTGATGTGATATGAGTTATTGGTGCTTCGTACATAAGTTGACAGCTAATT 26880  
 W N V I P S C Q Q K C D M P S L S N G L I S G S T F S I G G V I H L S C K S G P 215  
 CTTGGAATGTTATTCATCATGTCAACAAAAATGTGATATGCCGCTCTATCTAATGGATTAAATTCGGGATCATATTTCTATCGGTGGCGTTATACATCTTAGTTGTAAGGTTGTT 27000  
 T L T G S P S S T C I D G K W N P V L P I C V R T N E E F D P V D D G P D D E T 255  
 TTACATAACGGGCTCTCCATCATCCACATGTATCGACGGTAAATGGAATCCGCTACTCCCAATATGTGTACGAACTAACGAAGAATTGTATCCGATGGATGGTCCCGACGATGAGA 27120  
 D L S K L S K D V V Q Y E Q E I E S L E A T Y H I I I V A L T I M G V I F L I S 295  
 CAGATTGAGCAACTCTCGAAAGACGTTGTACAATATGAACAAGAAATAGAATCGTTAGAAGCAACTTATCATATAATCATAGTGGCGTTAACAATATTGGGCGTCATATTTTAACTCT 27240  
 V I V L V C S C D K N N D Q Y K F H K L L P \* 317  
 CCGTTATAGTATTAGTTTGTCTGTGACAAAAAATGAACCAATATAAGTTCATAAATGCTACCG TAAATATAAATCCGTTAAAAATGAATAATTATAACGAACAAGTATCAAA 27360  
 B6R→  
 M S S S V D V D I Y D A V R A F L L R H Y Y N K R F I V Y G 30  
 AGATTAAAGAATTATAGCTAGAATCAATTGAGATGTCTCTTCAGTGGATGTTGATATCTACGATGCCGTTAGAGCATTTTACTCAGGCATATTATAACAAGAGATTATTGTGTATG 27480  
 R S N A I L H N I Y R L F T R C A V I P F D D I V R T M P N E S R V K Q W V M D 70  
 GAAGAAGTAACGCCATATTACATAATATATACAGGCTATTACAGATGCGCGTTATACCGTTTCGATGATATAGTACGTAATGCAAAATGAATACGTTGTTAAACGAATGGGTGATGG 27600  
 T L N G I M M N E R D V S V S V G T G I L F M E M F F D Y N K N S I N N Q L M Y 110  
 ATACACTTAATGGTATAATGATGAATGAACGCGATGTTCTGTAAGCGTTGGCACCAGGAATACTATTATCGGAAATGTTTTTCGATTACAATAAAAAATAGTATCAACAATCAACTAATGT 27720  
 D I I N S V S I I L A N E R Y R S A F N D D G I Y I R R N M I N K L Y G Y A S L 150  
 ATGATATAATTAATAGCGTATCTATAATTTAGCTAATGAGAGATATAGAAGCGCTTTTAAACGACGATGGTATATACATCCGTAGAAATATGATTAACAAGTTGTACGATACCGATCTC 27840  
 B7R→  
 M Y K 3  
 T T I G T I A G G V C Y Y L L M H L V S L Y K \* 173  
 TAACTACTATTGGCAGGATCGCTGAGGTGTTTGTATTATCTGTGTATGATCTAGTTAGTTGTATAAATAATATTTCATATACTAGTTAAAAATTTAAGATT TAAATGTATAAA 27960  
 K L I T F L F V I G A L A S Y S N N E Y T P F N K L S V K L Y I D G V D N I E N 43  
 AAATAATAACGTTTATTTATTTGTAATAGGTGCATTAGCATCTATTGCAATAATGAGTACACTCCGTTTAAATAACTGAGTGTAAAACTCTATATAGATGGAGTAGATAATATAGAAAAT 28080  
 S Y T D D N N E L V L N F K E Y T I S I I T E S C D V G F D S I D I D V I N D Y 83  
 TCATATCTGATGATAAATGAATTTGGTGTAAATTTTAAAGAGTACACAATTTCTATTATTACAGAGTCATGCGACGTCGGATTGATTCCATAGATATAGATGTTATAAACGACTAT 28200  
 K I I D M Y T I D S S T I Q R R G H T C R I S T K L S C H Y D K Y P Y I H K Y D 123  
 AAAATTATTGATATACCATGACTCGTCTACTATTCAACGCGAGGTGACACGTCGTAGAATATCTACCAAAATATCATGCCATTATGATAAGTACCCCTTATATTACAAAATATGAT 28320  
 G D E R Q Y S I T A E G K C Y K G I K Y E I S M I N D D T L L R K H T L K I G S 163  
 GGTGATGAGCGACAATATTCTATTACTGCAGAGGGAAATGCTATAAAGGAATAAATAATGAATAAGTATGATCAACGATGATACTTATTGAGAAAACATACTCTTAAATTTGGATCT 28440  
 B8R→  
 M R 182/2  
 T Y I F D R H G H S N T Y Y S K Y D F \*  
 ACTTATATATTGATCGTCATGGACATAGTAATACATATTATTCAAAATATGATTTTTAAAAATTTAAAAATATATTATCACTTCAGTGACAGTAGTCAAATAACAACAACACCATGAGA 28560  
 Y I I I L A V L F I N S I H A K I T S Y K F E S V N F D S K I E W T G D G L Y N 42  
 TATATTATAATTTCTCGCAGTTTGTTCATTAAATAGTATACAGCTAAAAATAACTAGTTATAAGTTTGAATCCGTCAAATTTGATTCCAAAATGAATGGACTGGGGATGGTCTATACAAT 28680  
 I S L K N Y G I K T W Q T M Y T N V P E G T Y D I S A F P K N D F V S F W V K F 82  
 ATATCCCTTAAAAATATGGCATCAAGACGTGGCAAAACATGTATACAAATGTACCAGAAGGAACATACGACATATCCGCATTTCCAAAGAATGATTTCGTATCTTTCTGGGTTAAATTT 28800  
 E Q G D Y K V E E Y C T G L C V E V K I G P P T V T L T E Y D D H I N L Y I E H 122  
 GAACAAGGCGATTATAAAGTGAAGAGTATTGTACGGGACTATCGTCGAAGTAAAAATGGACCACCGACTGAACATTGACTGAATACGACGACCATATCAATTTGTACATCGAGCAT 28920  
 P Y A T R G S K K I P I Y K R G D M C D I Y L L Y T A N F T F G D S K E P V P Y 162  
 CCGTATGCTACTAGAGGTAGCAAAAGATTCTTATTACAAACGCGGTGACATGTGTATATCTACTTGTGTATACGCGTAACCTCACATTCGGAGATTCTAAAGAACCAGTACCATAT 29040  
 D I D D Y D C T S T G C S I D F V T T E K V C V T A Q G A T E G F L E K I T P W 202  
 GATATCGATGACTACGATTGACAGCTACAGGTTGACAGATAGACTTTGTCAACAGAAAAAGTGTGCGTGACAGCAGGGAGCCACAGAGGGTTTCTCGAAAAAATTAATCCATGG 29160  
 S S K V C L T P K K S V Y T C A I R S K E D V P N F K D K M A R V I K R K F N K 242  
 AGTTGCAAGATGTCTGACACCTAAAAAGAGTGTATATACATGCGCAATTAGATCAAAAGAAAGTGTCCCAATTTCAAGGACAAAATGGCCAGAGTATCAAGAGAAAAATTAATAAAA 29280  
 Q S Q S Y L T K F L G S T S N D V T T F L S M L N L T K Y S \* 272  
 CAGTCTCAATCTTATTAACTAAATTTCTCGGTAGCACATCAATGATGTTACCACCTTTTCTAGCATGCTTAACCTGACTAAATATTCATAACTAA TTTTATTAAATGATACAAAAACG 29400  
 B9R→  
 M R S L I I V L L F P S I I Y S M S I R Q 21  
 AAATAAACTGCATATTATACACTGGTTAACGCCCTTATAGGCTTAACCATTTTCAAGATGAGGTCCTGATTATAGTCTCTGTCTCCCTCTATCATCTACTCCATGTCTATTAGAC 29520  
 C E K T E E E T W G L K I G L C I I A K D F Y P E R T D C S V H L P T A S E G L 61  
 AATGTGAGAAGACTCAAGAGAAACATGGGATTGAAAAATAGGTTGTGTATAATTGCCAAAGATTTCTATCCCGAAAGAACTGATTGACAGTTTCACTCTCCCACTGCAAGTGAAGGAT 29640  
 B10R→  
 M D S G I Y E T P I N Y K K S N V S A V S V N N T I F V T G G L F I N 35

T G A A A C T G A A G G C A A T G A T T C A G G G A T A T A C G A A A C C C G A T A A A T T A T A A A A A A A G C A A T G T G T C C G C T G T T C C G T T A A T A A T A C T A T T T C G T A A C T G G C G G A T T A T T C A T A A A T 29760  
 N S N S T I V V N N M E K L D I Y K D K Q W S I E E M P M A R V Y T H G I D S T F 75  
 A A C T C T A A T A G C A C G A T C G T G G T T A A C A A T A T G G A A A A C T T G A C A T T T A A A G C A A A C A A T G T C G A T T A G A A A T G C C T A G G C T A G G T A T A C C G C A T C G A C T C G A C T T 29880  
 G M L Y F A G G L S V T E Q Y G N L E K N N E I S C Y N P R T N K W F D I S Y T 115  
 G G A A T G T T A T A T T T G C C G G A G G T C T A T C C G T T A C C G A A C A A T A T G T G A A T T A G A G A A A A C A C G A G A T A T C T T G T T A C A A C C T A G A A C G A A T A A G T G G T T G T A T T T C A T A T A C T 30000  
 I Y K I S I S S L C K L N N V F Y V F S K D I G Y V E K Y D G A W K L V H D R L 155  
 A T T T A T A A G A T A T C C A T A T C A T C A T T G T G T A A A C T A A A T A A C G T C T T C T A T G T A T T A G T A A G G A C A T T G G A T A T G T G G A A A A G A T G A T G G T G C A T G G A A G T T A G T A C A T G A T C G T C T C 30120  
 P A I K A L S T S P Y \* B11R→ 166  
 C C C G C T A T A A A G C A T T A T C A A C T T C C T T A T T G A T T G A A A A T G A A A A T A T A A A T A G T T T T A T G T A T A G C A G T A T T A C C C T A T A G T T T T A T T G C T T A C T A C T A A C A T G G A T A C A G A T 30240  
 T N V E D I I N E I D R E K E E I L K N V E I E N N K N I N K N H P S G Y I R E 45  
 T T A C A A A T G T A G A G A T A T C A T A A A T G A A A T A G A T A G A G A A A G A A A T A C T A A A A A A T G T A G A A T T G A A A A T A A T A A A A C A T T A C A A G A A T C A T C C A A G T G G A T A T A T T A G A G 30360  
 A L V I N T S S N S D S I D K E V I E C I C H D V G I \* 72  
 A A G C A C T C G T T A T T A A T A C C A G T A G T A A T A G T T C C A T T G A T A A A G A A G T T A T A G A A T G T A T C T G T C A C G A T G T A G A A T A T A G A T C A T A T C T A C T A A T T T T A T A A T C G A T A C A A A A 30480  
 B12R→  
 M E S F K Y C F D N D G K K W I I G N T L Y S G N S I L Y K 30  
 C A T A A A A A C A A C T C G T T A T T A C A T A G C A G G A T C C T T C A A G T A T G T T T T G A T A A C G A T G G C A A G A A T G G A T T A C T C G A A A C T T T A T A T T C T G T A A T T C A A T A C T A T A T A 30600  
 V R K N F T S S F Y N Y V M K I D H K S H K P L L S E I R F Y I S V L D P L T I 70  
 A G G T C A G A A A A A T T C A C T A G T T C G T T C A C A A T T A C G A T A A G A T A G A T C A C A A A C C A T C A G A A C C A T T G T T G T C G A A A T C A G A T T C T A T A T A T C T G T A T T G G A T C C T T T G A C T A 30720  
 D N W T R E R G I K Y L A I P D L Y G I G E T D D Y M F F V I K N L G R V F A P 110  
 T C G A C A A C T G G A C A C G G A A C G T G G T A T A A G T A T T T G C C T A T T C C A G A T C T G T A T G G A A T T G G A G A A C C G A T G A T T A T A T G T T C T C G T T A T A A G A A T T T G G G A A G A G T A T T C G C C C 30840  
 K D T E S V F E A C V T M I N T L E F I H S Q G F T H G K I E P R N I L I R N K 150  
 C A A A G G A T A C T G A A T C A G T C T T C G A A G C A T C G C T C A T A T A A C A C G T T A G A G T T T A T A C A C T C T C A A G G A T T A C C A T G G A A A A T A G A A C C G A G A A T A T A C T G A T T A G A A A T A 30960  
 R L C T L I D Y S R T N K L Y K S G N S H I D Y N E D M I T S G N I N Y M C V D N 190  
 A A C G T C T T T C A T A A T T G C A T A T T C T A G A A C T A A C A A A T A T A C A A G A T G G A A A C T C A C A T A G A T T A C A C A G G A C A T G A T A A C T T C A G G A A A T A C A A T T A T A T T G T G T A G A C A 31080  
 H L G A T V S R R R G D L E M L G Y C M I E W F G G K L P W K N E S S I K V I K Q 230  
 A T C A T C T T G G A C A A C A G T T T C A A G A C A G G A G A T T A G A A A T G T T G G G A T A T G C A T G A T A G A A T G T T C G G T G G C A A A C T T C C A T G G A A A A C G A A A G T A G T A T A A A A G T A A T A A A C 31200  
 K K E Y K K F I A T F F E D C F P E G N E P L E L V R Y I E L V Y T L D Y S Q T 270  
 A A A A A A A G A A T A T A A A A A T T T A T A G C T A C T T T C T T G A G G A C T G T T T C C T G A A G A A T G A A C C T C T G A A A T T A G T A G A T A T A G A A T T A G T A T A C A C G T T A G A T T A T T C T C A A A 31320  
 P N Y D R L R K L F I Q D \* 283  
 C T C C T A A T T A G C A C T A C G T A A C T G T T T A T A C A A G A T T G A A A T A T A T T C T T T T T T A T A G A G T G T G T A G T G T T A C G G A T A T T A A T A T T A G A C T A T C T C T A T C G C G T A C A C G 31440  
 B13R→  
 M D I F R E I A S S M K G E N V F I S P A S I S S V L T I L Y Y G A N 35  
 A C C A A T A T C G A T T A C T A T T G G A T A T C T C A G G A A A T C G C A T C T C T A T G A A A G A G A G A A T G T A T T C A T T C C A G C C T C A A T C T C G T C A G A T T G A C A A T A C T G A T T A T T G A G A C T A A 31560  
 G S T A E Q L S K Y V E K E E N M D K V S A Q N I S F K S I N K V Y G R Y S A V 75  
 T G G A T C C A C T G C T G A A C A G C A T A C A A A T A T G T A G A A A A G A G A G A A C A T G G A T A A G G T A G C C T C A A A A T A T C T A T T C A A A T C C A A A A A A A G T A T A T G G C G A T A T T C T G C C G T 31680  
 F K D S F L R K I G D K F Q T V D F T D C R T I D A I N K C V D I F T E G K I N 115  
 G T T T A A G A T T C C T T T T G A G A A A A T T G C G G A T A A G T T T C A A A C T G T T G A C T T C A C T G A T T G T C G C A C T A T A G A T G C A A T C A C A A G T G T G A G A T A T C T T A C T A G A G G G A A A A T C A A 31800  
 P L L D E P L S P D T C L L A I S A V Y F K A K W L T P F E K E F T S D Y P F Y 155  
 T C C A C T A T T G G A T G A A C C A T T G T C C T G A T A C C T G T C C T A G A A A T A G T G C C G T A T A C T T T A A G C A A A A T G G T T G A C G C C A T T C G A A A G A A T T A C C A G T G A T T A T C C C T T T T A 31920  
 V S P T E M V D V S M S M Y G K A F N H A S V K E S F G N F S I E L P Y V G 195  
 C G T A T C C G A C G G A A T G G T A G A T G T A A G T A T G A T G T C A T G T A T G T A C G G C A A G C A T T A A T C A C G A C T G T A A G A A A T C T G T C G G A A C T T T T C A A T C A T A G A A C T G C C A T A T T G T T G 32040  
 D T S M M V I L P D K I D G L E S I E Q N L T D T N F K K W C N S L E A T F I D 235  
 A G A T A C T A G T A T G A T G C T A C T T C C C A G A C A A G A T T G A T G G A T T A G A A T C C A T A G A A C A A A A T C A A C A G A T A C A A A T T T A A G A A T G G T G T A A C T C T C T G A A G C T A C G T T A T C G A 32160  
 V H I P K F K V T G S Y N L V D T L V K S G L T E V F G S T G D Y S N M C N S D 275  
 T G T T C A C A T T C C C A A G T T T A A G T A A C A G C C T C G T A A T C T G G T G G A T A C T C T A G T A A A G T C A G G A C T G A C A G A G G T G T T C G G T T C A A C T G G A G A T T A T A G C A A T A T G T G A A T T C A G A 32280  
 V S V D A M G I H K A A T Y I D V N E A E Y T E A A A A C T A C A L V S D C A S T I T A N E F 315  
 T G T A G T G T C G A C C T A T A G T T C C A A A A C G T A T A G A T G T C A A G A A G A T A T A C A A G A C C T G C A G C A A C T T G T G C A C T G G T G C A G A C T G C A T C A C A A T T A C A A A T G A G T T 32400  
 C V D H P F I Y V I R H V D G K I L F V G R Y C S P T T N C \* 345  
 C T G T G A G A T C A T C C G T C A T C T A T G T G A T T A G C A T G T T G A T G A A A A A T T C T T T C G T T G G T A G A T A T T G C T C C G A C A C T A A T T G T T A C C A T T T T T T T A A A A A A T A G A A A A A 32520  
 B14R→  
 M T A N F S T H V F S P Q H C G C D R L T S I D 24  
 C A T G T G G T A T T A G T C A G G T C G T T A T T C T T C C A A T T G C A A T T G G T A A G A T G A C G G C C A A C T T A G T A C C C A C G T C T T T C A C C

CATGTCCTCAAATAACGCTATCATCCGGATATAATATATTAGATATTTTATGGGAAAAACGAGGACGGGATAATGATAGAATTATACCGATAGATAATGGTAGCAATATGCTAATTC 33360

N P T Q S D S G I Y I C I T T N E T Y C D M M S L N L T I V S V S E S N I D L I 127  
TGAACCCGACACAATCAGACTCTGGTATTTATATATGCATTACCAGAACGAAACCTACTGTGACATGATGTCGTTAAATTTGACAATCGTGCTGTCTCAGAAATCAATATAGATCTTA 33480

S Y P Q I V N E R S T G E M V C P N I N A F I A S N V N A D I I W S G H R R L R 167  
TCTCGTATCCACAATAGTAATAGAGAGACTCTACTGGCGAAATGGTATGTGCCAATATTAAATGCATTTTATGGCTAGTAGTAACGTAACGCGAGATATTATATGGAGCGGGCATCGACGCCCTTA 33600

N K R L K Q R T P G I I T I E D V R K N D A G Y Y T C V L E Y I Y G G K T Y N V 207  
GAAATAGAGACTTAAACAACGGACACCTGGAATTATTACCATAGAAGATGTTAGAAAAATGATGCTGGTTATTATACATGTGTTTTAGAAATATATACCGTGGCAAAACATATAACG 33720

T R I V K L E V R D K I I P S T M Q L P D G I V T S I G S N L T I A C R V S L R 247  
TAACCAGAATTGTAATTTAGAGGTACGGGATAAAATAATACCTTCTACTATGCAATTACCAGATGGCATTTGTAAC17CAATAGGTAGTAATTTGACTATTGCATGCGAGATATCGTTGA 33840

P P T T D A D V F W I S N G M Y Y E E I D G D G D N G R I S V A N K I Y M T D K R 287  
GACCTCCCACAACGGATGACAGACTCTTTTGGATAAGTAATGGTATGTATTACGAAGAAGATATGCGGACGGAAACGGTAGAATAAGTGTAGCAATAAAATCTATATGACCGATAAGA 33960

R V I T S R L N I N P V K E E D A T T F T C M A F T I P S I S K A T V T V S I T \* 326  
GACGTGTTATTACATCCCGTTAAACATTAACTCTGTCAAGGAAGAAGATGCTACAACGTTTACGTGTATGGCGTTTACTATTCTAGCATCAGTAAACAGTTACTGTTAGTATAACGT 34080

GAATGTATGTTGTACATTTCCATGTCAATNAGTTTATAAGAA TTTTATACATTATCTTCCAACAAGCAATTGACGAACGTATTGCTATGATTAACCTCCACGATACATATGCATATTA 34200  
\* V N D E L L C N V F T N S H N V G V I S H M N N 340

TTAATCATTAACCTGCAGACTATACCTAGAGCTATTTTGACATACCTCGTGTCTTGTGTAATTGCAGTATCTATATTATTAAGTACGTAAATCTAGCTATAGTTTATTATTAAATTTT 34320  
I M L K C V I G L A I K V Y E H E Q T I A T D I N N F Y T F R A I T K N N L K L 316

AGATAATACCGTCTCCTTATTTTAAAAATTCGCCATCCTTTATTAAATCATGAATGGGAATTTCTATGTGATAGTTAATATATTGTGAACAACAAGACAGATATCTATAGGAAAG 34440  
Y Y V T E K N E K F I A V D K I F I A D H I P I E I D Y N I Y Q S C S C I D I F P F P 276

GGTGAATGCGATACATGATCTATGATGTTTAAAAACACACGCAAACTTTGAAGAATTTATATAAATCATTCCATCGATACATCCTTCTATGTTGAGATGTATATCCAGGAATTCGT 34560  
P I R Y M S R H L K L V C A F K S S N I Y I M G D I C G E I N L H I Y G P I R K 236

TTATTAATATCGGGAATGTATAAATAAATCATGCGCGAGCTTCTACCGGAGTTATATCAGTTTAACTTACAAAATGTAACCAATACCTTTGCATGACTTGTGTTGTTTC 34680  
N I D P F T Y V L V N G P L P A E V P T I D T K L K C F T V L V K A H S T K N P 196

GGCAACGTGATTTAACTTGACGAATGGATTAATACAATAGCATGATCCGCGATCTATTAAGTTTCTTAAACGCGCTTGTATGTTTACAGAGACTTATCTAAATTTCTA 34800  
L T L K F K V F P N I V I A H D A C R N L K K V K V G K Y T K V S V K D L N R T 156

GTGCTGTGATGTTTATAAATAACGGGATATAGAATCACTACCTTACATACCAATACATTTATCAGATCCAGATAATAACAAATTTTGTGCGCTAACTTCTATA 34920  
S T H T I F I V P Y L V S D G V K S V W N C K I L D L Y Y V F K T A R V L E I N 116

TTGTTATATATTTTACAATGGTTATGATATCATGTAATAACTTGAATCTAACGCACATCGTGTACGTTTATACAATGTGATTTAGTGTAGTATATCTACACATGATTTTTCGCG 35040  
N Y I K C N T I I D H L L K S D L A C R R V N I C N H N L T T Y R C M Y K E A S 76

CTATAGTATCTGGACTAGTGATAAACTATCGTTATATCTGCTTCAATGAATCATCGAGATATTGCTCTCTGICATATTATACACCTGCATAAACTTTCTAGACATCTTACAATCC 35160  
Y Y E P S T I F S D N Y R D E I F E D L Y Q E R D Y E Y V Q M F K R S S 36

GTGTTATTTTAGGATCATATTTACATATTTACGGGTATATCAAGATGTTAGATTAGTTAATGGGAATCGTCTATAATAATGAATATTAACAATATATAGGAGCTTTTACCACAAGC 35280  
←B16L

B17R→  
M S R R L I Y V L N I N R E S T H K I Q E N E I Y T Y F S H C N I D H T S 37  
ATCATAAAATAGATCGTCGCTGATTTATGTTTAAATATCAACCGCGAATCAACTCATAAAATACAGAGAATGAATATATACATATTTTAGTCATTGCAATATAGACCATAGCTCT 35400

T E L D F V V K N Y D L N R R Q P V T G Y T A L H C Y L Y N N Y F T N D V L K I 77  
ACGAACTGATTTTGTAGTTAAAACTATGATCTAAACAGACGACAACCTGTAACCTGGGTACTGCACTACACTGCTATTTGTATAATAATTACTTTTACAAACGATGTACTGAAGATA 35520

L N H G V D V T M K T S S G R M P V Y I L L T R C C N I S H D V V I D M I D K 117  
TTATTAATCATGGAGTGGATGTAACGATGAAAACAGTAGCGGACGTATGCTGTTATATATTGCTTACTAGATGTTGCAATATTTACATGATGTAGATATGATAGACAAA 35640

D K N H L L H R D Y S N L L L E Y I K S R Y M L L K E E D I D E N I V S T L L D 157  
GATAAAACCACTTATTACATAGAGACTATTCCAACCTATTACTAGAGTATATAAAATCTCGTTACATGTTATTAAGGAAGAGGATATCGATGAGAATAGTATCCACTTTTATTAGAT 35760

K G I D P N F K Q D G Y T A L H Y Y Y L C L A H V Y K P G E C R K P I T I K K A 197  
AAGGAATCGATCTTAACCTTAAACAGACGGATATACAGCGTTACATTTATTTATTTGTTGCTCGCACACGTTTATAAACCCAGGTGAGTAGAAAACCGATACAGTAAACAAAGGCC 35880

K R I I S L F I Q H G A N L N A L D N C G N T P F H L Y L S I E M C N N I H M T 237  
AAGCGAATTATTTCTTTTATTAACACATGGAGCTAATCTAACCGCTTATAGATAATTGTTGTAATACACCATTCCTATTGTTATCTAGTATGAAATGTGTAATAATATTCATATGACT 36000

K M L L T F N P N F E I C N N H G L T P I L C Y I T S D Y I Q H D I L V M L I H 277  
AAAATGCTGTGACTTTTAAATCCGAATTTGAAATATGTAATAATCATGGATTAACGCCATACTATGTTATATAACTTCCGACTACATACAAACAGCATATTTCTGTTATGTTAATACAT 36120

H Y E T N V G E M P I D E R R I I V F E F I K T Y S T R P A D S I T Y L M N R F 317  
CACTATGAAACAAATGTGGAGAAATGCGGATAGATGAGCGTGTATATGCTATTGCGATTTATCAAAACATATTCTACAGCTCTGCAGATTCGATAACTTATTGATGAATAGGTTT 36240

K N I D I Y T R Y E G K T L L H V A C E Y N N T H V I D Y L I R I N G D I N A L 357  
AAAAATATAGATATTTATACCCGCTATGAAGAAAGACATTATTACAGTACGATGTGAATATAATAACACACGTAATAGATTATCTTATACGATCAACGGAGATATAAATGCGTTA 36360

T D N N K H A T Q L I I D N K E N S P Y T I N C L L Y I L R Y I V D K N V I R S 397  
ACCGAATAAACAAACCGCTACACAACCTATTATAGATAACAAAGAAAATTCGCCATATACCAATTAATTTGTTACTGTATATACCTTAGATATATTGTAGATAAGAATGTGATAAGATCG 36480

L V D Q L P S L P I F D I K S F E K F I S Y C I L L D D T F Y N R H V R N R D S 437  
TTGGTGGATCAACTTCCATCTACCTATCTTCGATATAAAATCATTGTAGAAATTCATATCTACTGTATACTTTTATAGATGACACATTTTACAATAGACAGCTTAGGAATCGCGATTCT 36600

K T Y R Y A F S K Y M S F D K Y D G I I T K C H K E T I L L K L S T V L D T T L 477  
AAACGATTCGATACGATTTTCAAAATACATGTCGTTGATAAATACGATGTTAATAAATAATGTCATAAGAAACAATATTGCTCAAACTATCCACTGTTCTAGACACTAGACTA 36720

Y A V L R C H N S K K L R R Y L T E L K K Y N N D K S F K I Y S N I M N E R Y L 517  
TATGAGTTTAAAGATGCCAATATTCGAAAAGTTAAGAAGATACCTACCGAGTTAAAAAATATAATAACGATAAGTCCTTTAAATATATTTCTAATATTATGAATGAGAGATACCTT 36840

N V Y Y K D M Y G S K V Y D K L F P V F T D K N C L L T L L P S E I I Y E I L Y 557  
AATGTATATTATAAAGATATGTCAGGTCAAAGGTATATGATAAACCTATTTCCTGTTTTCACAGATAAAAAATGTCTACTAACAATAGTACCTTCAGAAATATATACGAAATATTATAC 36960

M L T I N D L Y N I S Y P P T K V \* 574  
ATGCTGACAAATTAACGATCTTTATAATATATCGTATCCACTACCAAGTATAGTTGTA TTTTCTCATGCGATGTGTGTAAAAAACGATATTATA TAAATATTTTGTGCGGTATAA 37080  
B18R→

M T M K M M V H I Y F V S L L L L L L F H S Y A I D I E N E I T E F F N K M R D 39  
 TAAAGATGACGATGAAAATGATGGTACATATATATTCGTATCAITATTGTTATTGCTATTCCACAGTTACGCCATAGACATCGAAAATGAAATCACAGAATTCTTCAATAAAATGAGAG 37200  
 T L P A K D S K W L N P A C M F G G T M N D I A A L G E P F S A K C P P I E D S 79  
 ATACTCTACCACTAAAGACTCTAAATGGTTGAATCCAGCATGTATGTTCCGAGGCACAAATGAATGATATAGCCGCTCTAGGAGGCCATTACGCCAAAGTGCTCTCTATTGAAGACA 37320  
 L L S H R Y K D Y V V K W E R L E K N R R R Q V S N K R V K H G D L W I A N Y T 119  
 GTCTTTTATCGCAGATATAAAGACTATGTGGTTAAATGGGAAAGGCTAGAAAAAATAGACGGCGACAGGTTTCTAATAAACGTGTAAACATGGTGATTATGGATACGCAACTATA 37440  
 S K F S N R R Y L C T V T T K N G D C V Q G G I V R S H I R K P P S C I P K T Y E 159  
 CATCTAAATTCAGTAACCGTAGGTATTGTGCACCGTAACACAAAAGATGGTGACTGTGTTCCAGGGTATAGTTAGATCTCATATTAGAAAACCTCCTTCATGCATTCCAAAACATATG 37560  
 L G T H D K Y G I D L Y C G I L Y A K H Y N N I T W Y K D N K E I N I D D I K Y 199  
 AACTAGGTACTCATGATAAGTATGGCATAGACTTATACTGTGAATCTTCTACGCAAAACATTATAATAATAAATACTGGTATAAAGATAATAAGGAAATTAATATCGACGACATTAAGT 37680  
 S Q T G K E L I I H N P E L E D S G R Y D C Y V H Y D D V R I K N D I V V S R C 239  
 ATTACAAAACGGGAAAGGAATTAATTATTCATAATCCAGAGTTAGAGATAGCGGAAGATACGACTGTTACGTTTCATTACGACGACGTTAGAAATCAAGAAATGATATCGTAGTATCAAGAT 37800  
 K I L T V I P S Q D H R F K L I L D P K I N V T I G E P A N I T C T A V S T S L 279  
 GTAAAATACTTACGGTTATACGGTCAACAAGACCACAGGTTTAACTAATACTAGATCCAAAATCAACGTAAACGATAGGAGAACCCTGCAATATAACATGCAGCTGCTGTGCAACGTCAT 37920  
 L I D D V L I E W E N P S G W L I G F D F D V Y S V L T S R G G I T E A T L Y F 319  
 TATGTATTGACGATGACTGATTGAATGGGAAATCCATCCGATGGCTTATAGGATTTCGATTTTGATGTACTCTGTTTAACTAGTAGAGGCGGATTACCCGAGGCGACCTTGACT 38040  
 E N V T E E Y I G N T Y K C R G H N Y Y F E K T L T T T V V L E \* 351  
 TTGAAAATGTTACTGAAGAATATATAGGTAATACATATAAATGTCGTGGACACAATATTATTGAAAAAACCTTACAACACAGTAGTATTGGAGTAAATATACAATGCATTTTAT 38160  
 ATACATTACTGAATAATTATTATTATTATTATATCGTATTTGTCTATAACCGGACTATCTAGGTATTTGTATCTCACCAGTAGAGAACATATAAATGTAGACTCTATTAAACAGTTGT 38280  
 GTAAAATATCAGATCCTAATGCATGTTATAGATGTGGATGTACGGCTTACATGAGTACTTTTATAATTATAGATCAGTCAACGGAAATACAAGTATAGATACAACGGTTACTATCAAT 38400  
 ATTATTTTCTAGCGATTATGAAAATTATAATGAATATTATTATGATGATTATGATAGAAGTGGTATGAACAGTGAGAGTGATAATATATCAATCAAAACAGAATATGAATCTATGATG 38520  
 AAACACAAGATCAAGTACACAACCTAGTAGGTTACGACATTAACCTACAAACCAATGAGGATGATTTTATGGCTATGATAGATCAGTGGGTGTCATGATTATATAGATGAATCAATTA 38640  
 TAAAGTAGTATATGGAAGAGAGTCTCACGTAAGATGGCGGATATATGGCAAGAACATAATGATGGCGTATACAGTATAGGAAAGGAGTGATAGATAATATATACGAAGACAACCATAC 38760  
 CGTAGACGAATTTCAAGATAGACAGCGTATCAGATGTAGATGACGGGAACACATATCTCCGATAACTAATGATGTATCTACACAACATGGGAAAGAAATCAGAGTTAGATAGATA 38880  
**B19R→**  
 M E M Y P R H R Y S K H S V F K G F S D K V R K N D L D M N V V K E L L S N G A 40  
 CATGGAATGTATCCTCGTCATAGATATAGTAAGCATTCTGTCTTTAAGGATTTCTGACAAAGTTAGAAAAATGATTAGACATGAATGGTAAAAGAAATTACTTTCTAACCGTGC 39000  
 S L T I K D S S N K D P I T V Y F R R T I M N L E M I D E R K Y I V H S Y L K N 80  
 ATCTCTAACAAATTAAGGATAGCAGTAATAAGGATCCAATAACCGTTTATTTCGAAGAAGCATAATGAATTTAGAAATGATTGATGAACGAAAGTATATAGTACACTCCTATCTAAAAA 39120  
 Y K N F D Y P F F R K L V L T N K H C L N N Y N I S D S K Y G T P L H I L A S 120  
 TTATAAAAATTCGATTATCCATTTTCAGGAAGTTAGTTTGACTAATAAACATTTGTCTCAACAATTTATATAATATAAGCGACAGCAAAATATGGAACACCGCTACATATTTGGCGTC 39240  
 N K K L I T P N Y M K L L V Y N G N D I N A R G E D T Q M R T P L H K Y C V N L 160  
 TAATAAAAATTAATAACTCTAATACATGAAGTATTAGTGATAACCGAAATGATATAACCGCAGGAGTGAAGATACAAAATGCGAACTCCATTACAAAATTTGTGTAATTT 39360  
**B20R→**  
 Y I I I L N M V S D T I M K R L \* 176  
 GTATATCATAATTAAGATATGGTATCCGATACATATAATGAAAGATTATAGACGCTTTATAGAGTTAGAGCGCATCTAACTATTCCAATGACGATGGAATGATACAGTAGTTTAC 39480  
 C I H S N A E Y G Y N N I T N I K I I R K L L N L S R R A S H N L F R D R V M H 46  
 TGTATACACTCAATGCCGAATATGGTTATAACAATATTACTAACATAAGATAATACGTAACCTACTTATCTTAGTAGACGTGCGTCACATAATCTATTAGAGATCGAGTCATGCAC 39600  
 D Y I S N T Y I D L E C L D I I R S L D G F D I N G Y F E G R T P L H C A I Q H 86  
 GATTATATAAGTAATACATATATTGATCTTGAGTGTTAGATATTATTAGATCGTTGGATGGATTGATATCAATGGTTACTTTGAAGGACGTACACACTTCATTGCGCTATACAACAT 39720  
 N F T Q I A K Y L L D R G A D I V V P N T L I I H Q Y I Q \* 115  
 AACTTCACTCAGATTGCTAAGTACTTATTAGATCGAGGAGCTGATATAGTCGTACCCAACACATTGATTATACATCAGTACATACAGTAAATAGCATAGATATGGAGGAGTACAAATA 39840  
**B21R→**  
 S N K V I R Y N T V N N I W E T L P N F W T G T I N P G V V S H K D D I Y V V C 47  
 TTTCAAATAAAGTTATAAGGTACAACTGTCAATAATATATGGAAACATTACCTAACTCTGAGCTGGAACATAAAATCCAGGCGTGGTCTCGCATAAAGATGATATATGTTGTAT 39960  
 D I K D E K N V K T C I F R Y N T N T Y N G W E L V T T T E S R L S A L H T I L 87  
 CGGACATCAAGATGAAAAAATGTTAAACTTGTATATTAGATATAACAGGAATACGTATAACGGATGGGAATTGGTCACGACGACAGAAAGCAGATTATCAGCTCTGCATCTATTTC 40080  
 Y N N T I M M L H C Y E S Y M L Q D T F N V Y T R E W N H M C H Q H S N S Y I M 127  
 TTTATAACAAATACCAATGATCTGTTATGAATCGTATATGTTACAAGATACATTTATGTGTACACTCGGAATGGAATCATATGTGTATCAACATTCGAATAGTTATATCA 40200  
 Y N I L P I Y \* 134  
 TGTACATATACCTACCATCTACTAAATATAATAGAATAAAATAAATGAGTATGATCAATTTTATAGAACGATTGATTTTATCATTACCGCTTCATTCTTATATCTTTGCTTACGGAACC 40320  
 TATATTIAGAAACATCTACTAACGAT TTTTATGCTTGCAATTATTAATGGTATGTAATATGATTGATTGTGTACGCAATACCAATTTGTTAAGTATGAATACGGGTACAAAACATAAACT 40440  
**B22R→**  
 M D I F K E L I L K H T D E N V L I S P V S I 23  
 GAAGTTTAACTATTATTATTATGATATATATCGTTATTGTTGGTCTATACCATGGATATCTTTAAAGAACTAATCTTAAACACACGGATGAAAATGTTTGTATTCTCCAGTTCTA 40560  
 L S T L S I L N H G A A G S T A E Q L S K Y I E N M N E N T P D D N N D M D V D 63  
 TTTTATCTACTTTATCTATTCTAAATCATGGAGCAGCTGGTTCTACAGCTGAACAATATCAAAATATATAGAGAATATGAATGAGAATACACCCGATGACAATATGACATGGACGTAG 40680  
 I P Y C A T L A T A N K I Y G S D S I E F H A S F L Q K I K D D F Q T V N F N N 103  
 ATATTCGGTATTGTGCGCACTAGCTACCGCAATAAAATATACGGTAGCGATAGTATCGAGTTCACGCGCTCCTCTACAAAAAATAAAGACGATTTTCAAACTGTAACTTTAATA 40800  
 A N Q T K E L I N E W V K T M T N G K I N S L T S P L S I N T R M T V V S A V 143  
 ATGCAACCAACAAAGGAATCAATCAACGAATGGTTAAGACAATGACAAATGGTAAAAATTAATCTTATGACTAGTCCGCTATCCATTAATCTCGTATGACAGTTGTTAGCGCG 40920  
 H F K A M W K Y P F S K H L T Y T D K F Y I S K N I V T S V D M M V S T E N N L 183  
 TCCATTTTAAAGCAATGTGGAATATCCATTTTCTAAACATCTTACATATACAGACAAGTTTATATTCTAAGAATATAGTTACCAGTGTGATATGATGGTGAGCACTGAGAATAACT 41040

Q Y V H I N E L F G G F S I I D I P Y E G N S S M V I I L P D D I E G I Y N I E 223  
 TGCAATATGTACATATTAAATGAATTATTCGGAGGATTCTCTATTATCGATATTCATACGAGGAACTCTAGTATGGTAATTATACTACCGGACGACATAGAAGGTATATATAACATAG 41160

K N I T D E K F K K W C G M L S T K S I D L Y M P K F K V E M T E P Y N L V P I 263  
 AAAAAATATAACAGATGAAAAATTTAAAAATGGTGTGGTATGTTATCTACTAAAAGTATAGACTTGTATATGCCAAAGTTTAAAGTGGAATGACAGAACCGTATAATCTGGTACCGA 41280

L E N L G L T N I F G Y Y A D F S K M C N E T I T V E K F L H T T F I D V N E E 303  
 TTTTAGAAAAATTTAGGACTTACTAATATATTCGGATATTATGCAGATTTAGCAAGATGTGAATGAACTATCACTGTAGAAAAATTTCTACATACGACGTTTATAGATGTTAATGAGG 41400

Y T E A S A V T G V F M T N F S M V Y R T K V Y I N H P F M Y M I K D N T G R I 343  
 AGTATACAGAAGCATCGCCGTTACAGGAGTATTATGACTAATTTTCGATGGTATATCGTACGAAGGTCTACATAAACCATCCATTATGTACATGATTAAAGACAACACAGGACGTA 41520

L F I G K Y C Y P Q \* 353  
 TACTTTTATAGGGAATACTGCTATCCGCAATAAATATAACAAATAGACTTTTATCACGTTTATCTATGTCTAAATATTACAAATAGTAATAGTATAAACTAAAGCTGATAACTTAA 41640

**B23R→**

M M I Y G L I A C L I F V T S 15  
 AAAAAATAATATCATTACAAATTAATAGTATAAACTAAAAATTAACAAATCGTTATTATAAGTAATATCAAAATGATGATATACGGATTAATAGCGTGCTTATATTCGTGACTTC 41760

S I A S P L Y I P V I P P I S E D K S F N S V E V L V S L F R D D Q K D Y T V T 55  
 ATCCATCGCTAGTCCACTTTATATTCCTCCGTTATCCACCCATTTCCGAAGATAAATCGTTCAATAGTGTAGAGGTATTAGTTTCCTTGTGTAGAGATGACCAAAAAAGACATACGGTAAC 41880

S Q F N N Y T I D T K D W T I G V L S T P D G L D I P L T N I T Y W S R F T I G 95  
 TTCTCAGTTCAATACTACACTATCGATACCAAGACTGGACTATCGCGTACTATCCACACCTGATGGTTTGGATATACCATTGACTAATATAACTTATTTGGTCACGTTTACTATAGG 42000

R A L F K S E S E D I F Q K K M S I L G V S I E C K K S 125  
 TCGTGCATTGTTCAAATCAGAGTCTGAGGATATTTCCAAAAGAAAAATGAGTATTCTAGGTGTTTCTATAGAATGTAAGAAGTCGTCGAC 42090

Fig. 2. Nucleotide sequence of 42090 nucleotides of vaccinia virus (strain WR) starting from the left end of the *SalI* L restriction fragment and extending rightwards almost to the ITR. Deduced amino acid sequences of major ORFs are shown above or below the nucleotide sequence for rightward or leftward transcribed genes, respectively. Gene name and direction of transcription are included at the beginning of each ORF (for comparison with gene names from strain Copenhagen see Table 3). Underlined amino acids indicate uncharged or hydrophobic regions, or potential sites for *N*-linked glycosides. Underlined nucleotides indicate potential early transcriptional termination signals (TTTTNT) or late transcriptional initiation sites (TAAAT). Numbers on the right indicate nucleotide number, starting from the centre of the genome, or amino acid number for the individual ORFs.

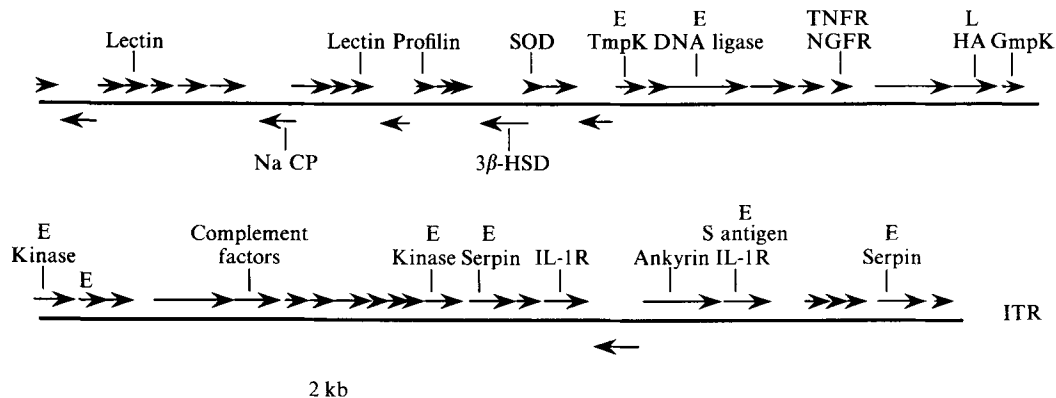


Fig. 3. Summary of the positions and direction of the ORFs, the principal homologies and the transcriptional regulation where experimentally determined. The *HindIII* site between the *HindIII* A and B fragments lies approximately at the right end of the top line. E, early gene; L, late gene; E and L, constitutive gene. Abbreviations: NaCP, sodium channel protein; IL-1R, interleukin-1 receptor; S antigen, soluble antigen.

starting from the left end of the 42 kb, the sizes of the predicted primary translation products and homologies with proteins from the SWISSPROT (version 14) database and from our database of vaccinia virus proteins. A brief description of some of the ORFs and homologies is presented.

#### (i) Genes potentially encoding enzymes

TmpK (SalF11R), DNA ligase (SalF13R) and two potential serine-threonine protein kinases (B1R and B12R) have been described previously (Smith *et al.*,

1989c, b; Kerr & Smith, 1989; Howard & Smith, 1989; Traktman *et al.*, 1989; Rempel *et al.*, 1990). Enzyme activity has been demonstrated for DNA ligase (Kerr & Smith, 1989; Colinas *et al.*, 1990) and TmpK (S. Hughes, unpublished results). In addition three other genes are predicted to encode proteins related to known enzymes. Gene SalF7L is predicted to encode a 39.3K protein with a similar size and 34% amino acid identity to human and bovine 3  $\beta$ -HSD (The *et al.*, 1989; Zhao *et al.*, 1989). This is a key enzyme in steroid biosynthesis and disease (adrenogenital syndrome) ensues from its deficiency in humans (Bongiovanni, 1962).



Table 1. *Open reading frames and protein homologies*

ORF	Size	Position		Homology/function	FASTA score*	Amino acid identity*	Reference
SalL1R	>14.9K	2	394	Ribosomal protein L7	91	28.9% in 83	Meyuhas & Klein (1990)
SalL2L	30.9K	1176	367				
SalL3R	20.5K	1294	1848				
SalL4R	19.5K	1875	2378	Fowlpox virus <i>Bam</i> HI ORF 11 (lectin)	114	20.9% in 86	Tomley <i>et al.</i> (1988)
				Fowlpox virus <i>Bam</i> HI ORF 8 (lectin)	147	21.5% in 107	Tomley <i>et al.</i> (1988)
				Fowlpox virus <i>Bam</i> HI ORF 2 (lectin)	132	21.3% in 150	Tomley <i>et al.</i> (1988)
				Cartilage proteoglycan (chicken)	118	22.8% in 101	Sai <i>et al.</i> (1986)
				CD23 (human)	116	16.1% in 155	Ikuta <i>et al.</i> (1987)
SalL5R	20.0K	2425	2952				
SalL6R	25.1K	3022	3684				
SalL7R	29.9K	3751	4539				
SalL8L	31.6K	5671	4841	Sodium channel protein I (rat)	104	21.7% in 157	Noda <i>et al.</i> (1986)
SalF9R	33.6K	5574	6458				
SalF1R	16.1K	6458	6883				
SalF2R	18.1K	6912	7388	CD23 (human)	174	26.1% in 92	Ikuta <i>et al.</i> (1987)
				Fowlpox virus <i>Bam</i> HI ORF 2 (lectin)	175	27.1% in 70	Tomley <i>et al.</i> (1988)
				Fowlpox virus <i>Bam</i> HI ORF 8 (lectin)	118	25.8% in 93	Tomley <i>et al.</i> (1988)
				Fowlpox virus <i>Bam</i> HI ORF 11 (lectin)	139	21.6% in 134	Tomley <i>et al.</i> (1988)
				Hepatic lectin 2 (human)	110	26.5% in 68	Spieß & Lodish (1985)
				Antifreeze protein (sea raven)	142	22.4% in 98	Ng <i>et al.</i> (1986)
				Cartilage-specific proteoglycan (human)	137	29.9% in 67	Baldwin <i>et al.</i> (1989)
SalF3L	25.0K	8148	7492	Shope fibroma virus T1	134	21.4% in 145	Upton <i>et al.</i> (1987)
SalF4R	15.0K	8312	8710	Profilin (human)	204	31.1% in 135	Kwiatkowski & Bruns (1988)
				Profilin (bovine)	201	31.1% in 135	Nystrom <i>et al.</i> (1979)
				Profilin (mouse)	197	30.4% in 135	Sri Widada <i>et al.</i> (1989)
SalF5R	22.5K	8751	9332				
SalF6R	8.7K	9343	9576				
SalF7L	39.3K	10717	9680	3- $\beta$ Hydroxy steroid dehydrogenase (human)	476	33.6% in 351	The <i>et al.</i> (1989)
				3- $\beta$ Hydroxy steroid dehydrogenase (bovine)	443	33.0% in 352	Zhao <i>et al.</i> (1989)
SalF8R	13.6K	10764	11138	Cu-Zn superoxide dismutase (sheep)	121	38.6% in 57	Schinina <i>et al.</i> (1986)
				Cu-Zn superoxide dismutase (bovine)	120	36.8% in 57	Steinman <i>et al.</i> (1974)
SalF9R	27.6K	11131	11850				
SalF10L	29.1K	12659	11904				
SalF11R	26.1K	12690	13370	Thymidylate kinase (yeast)	371	40.5% in 190	Jong <i>et al.</i> (1984)
SalF12R	18.7K	13421	13906				
SalF13R	63.3K	13942	15597	DNA ligase ( <i>S. pombe</i> )	520	29.8% in 389	Barker <i>et al.</i> (1987)
				DNA ligase ( <i>S. cerevisiae</i> )	537	30.9% in 414	Barker <i>et al.</i> (1985)
SalF14R	37.7K	15653	16654				
SalF15R	22.7K	16727	17296				
SalF16R	12.0K	17605	17913	Nerve growth factor receptor	97	36.2% in 47	Johnson <i>et al.</i> (1986)
				Tumour necrosis factor receptor	143	33.3% in 61	Smith <i>et al.</i> (1990)
				OX40 antigen	104	33.3% in 45	Mallet <i>et al.</i> (1990)
				Shope fibroma virus T2	123	31.1% in 61	Upton <i>et al.</i> (1987)
SalF17R	64.4K	18423	20114	Myxoma virus M9 protein	238	23.1% in 212	Upton <i>et al.</i> (1988)
				Chromosome maintenance protein ( <i>S. pombe</i> )	100	25.0% in 88	Adachi & Yanagida (1989)
SalG1R	34.6K	20167	21108	Vaccinia virus haemagglutinin, Ig superfamily			Shida (1986)
SalG2R	17.4K	21256	21708	Guanylate kinase (yeast)	231	29.3% in 140	Berger <i>et al.</i> (1989)
B1R	34.2K	21862	22761	Protein kinase C (rat)	128	22.7% in 185	Ono <i>et al.</i> (1988)
				Myosin light chain kinase, smooth muscle	130	20.6% in 165	Takio <i>et al.</i> (1986)
				Vaccinia virus B12R	338	37.8% in 172	Howard & Smith (1989)
B2R	24.6K	22854	23510				
B3R	19.4K	23549	24049				
B4R	65.4K	24578	26251	Cowpox virus range protein	260	25.6% in 219	Spehner <i>et al.</i> (1988)
				Fowlpox virus <i>Bam</i> HI ORF7	145	19.1% in 145	Tomley <i>et al.</i> (1988)
				Shope fibroma virus T5	171	20.0% in 370	Upton <i>et al.</i> (1987)
				Vaccinia virus M1	112	23.8% in 172	Tamin <i>et al.</i> (1988)
				Vaccinia virus 74K <i>Hind</i> III C	121	19.4% in 443	Kotwal & Moss (1988a)
				Vaccinia virus 27.4K <i>Hind</i> III C	258	27.9% in 219	Kotwal & Moss (1988a)
B5R	35.1K	26358	27308	Coagulation factor XIII (human)	297	27.2% in 246	Ichinose <i>et al.</i> (1986)
				Complement factor H precursor (human)	226	24.2% in 240	Schulz <i>et al.</i> (1986)
				Vaccinia virus 28K <i>Hind</i> III C	328	28.9% in 256	Kotwal & Moss (1988b)
				Various complement proteins			
B6R	20.1K	27393	27911				
B7R	21.3K	27952	28497	Vaccinia virus 20.7K <i>Hind</i> III C	200	23.7% in 152	Kotwal & Moss (1988a)
B8R	31.0K	28555	29370	Shope fibroma virus T7	224	25.3% in 182	Upton <i>et al.</i> (1987)

Table 1. *Continued*

ORF	Size	Position		Homology/function	FASTA score*	Amino acid identity*	Reference
B9R	8.8K	29460	29690	Shope fibroma virus T4	146	33.3% in 69	Upton <i>et al.</i> (1987)
B10R	18.9K	29656	30153	Shope fibroma virus T8	127	24.3% in 148	Upton <i>et al.</i> (1987)
B11R	8.2K	30228	30443				
B12R	33.3K	30513	31361	Vaccinia virus B1R	338	37.8% in 172	Howard & Smith (1989)
				Various protein kinases			
B13R	38.5K	31457	32491	Cowpox virus 38K serpin	1597	92.4% in 344	Pickup <i>et al.</i> (1986)
				Vaccinia virus B22R	813	45.7% in 352	Smith <i>et al.</i> (1989a)
				Vaccinia virus K2L	248	21.2% in 302	Bournsnel <i>et al.</i> (1988)
				Various serpins			
B14R	17.3K	32569	33015	Shope fibroma virus T3a	166	30.8% in 117	Upton <i>et al.</i> (1987)
B15R	36.5K	33102	34079	Interleukin-1 receptor (mouse)	198	21.4% in 266	Sims <i>et al.</i> (1988)
				Interleukin-1 receptor (human)	214	20.7% in 323	Sims <i>et al.</i> (1989)
				Ig superfamily			Smith & Chan (1991)
B16L	39.5K	35150	34131				
B17R	67.9K	35290	37011	Erythrocyte ankyrin (human)	162	17.7% in 311	Lux <i>et al.</i> (1990)
B18R	40.7K	37086	38138	Interleukin-6 receptor (human)	98	21.4% in 56	Yamasaki <i>et al.</i> (1988)
				Interleukin-1 receptor (mouse)	166	19.0% in 273	Sims <i>et al.</i> (1988)
				Ig superfamily			Smith & Chan (1991)
B19R	20.9K	38882	39409				
B20R	13.4K	39463	39807				
B21R	16.0K	39822	40223				
B22R	40.4K	40494	41552	Cowpox virus 38K serpin	818	46.3% in 352	Pickup <i>et al.</i> (1986)
				Vaccinia virus B13R	813	45.7% in 352	Smith <i>et al.</i> (1989a)
				Various serpins			
B23R	> 13.9K	41717	42085				
<i>Minor open reading frames</i>							
SalLa	9.9	485	748				
SalLb†	8.6	2585	2358				
SalLc†	7.9	4049	3840				
SalLd†	9.5	4523	4275				
SalFa†	9.8	6720	6454				
SalFb†	11.8	11774	11460				
SalFc	7.5	14730	14930				
SalFd†	10.8	17713	17444				
SalFe	8.9	17621	17863				
Ba†	11.8	22685	22365				
Bb†	10.8	23644	23360				
Bc†	7.8	27670	27458				
Bd†	10.4	33030	32758				
Be†	8.4	33643	33419				

\* FASTA scores and amino acid homologies (%) refer to the initial and not optimized alignments.

† ORFs running leftwards.

Gene SalF8R encodes a 13.6K protein with 39% amino acid identity with Cu-Zn superoxide dismutase (SOD), an enzyme that catalyses the dismutation of superoxide  $O_2^-$  to oxygen and hydrogen peroxide (McCord & Fridovich, 1969). The three-dimensional structure of SOD shows the protein to contain eight anti-parallel  $\beta$ -strands and three protruding loops, one of which binds the Zn atom (Tainer *et al.*, 1983). Alignment of the amino acid sequences of several Cu-Zn SODs with SalF8R (Fig. 4) indicates that although the vaccinia virus protein is predicted to contain the eight  $\beta$ -strands and probably has a similar overall globular structure, it lacks the Zn-binding region of loop 6,5 and the loop 7,8 near the C terminus. The majority of the conserved histidine residues involved in ion binding are also missing and

there is a C-terminal extension of 16 residues. These observations make it unlikely that this protein has SOD activity and, consistent with this, no additional SOD activity was found in extracts of virus-infected cells (J. Cavet, unpublished data). Although purified poxviruses have been reported to contain copper (Hoagland *et al.*, 1941), it seems unlikely that this is attributable to the product of SalF8R in this strain of virus.

Gene SalG2R has 30% amino acid similarity over 140 residues to guanylate kinase of *Saccharomyces cerevisiae* (Berger *et al.*, 1989). However, SalG2R lacks the N-terminal 42 residues present in the yeast protein including the nucleotide-binding site. Inspection of the amino acid sequence upstream of the first methionine and in another reading frame located sequences homolo-

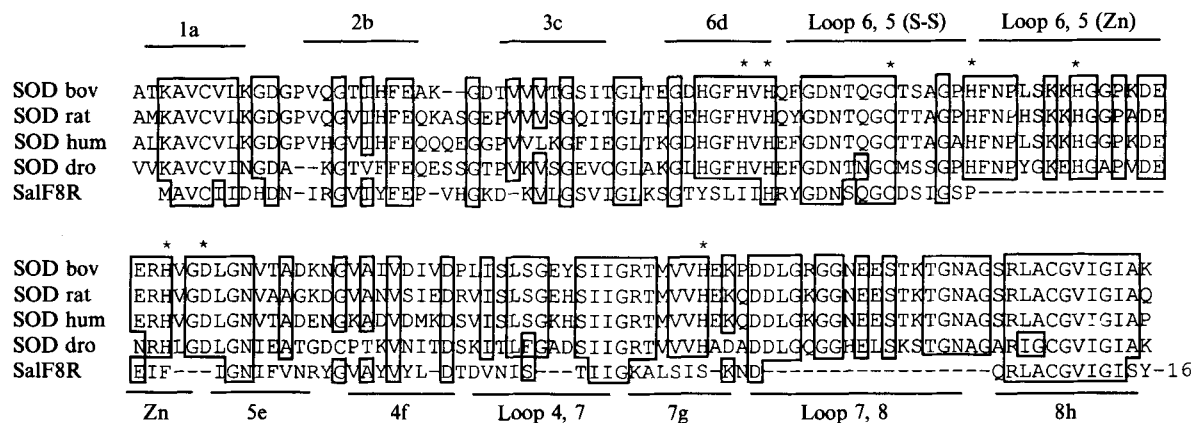


Fig. 4. Amino acid alignment of the predicted sequence of vaccinia virus protein SalF8R and the Cu-Zn superoxide dismutases from drosophila, rat, man and cow. The amino acids are boxed where five or more sequences are identical. Dashes indicate gaps introduced to make better alignment. Stars indicate histidine residues involved in Cu or Zn binding. Lines mark the eight anti-parallel  $\beta$ -strands and loops known from the three-dimensional structure and are named as previously described (Tainer *et al.*, 1983).

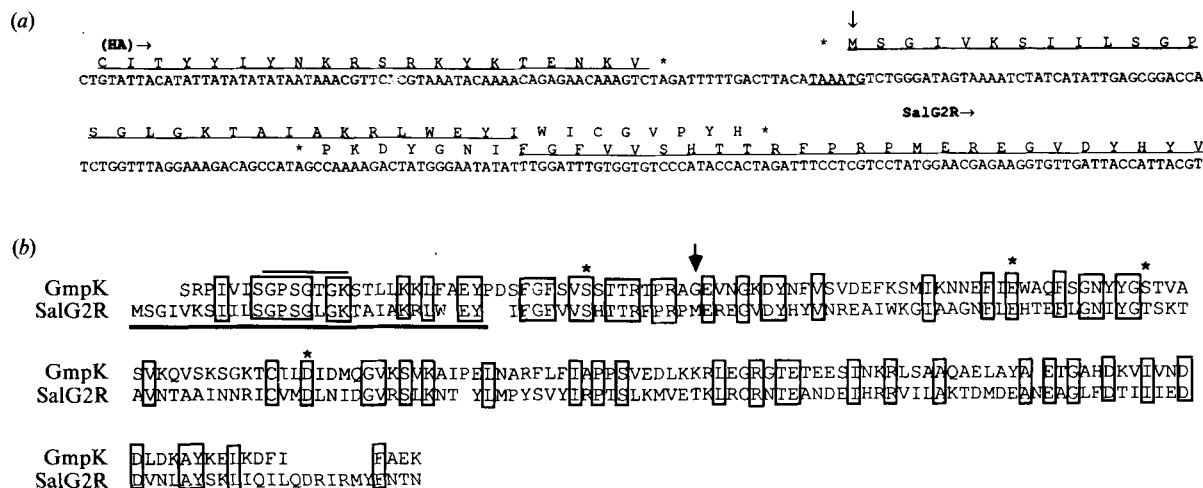


Fig. 5. (a) Nucleotide sequence from near the 3' end of the HA gene (SalG1R) to the 5' terminus of SalG2R (GmpK). ORF SalG2R starts at the ATG codon 11 amino acids from the end of the bottom line. Underlined upstream sequences are those present in the same or -1 reading frame that have homology to yeast GmpK (Berger *et al.*, 1989). (b) Amino acid alignment of yeast and vaccinia virus GmpK including the sequences present upstream of SalG2R in the -1 reading frame. Identical amino acids are boxed. The arrow indicates the beginning of SalG2R. Line above the amino acid sequence marks the nucleotide-binding site and the lower line marks the region in the -1 reading frame. Stars indicate the amino acids implicated in GMP-binding (Stehle & Schulz, 1990).

gous to the N terminus of the yeast protein (Fig. 5). This putative frameshift mutation in WR is also contained in the sequence of the Copenhagen strain (Goebel *et al.*, 1990) and the partial sequence from vaccinia virus strain IHD-J by translation of the nucleotide sequence downstream of HA (Shida, 1986). It seems likely that in an ancestral poxvirus, this gene may have been contiguous and the present discontinuous ORF presumably arose by a frameshift mutation. If guanylate kinase is found to be active in other poxviruses, it represents another poxvirus enzyme involved in biosynthesis of precursors for DNA synthesis. Others include thymidine kinase (Dubbs & Kit, 1964), TmpK (S. Hughes, unpublished results) and

ribonucleotide reductase (Slabaugh *et al.*, 1984). The virus also contains an ORF homologous to dUTPase (McGeoch, 1990).

#### (ii) Potential membrane-associated or secretory glycoproteins

Several of the predicted ORFs contain hydrophobic signal or anchor sequences and may be either secreted from the infected cell or remain associated with virus or cellular membranes (Fig. 2). These are detailed further where there are additional features of interest.

Genes SalL4R and SalF2R encode related proteins with a single large hydrophobic domain near the N

terminus. These proteins are similar to the family of lectins and to three proteins encoded by fowlpox virus (Tomley *et al.*, 1988). The topological arrangement of the encoded proteins in the cell membrane may be that of class II membrane glycoproteins in which the hydrophobic N-terminal region functions as a signal and anchor and the N terminus remains uncleaved on the cytoplasmic side of the plasma membrane while the C terminus is exposed on the cell surface (Blobel, 1980).

SalL8L encodes an extremely hydrophobic protein. At the N terminus there is a potential signal sequence that is followed by a relatively charged region. The remaining 60% of the protein consists of predominately hydrophobic amino acids that might function as multiple transmembrane regions. The protein has weak amino acid homology to a family of sodium channel proteins (Noda *et al.*, 1986) which also have multiple transmembrane domains and create a transmembrane pore through which Na<sup>+</sup> ions may be transported.

Gene SalG1R encodes the previously described HA (Shida, 1986) that is expressed late during infection on the cell membrane (Ichihashi, 1977) and in extracellular enveloped virus (Payne & Norrby, 1976). This gene has now been sequenced from four strains of vaccinia virus, WR (this paper), IHD-J (Shida, 1986), Copenhagen (Goebel *et al.*, 1990) and Tian Tan (Jin *et al.*, 1989), and is a member of the immunoglobulin (Ig) superfamily (Jin *et al.*, 1989). Only minor amino acid differences exist between the different strains and these are all located outside the single Ig domain of the HA molecule. Curiously, as the result of an internal deletion of three nucleotides, the WR HA lacks glutamic acid 215 found in the other vaccinia virus strains.

Gene B5R is predicted to encode a polypeptide with properties of a class I membrane glycoprotein. There is a large region of hydrophobic amino acids at the N and C terminus and the predicted external domain contains three potential N-linked glycosylation sites. The protein has homology to the superfamily of proteins that contain 60 amino acid repeats (Campbell *et al.*, 1988), and the closest matches are human coagulation factor XIIIb chain precursor (27.2% identity over 246 residues) and human complement factor H precursor (24.2% over 240 residues). Vaccinia virus encodes another member of this family of proteins (C28K) that has closest homology to the C4b-binding protein, is secreted from the infected cell and interferes with the classical pathway of complement activation (Kotwal & Moss, 1988b; Kotwal *et al.*, 1990). B5R differs from C28K in that it is probably membrane-anchored rather than secreted and is more closely related to factor H than the C4b-binding protein.

SalL3L (formerly named SalL4L), B7R, B8R and B9R all have homology to predicted proteins from the ITR of SFV (Howard *et al.*, 1991).

Ankyrin repeats of B17R				
Ankyrin	1	10	20	30
Consensus (>8/11)	G TPLH A	H	LL GA	
Consensus (>3/5)	G T L H Y			
B17R 57	TGYTALHCYLYNNYFTNDVLKILLNHGVDVTMK			
168	DGYTALHYYLCLAHVYKPGCECRKPITIKKAKR			
218	CGNTPFHLVLSIEMCNNIHMTKMLLTFFNPFEEI			
253	HGLTPILCYITSDYIQHDILVMLIHYYETNVGE			
327	EGKTLHCVACEYNNTHVIDYLIRINGDINALTD			

Fig. 6. Ankyrin-like repeats of B17R. The consensus sequence for ankyrin repeats is shown on the top line. Below this the consensus sequence of the repeats of B17R is shown and then the individual amino acid sequences for each of the five repeats of B17R. The numbers on the left indicate the starting positions for each of the amino acid repeats.

B15R and B18R both encode potential membrane-associated or secretory glycoproteins that are related to each other, to the Ig superfamily and to the human and murine interleukin-1 receptors (Smith & Chan, 1991). The sequence of B18R has been reported for another strain of virus (DIE) and contains three amino acid differences from the WR strain (Ueda *et al.*, 1990). This protein is located on the cell surface early during infection and additionally is the previously described soluble (S) antigen.

#### (iii) Other homologies

SalL1R contains an RNA-binding motif (Query *et al.*, 1989), noted by Goebel *et al.* (1990) and has 29% amino acid identity over 83 amino acids with murine 60S ribosomal protein L7 (Meyuhas & Klein, 1990). SalF4R encodes a 15.0K protein with 31% amino acid homology to mammalian profilins (Kwiatkowski & Bruns, 1988; Nystrom *et al.*, 1979; Sri Widada *et al.*, 1989). Profilin is an actin-binding protein that regulates the polymerization of soluble to filamentous actin. The vaccinia virus profilin has actin-binding properties (J. Rodriguez, unpublished data). The homology of SalF16R to a family of cysteine-rich proteins including TNFR and NGFR, and of B4R to poxvirus host range proteins and to proteins containing ankyrin or ankyrin-like repeats has been described (Howard *et al.*, 1991). B17R is also related to human ankyrin (FASTA score of 162 with 17.1% identity over 311 amino acids) and contains several ankyrin like repeats (Fig. 6). B13R and B22R encode proteins related to serpins (Kotwal & Moss, 1989; Smith *et al.*, 1989a).

#### Gene families

Several ORFs from within the region described have homology to other ORFs from this region or elsewhere in the virus genome and in some cases to genes from other poxviruses (Table 2). Some of these gene families have already been described, e.g. the serpins (Kotwal & Moss, 1989; Smith *et al.*, 1989a), protein kinases (Howard &

Table 2. *Vaccinia virus gene families*

Family	Genes	Other poxvirus genes	Homology/function	Reference
1	K2L B13R B22R	CPV 38K FPV	Serpins	Pickup <i>et al.</i> (1986) Boursnell <i>et al.</i> (1988) Tomley <i>et al.</i> (1988) Kotwal & Moss (1989) Smith <i>et al.</i> (1989a)
2	B1R B12R		Protein kinases	Howard & Smith (1989) Traktman <i>et al.</i> (1989)
3	SalL4R SalF2R	FPV ORFs 2, 8 and 11	Lectins	Tomley <i>et al.</i> (1988)
4	B5R C28K		Complement factors	Kotwal & Moss (1988b) Kotwal <i>et al.</i> (1990)
5	B14R K7R C17K SalF15R	SFV T3a		Howard <i>et al.</i> (1991) Upton <i>et al.</i> (1987) Boursnell <i>et al.</i> (1988) Kotwal & Moss (1988a)
6	SalF17R C59K C24K B10R B21R	SFV T6, T8 and T9		Howard <i>et al.</i> (1991) Kotwal & Moss (1988a) Upton <i>et al.</i> (1987)
7	C42K C37K			Kotwal & Moss (1989) Venkatesan <i>et al.</i> (1982) Kotwal & Moss (1988a)
8	K4L F37K		Envelope (37K)	Boursnell <i>et al.</i> (1988) Hirt <i>et al.</i> (1986)
9	B4R C74K C27K B17R M1L	SFV T5 FPV 47K CPV hr	Ankyrin Host range	Howard <i>et al.</i> (1991) Kotwal & Moss (1988a) Upton <i>et al.</i> (1987) Tomley <i>et al.</i> (1988) Tamin <i>et al.</i> (1988)
10	B7R C21K	SFV T2		Howard <i>et al.</i> (1991) Upton <i>et al.</i> (1987)
11	B15R B18R SalG1R		IL-1r Ig superfamily Haemagglutinin	Smith & Chan (1991) Jin <i>et al.</i> (1989) Shida (1986)
12	D9R D10R	FPV D9, D10		Niles <i>et al.</i> (1986) Binns <i>et al.</i> (1990)

Smith, 1989), D9R and D10R (Binns *et al.*, 1990), and four families which have homologues encoded in the ITR of the SFV genome (Howard *et al.*, 1991). One of these families includes host range proteins from cowpox and vaccinia virus, and several of the members have repeated sequences with homology to ankyrin or ankyrin-like repeats. The B17R gene product may be added to this family. Another gene family including the SFV T6, T8 and T9 proteins is now expanded to include the B21R protein. This has 30% amino acid identity over 66 residues with SalF17R. Because of the low homology it was left out of the alignment (Howard *et al.*, 1991), as were B10R and C24K. The two proteins related to complement factors (C28K and B5R) form a new family, as do the two proteins related to lectins (SalL4R and SalF2R). Other pairs of related genes include K4L and F37K (Boursnell *et al.*, 1988), and C42K and C37K (Kotwal & Moss, 1988a). The penultimate gene family shown in Table 2 includes B15R, B18R and HA, all of which encode members of the Ig superfamily. Both B15R

and B18R are more closely related to the interleukin-1 receptor than to each other, or to the HA (Smith & Chan, 1991).

#### *Comparison with vaccinia virus Copenhagen*

During the preparation of this manuscript the complete sequence of vaccinia virus strain Copenhagen was published (Goebel *et al.*, 1990). A comparison of the ORFs from the region of WR reported here and the corresponding region of Copenhagen is shown in Table 3. Because of differing nomenclature, the ORFs have been aligned by amino acid homology and provide a comparison of the names assigned to equivalent genes from the different virus strains.

Starting from the centre of the *Hind*III A fragment and working towards the terminus, the great majority of ORFs from the two virus strains are very similar in size and in amino acid sequence (greater than 97% amino acid identity) until after B18R in WR (equivalent to

Table 3. Comparison of WR and Copenhagen ORFs

WR			Copenhagen			
Gene	$M_r \times 10^{-3}$	Amino acids	Gene	$M_r \times 10^{-3}$	Amino acids	Amino acid identity
SalL1R	> 14.9	131	A31R	14.2	124	98.4% over 124
SalL2L	30.9	270	A32L	34.4	300	100% over 270
SalL3R	20.5	185	A33R	20.5	185	100%
SalL4R	19.5	168	A34R	19.5	168	98.8%
SalL5R	20.0	176	A35R	20.0	176	99.4%
SalL6R	25.1	221	A36R	25.1	221	100%
SalL7R	29.9	263	A37R	29.9	263	98.1%
SalL8L	31.6	277	A38L	31.6	277	97.5%
SalL9R	33.6	295	A39R	45.7	403	98.4% over 257
SalF1R	16.1	142	A39R	45.7	403	99.3% over 142
SalF2R	18.1	159	A40R	19.3	168	86.7% over 158
SalF3L	25.0	219	A41L	25.1	219	97.7%
SalF4R	15.0	133	A42R	15.0	133	100%
SalF5R	22.5	194	A43R	22.6	194	99.5%
SalF6R	8.7	78	—	—	—	—
SalF7L	39.3	346	A44R	39.4	346	98.6%
SalF8R	13.6	125	A45R	13.8	125	98.4%
SalF9R	27.6	240	A46R	24.7	214	91.6%
SalF10L	29.1	252	A47L	28.3	244	98.4% over 244
SalF11R	23.2	204	A48R	23.2	204	100%
SalF12R	18.7	162	A49R	18.8	162	99.4%
SalF13R	63.3	552	A50R	63.4	552	99.6%
SalF14R	37.7	334	A51R	37.7	334	99.4%
SalF15R	22.7	190	A52R	22.7	190	99.5%
SalF16R	12.0	103	A53R	12.0	103	99.0%
SalF17R	64.4	564	A55R	64.7	564	99.1%
SalG1R	34.6	314	A56R	34.8	315	98.1%
SalG2R	17.4	151	A57R	17.4	151	99.3%
B1R	34.2	300	B1R	34.3	300	99.7%
B2R	24.6	219	B2R	24.6	219	99.1%
B3R	19.4	167	B3R	14.4	124	96.8% over 124
B4R	65.4	558	B4R	65.3	558	98.7%
B5R	35.1	317	B5R	35.1	317	98.7%
B6R	20.1	173	B6R	20.1	173	100%
B7R	21.3	182	B7R	21.3	182	100%
B8R	31.0	272	B8R	31.2	272	98.2%
B9R	8.8	77	B9R	8.8	77	98.7%
B10R	18.9	166	B10R	18.9	166	100%
B11R	8.2	72	B11R	9.9	88	98.6%
B12R	33.3	283	B12R	33.4	283	99.3%
B13R	38.5	345	B13R	12.8	116	94.0% over 116
B13R	38.5	345	B14R	24.9	222	96.8% over 222
B14R	17.3	149	B15R	17.4	149	98.7%
B15R	36.5	326	B16R	32.5	290	97.9% over 284
B16L	39.5	340	B17L	39.5	340	98.5%
B17R	69.7	574	B18R	68.1	574	97.6%
B18R	40.7	351	B19R	40.9	353	99.1%
B19R	20.9	176	—	—	—	—
B20R	13.4	115	—	—	—	—
B21R	16.0	134	—	—	—	—
B22R	40.4	353	C12L	40.4	353	98.3%
B23R*	21.5	190	C14L	9.3	82	100% over 69

\* The size of B23R is taken from the complete sequence of this ORF (Kotwal & Moss, 1989).

B19R in Copenhagen) where the two sequences diverge completely (Table 3). This divergence results from terminal transposition and deletion event(s) since serpin B22R and the adjacent ORF B23R of WR are located at the opposite end of the Copenhagen genome (C12L and C14L, respectively). However, the transposition is

incomplete since there are no homologues of WR ORFs B19R, B20R and B21R anywhere in the Copenhagen genome and only part of B23R has been transposed. Evidently a DNA breakage occurred in the coding region of B23R during a transposition or deletion event. A further complication is the observation that in the

Copenhagen strain the genes C14L and C12L are separated by a short ORF (C13L), whereas in WR their homologues (B22R and B23R) are contiguous.

Within the conserved region to the left of B19R in WR (B20R in Copenhagen), there are some notable differences (Table 3). These include the presence of one ORF (A39R) in Copenhagen that corresponds to two ORFs in WR (SalL9R and SalF1R). Conversely, the serpin gene B13R of WR (Kotwal & Moss, 1989; Smith *et al.*, 1989a) is present as two ORFs in Copenhagen (B13R and B14R). An equivalent of WR SalF6R is not described in Copenhagen because the size of the ORF is less than the 65 codon cut off, and the homologue of Copenhagen A54L is considered only a minor ORF in WR (Table 1), because it overlaps a larger, complementary ORF (SalF16R) that has homology to TNFR (Howard *et al.*, 1991). Three Copenhagen ORFs are larger than the WR equivalents: A32L (34.4K), A40R (19.3K) and B11R (9.9K) correspond to WR SalL2R (30.9K), SalF2R (18.1K) and B11R (8.2K), respectively. On the other hand, three WR ORFs are larger than the Copenhagen homologues: SalF9R (27.6K), SalF10L (29.1K) and B3R (19.4K) match Copenhagen ORFs A46R (24.7K), A47L (28.3K) and B3R (14.4K), respectively.

These differences result from a variety of sequence alterations. The simplest is the presence of additional amino acid residues at either terminus deriving from extra in-frame methionine codon(s) upstream, or the lack of a translational termination codon. Two examples of N-terminal additions are seen with SalL2L and SalL9R. The former is 30 residues shorter than Copenhagen A32L, whereas the latter is 38 residues longer than Copenhagen A39R. The TmpK gene (SalF11R and A48R) provides another example. In WR there are an additional 23 N-terminal amino acids that are located upstream of the early RNA start site and were predicted, therefore, not to be part of the protein (Smith *et al.*, 1989c). In Copenhagen these sequences are not included in the TmpK ORF because the extra upstream methionine codon is out of frame owing to the loss of one nucleotide. Different C termini are seen with several ORFs. The simplest change occurs with B3R which in Copenhagen is 43 amino acids shorter, due to a frameshift mutation that creates an adjacent in-frame termination codon.

Several of the sequence changes result from short deletions in one sequence that may have arisen from recombination between short direct repeats. In Fig. 7 there are five examples of penta- or hexanucleotide direct repeats located eight to 10 nucleotides apart in one sequence that have apparently recombined in the other sequence to delete one copy of the repeat and the intervening DNA. These putative recombination events explain why genes SalL9R and SalF1R are separate

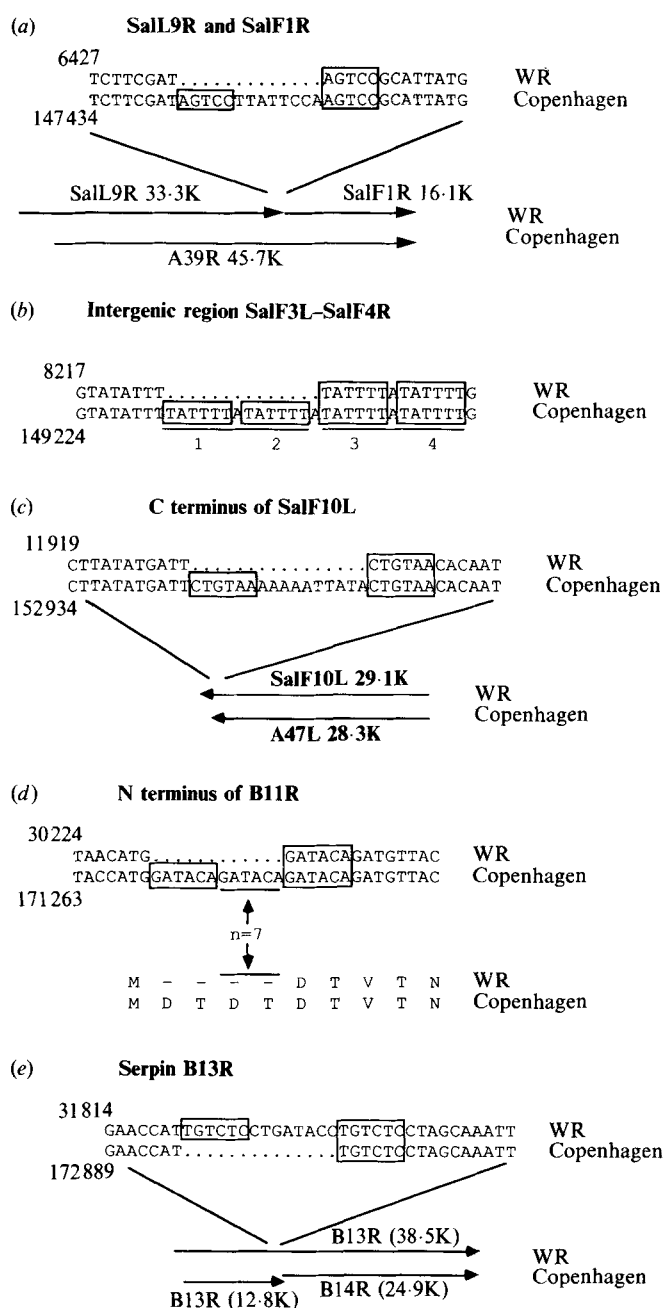


Fig. 7. Comparison of selected regions of the sequence of vaccinia virus WR presented in Fig. 2 with the corresponding sequence from vaccinia virus Copenhagen (Goebel *et al.*, 1990). Regions are chosen to show sequence variation occurring between direct repeats of five to six nucleotides (boxed) located eight to 10 nucleotides apart. Nucleotide numbers refer to sequences shown in Fig. 2 for WR and as described for Copenhagen (Goebel *et al.*, 1990). (a) A region near the junction of WR ORFs SalL9R and SalF1R which in Copenhagen is a single ORF, A39R. (b) Part of the non-coding region between SalF3L and SalF4R. (c) C-terminal region of WR SalF10L. (d) N-terminal region of B11R. The hexanucleotide GATACA is present nine times in Copenhagen but only three copies are shown for brevity. WR contains a single copy. The predicted amino acid sequence is written underneath. (e) A region of serpin B13R (WR) which in Copenhagen contains a frameshift forming two ORFs.

ORFs in WR but a single ORF (A39R) in Copenhagen (Fig. 7a), and why the serpin B13R of WR is broken into ORFs B13R and B14R in Copenhagen (Fig. 7e). Such events also explain the C-terminal sequence differences between SalL10L and A47L (Fig. 7c), and the differences at the N terminus of the B11R genes. Immediately after the initiating ATG codon of Copenhagen B11R the hexanucleotide GATACA (encoding the dipeptide DT) is repeated nine times. In WR only a single copy of GATACA remains (Fig. 7d). Although the sequence deleted between the first and ninth hexanucleotide repeat is much greater than that deleted in the other examples shown, the intervening DNA could have been removed stepwise using the other copies of the direct repeat, so that comparable size deletions were made at each stage. The above examples all occur within coding regions, but this phenomenon is present in non-coding regions too, for instance in the intergenic region between SalF3L and SalF4R (Fig. 7b). It should be emphasised that although there are several examples of this type of putative recombination, other length variations (larger or smaller) between the WR and Copenhagen sequences are present that have apparently occurred without the presence of direct repeats.

## Discussion

These data complete the sequence of a large region of the vaccinia virus (strain WR) genome near the right ITR that had hitherto not been analysed in detail. The region adjacent to the right ITR is variable in orthopoxviruses and undergoes deletions and terminal transpositions (Mackett & Archard, 1979; Moyer *et al.*, 1980; Lake & Cooper, 1980; Archard *et al.*, 1984; Pickup *et al.*, 1984, 1986), indicating that many of the ORFs may be non-essential for virus replication in cultured cells. Overall the gene arrangement is similar to that discovered in other regions of the genome; namely the genes are tightly packed, are transcribed from either DNA strand in the genomic centre (Niles *et al.*, 1986) and near the ITR are transcribed outward towards the terminal hairpin (Boursnell *et al.*, 1988; Kotwal & Moss, 1988a). The arrangement of ORFs described here and the limited transcriptional mapping, indicate that early transcription in opposing directions in the same region is not favoured, nor is the production of overlapping early transcripts since the majority of early genes contain termination signals after their coding regions. However there are exceptions, for instance, the DNA ligase mRNA terminates after the downstream gene (SalF14R) and not at the end of the ligase ORF (Smith *et al.*, 1989b). Similarly, mRNAs from early genes D4R and D5R have coterminal 3' ends but initiate at unique sites (Lee-Chen

& Niles, 1988). Excluding these exceptions, it is possible that the arrangement is advantageous by minimizing interference from upstream transcription units on the transcriptional activity of a downstream promoter (Ink & Pickup, 1990), reducing activation of non-specific antiviral mechanisms by dsRNA, or reducing torsional stress induced by RNA polymerase complexes moving towards each other on the same template. In contrast, the greater length, degree of overlap and complementarity of late transcripts suggests these features are not a major constraint on late gene expression.

Protein homology searches have revealed many striking matches to known proteins. The types of protein identified may be grouped roughly into (i) enzymes, (ii) factors likely to aid virus replication in the mammalian host by immune evasion or suppression, and (iii) proteins mediating virus-cell interactions. Predicted proteins with homology to enzymes include, DNA ligase, TmpK, guanylate kinase (GmpK), SOD, 3 $\beta$ -HSD and protein kinases. Enzyme activity has been demonstrated for DNA ligase (Kerr & Smith, 1989; Colinas *et al.*, 1990) and TmpK (S. Hughes, unpublished). SOD, GmpK and one of the proteins homologous to protein kinases (B12R) (Howard & Smith, 1989) are unlikely to be active. The role of 3 $\beta$ -HSD in virus replication is unknown.

Proteins that may cause virus immune evasion or suppression are perhaps the most fascinating group and indicate how well the virus is adapted to survive attack by the mammalian immune system. There are two genes related to complement control proteins (C28K and B5R). The former is a non-essential secretory protein that prevents activation of the complement cascade by the classical pathway (Kotwal & Moss, 1988b; Kotwal *et al.*, 1990). B5R is probably membrane-associated and has closer homology to factor H, a regulator of the alternative pathway, suggesting a possible interference with activation of the complement cascade by this mechanism. If it is present in virus particles it might prevent virion lysis and enable virus binding to target cells via complement receptors. Three genes encode proteins that may bind cytokines and inhibit inflammation. Two of these (B15R and B18R) are members of the Ig superfamily and have closest homology to the receptors for interleukins 1 and 6 (Smith & Chan, 1991), and the third is a discontinuous ORF with homology to TNFR (Howard *et al.*, 1991). If these proteins are present at the cell surface or are secreted, they may bind and sequester the ligands (IL-1, IL-6 or TNF) and diminish a normal inflammatory response during virus infection. The virus encodes a family of serpins (Boursnell *et al.*, 1988; Kotwal & Moss, 1989; Smith *et al.*, 1989a; Goebel *et al.*, 1990). In cowpox virus the equivalent of serpin B13R causes a haemorrhagic pox phenotype (Pickup *et al.*, 1986) and prevents migration of the white cells into the infected lesion *in vivo*



(Palumbo *et al.*, 1989) and *in vitro* (Chua *et al.*, 1990). Vaccinia virus serpins B13R and B18R cause a reduced antibody response to a foreign antigen expressed by recombinant vaccinia virus (Zhou *et al.*, 1990). They have also been proposed to prevent the processing of intracellular virus antigens and explain why some epitopes are not presented to class I major histocompatibility complex-restricted T cells from vaccinia virus-infected cells (Coupar *et al.*, 1986; Townsend *et al.*, 1988; Smith *et al.*, 1989a). A study of these virus evasion strategies may contribute to our knowledge of normal immune mechanisms.

Virus pathogenesis might also be influenced by several enzymes including TmpK, GmpK and SOD. The former two would (if active) provide an increased nucleotide pool and aid virus replication particularly in quiescent cells. TmpK is active (S. Hughes, unpublished results) and deletion causes attenuation (G. L. Smith, unpublished data) similar to that attributable to loss of TK (Buller *et al.*, 1985). GmpK and SOD are unlikely to be active in this strain of vaccinia virus (for reasons described above), but might be active in other more virulent poxviruses, for instance ectromelia and variola viruses, and contribute to the increased virulence of such viruses. Virus-encoded SOD activity could be advantageous for virus survival and replication in macrophages, a cell type that is infected by poxviruses and which may aid virus dissemination (Fenner, 1990). Phagocytosis by macrophages is usually followed by the production of toxic, oxidative free-radicals that destroy the engulfed organism. SOD dismutates the toxic radicals and so may provide a defense against destruction within macrophages.

The third type of gene provides information about virus interactions with the host cell. Profilin is an actin-binding protein that regulates actin polymerization into filaments. The finding of a related protein encoded by vaccinia virus is relevant to the observed interaction of virus with actin-containing cytoskeletal filaments (Hiller *et al.*, 1979; Hiller & Weber, 1982). B4R (Howard *et al.*, 1991) and B17R belong to a family of poxvirus proteins (Table 3) that contain repeated sequences found in ankyrin (a protein that mediates interaction between the cytoskeleton and the surface membrane; Lux *et al.*, 1990), the cell-cycle control protein cdc10 of *Schizosaccharomyces pombe* (Aves *et al.*, 1985) and the human proto-oncogene *bcl-3* (Ohno *et al.*, 1990). Two of the poxvirus proteins in this family determine host range and are essential for the replication of vaccinia virus in human cells (Gillard *et al.*, 1986), or cowpox virus in Chinese hamster ovary cells (Spehner *et al.*, 1988). The proteins encoded by SalL4R and SalF2R are homologous to lectins (Drickamer, 1988) and, if they function as such, may bind virus particles or infected cells to uninfected

cells to mediate virus spread. Lastly, the product of Sal8L encodes a protein with multiple hydrophobic transmembrane domains and has homology to sodium ion transport proteins. It is known that vaccinia virus infection induces changes in the membrane permeability of cells (Carrasco & Esteban, 1982) and the product of the SalL8L gene is a candidate for involvement in this process.

Comparison of the region of WR described here with vaccinia virus strain Copenhagen has demonstrated several differences. There are examples of frameshifts, deletions and transpositions causing the nomenclature adopted for ORFs within a given restriction fragment to be different for the two viruses. To overcome this difficulty the genes from the two viruses are aligned in Table 3 based upon amino acid homology so the names can be directly compared. Near the ITR the genes are either completely different or have been transposed to the opposite end of the genome; for instance, serpin B22R of WR is gene C12L of Copenhagen. Several of the frameshift differences in the conserved internal region (Table 3) can be attributed to short deletions between direct repeats eight to 10 nucleotides apart (Fig. 7), possibly by recombination. Without direct sequencing of virus DNA (for instance by using material generated by the polymerase chain reaction), it is unclear whether the deletion events have occurred in virus-infected cells or in cloned virus restriction fragments in plasmid vectors. There is ample precedent for recombination between direct repeats in poxvirus genomes such as in the ITR (Baroudy & Moss, 1982) or in artificially constructed gene duplications (Ball, 1987; Spyropoulos *et al.*, 1988; Shuman *et al.*, 1989; Spehner *et al.*, 1990; Kerr & Smith, 1991), although these direct repeats are all much larger than those described here. Irrespective of the situation in which these deletions have occurred, the requirements seem to be penta- or hexanucleotide direct repeats separated by eight to 10 nucleotides.

In summary, the region described contains greater than 50 tightly packed ORFs, which are predominately transcribed towards the genomic terminus and which display an extraordinary array of homologies to known protein sequences. In several cases the function of the encoded protein has been directly deduced from this homology. Our understanding of the replicative and pathogenic mechanisms of large DNA viruses has been greatly expanded by the complete nucleotide sequences of their genomes (Davison & Scott, 1986; Baer *et al.*, 1984; McGeoch *et al.*, 1988; Chee *et al.*, 1990; Goebel *et al.*, 1990). This has been particularly true in orthopoxviruses where the lack of splicing, high adenine and thymidine content (67%), distinctive codon usage and known transcriptional signals aid the interpretation of these raw nucleotide sequence data.

This work was supported by MRC grants G8703796CA and G8901790. We thank members of our laboratory for critical reading of the manuscript. S. T. H. is a Commonwealth Scholarship Commission Research Student and G. L. S. is a Lister Institute-Jenner Research Fellow.

## References

- ADACHI, Y. & YANAGIDA, M. (1989). High order chromosome structure is affected by cold-sensitive mutations in a *Schizosaccharomyces pombe* gene *crm1<sup>+</sup>* which encodes a 115-kD protein preferentially localised in the nucleus and its periphery. *Journal of Cell Biology* **108**, 1195–1207.
- AHN, B.-Y., GERSHON, P. D., JONES, E. V. & MOSS, B. (1990). Identification of *rpo30*, a vaccinia virus RNA polymerase gene with structural similarity to a eucaryotic transcription elongation factor. *Molecular and Cellular Biology* **10**, 5433–5441.
- AMEGADZIE, B. Y., HOLMES, M. H., COLE, N. B., JONES, E. V., EARL, P. L. & MOSS, B. (1991). Identification, sequence and expression of the gene encoding the second-largest subunit of the vaccinia virus DNA-dependent RNA polymerase. *Virology* **180**, 88–98.
- ARCHARD, L. C., MACKETT, M., BARNES, D. E. & DUMBELL, K. R. (1984). The genome structure of cowpox virus white pock variants. *Journal of General Virology* **65**, 875–886.
- AVES, S. J., DURKACZ, B. W., CARR, A. & NURSE, P. (1985). Cloning, sequencing and transcriptional control of the *Schizosaccharomyces pombe* *cdc10* 'start' gene. *EMBO Journal* **4**, 457–463.
- BAER, A., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHWELL, S. C., SÉGUIN, C., TUFFNELL, P. S. & BARRELL, B. G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature, London* **310**, 207–211.
- BALDICK, C. J. & MOSS, B. (1987). Resistance to rifampicin conferred by a single nucleotide substitution near the predicted NH<sub>2</sub> terminus of a gene encoding an M<sub>r</sub> 62,000 polypeptide. *Virology* **156**, 138–145.
- BALDWIN, C. T., REGINATO, A. M. & PROCKOP, D. J. (1989). A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain. *Journal of Biological Chemistry* **264**, 15747–15750.
- BALL, L. A. (1987). High-frequency homologous recombination in vaccinia virus DNA. *Journal of Virology* **61**, 1788–1795.
- BARKER, D. G., WHITE, J. H. M. & JOHNSTON, L. H. (1985). The nucleotide sequence of the DNA ligase gene (CDC9) from *Saccharomyces cerevisiae*: a gene which is cell-cycle regulated and induced in response to DNA damage. *Nucleic Acids Research* **13**, 8323–8337.
- BARKER, D. G., WHITE, J. H. M. & JOHNSTON, L. H. (1987). Molecular characterisation of the DNA ligase gene, CDC17, from the fission yeast *Schizosaccharomyces pombe*. *European Journal of Biochemistry* **162**, 659–667.
- BAROUDY, B. M. & MOSS, B. (1982). Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over. *Nucleic Acids Research* **10**, 5673–5679.
- BAROUDY, B. M., VENKATESAN, S. & MOSS, B. (1982). Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. *Cell* **28**, 315–324.
- BARTON, G. J. & STERNBERG, M. J. E. (1987). A strategy for the rapid multiple alignment of protein sequences. *Journal of Molecular Biology* **198**, 327–337.
- BELLE ISLE, H., VENKATESAN, S. & MOSS, B. (1981). Cell-free translation of early and late mRNAs selected by hybridization to cloned DNA fragments derived from the left 14 million to 72 million Daltons of the vaccinia virus genome. *Virology* **112**, 306–317.
- BERGER, A., SCHILTZ, E. & SCHULZ, G. E. (1989). Guanylate kinase from *Saccharomyces cerevisiae*. Isolation and characterization, crystallization and preliminary X-ray analysis, amino acid sequence and comparison with adenylate kinases. *European Journal of Biochemistry* **184**, 433–443.
- BERTHOLET, C., DRILLIEN, R. & WITTEK, R. (1985). One hundred base pairs of 5' flanking sequence are sufficient to temporally regulate late transcription. *Proceedings of the National Academy of Sciences, U.S.A.* **82**, 2096–2100.
- BINNS, M. M., BRITTON, B. S., MASON, C. & BOURSNEILL, M. E. G. (1990). Analysis of the fowlpox virus genome corresponding to the vaccinia virus D6 to A1 region: location of, and variation in, non-essential genes in poxviruses. *Journal of General Virology* **71**, 2873–2881.
- BLOBEL, G. (1980). Intracellular protein topogenesis. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 1496–1500.
- BONGIOVANNI, A. M. (1962). The adrenogenital syndrome with deficiency of 3 $\beta$ -hydroxysteroid (3 $\beta$ -HSD) dehydrogenase. *Journal of Clinical Investigation* **41**, 2086–2092.
- BOURSNEILL, M. E. G., FOULDS, I. J., CAMBELL, J. I. & BINNS, M. M. (1988). Non-essential genes in the vaccinia virus HindIII K fragment: a gene related to serine protease inhibitors and a gene related to the 37K vaccinia virus major envelope antigen. *Journal of General Virology* **69**, 2995–3003.
- BOYLES, S. S. & MOSS, B. (1986). Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding the 147-kDa and 22-kDa subunits. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 3141–3145.
- BULLER, R. M. L., SMITH, G. L., CREMER, K., NOTKINS, A. L. & MOSS, B. (1985). Decreased virulence of recombinant vaccinia virus expression vectors is associated with a TK negative phenotype. *Nature, London* **317**, 812–815.
- CAMPBELL, R. D., LAW, S. K. A., REID, K. B. M. & SIM, R. B. (1988). Structure, organization, and regulation of the complement genes. *Annual Review of Immunology* **6**, 161–195.
- CARRASCO, L. & ESTEBAN, M. (1982). Modification of membrane permeability in vaccinia virus-infected cells. *Virology* **117**, 62–69.
- CHEE, M. S., BANKIER, A. T., BECK, S., BOHNI, R., BROWN, C. M., CERNY, R., HORSNELL, T., HUTCHINSON, C. A., III, KOUZARIDES, T., MARTIGNETTI, J. A., PREDDIE, E., SATCHWELL, S. C., TOMLINSON, P., WESTON, K. M. & BARRELL, B. G. (1990). Analysis of the protein coding content of the human cytomegalovirus strain AD169. *Current Topics in Microbiology and Immunology* **154**, 125–169.
- CHUA, T. P., SMITH, C. E., REITH, R. W. & WILLIAMSON, J. D. (1990). Inflammatory responses and the generation of chemoattractant activity in cowpox virus infected tissues. *Immunology* **69**, 202–208.
- COLINAS, R. J., GOEBEL, S. J., DAVIS, S. W., JOHNSON, G., NORTON, E. K. & PAOLETTI, E. (1990). A DNA ligase gene in the Copenhagen strain of vaccinia virus is non-essential for viral replication and recombination. *Virology* **179**, 267–275.
- COUPAR, B. E. H., ANDREW, M. E., BOTH, G. W. & BOYLE, D. B. (1986). Temporal regulation of the influenza haemagglutinin expression in vaccinia. *European Journal of Immunology* **16**, 1479–1487.
- DAVISON, A. J. & MOSS, B. (1989). The structure of vaccinia virus late promoters. *Journal of Molecular Biology* **210**, 771–784.
- DAVISON, A. J. & SCOTT, J. E. (1986). The complete DNA sequence of varicella-zoster virus. *Journal of General Virology* **67**, 1759–1816.
- DAYHOFF, M. O., BARKER, W. C. & HUNT, L. T. (1983). Establishing homologies in protein sequences. *Methods in Enzymology* **91**, 524–545.
- DRICKAMER, K. (1988). Two distinct classes of carbohydrate-recognition domains in animal lectins. *Journal of Biological Chemistry* **263**, 9557–9560.
- DUBBS, D. R. & KIT, S. (1964). Isolation and properties of vaccinia virus mutants deficient in thymidine kinase inducing activity. *Virology* **22**, 214–225.
- EARL, P. L., JONES, E. V. & MOSS, B. (1986). Homology between DNA polymerases of poxviruses, herpesviruses, and adenoviruses: nucleotide sequence of the vaccinia virus DNA polymerase gene. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 3659–3663.
- FENNER, F. (1990). Poxviruses. In *Virology*, pp. 2113–2133. Edited by B. N. Fields. New York: Raven Press.

- GILLARD, S., SPEHNER, D., DRILLIEN, R. & KIRN, A. (1986). Localization and sequence of a vaccinia virus gene required for multiplication in human cells. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 5573–5577.
- GOEBEL, S. J., JOHNSON, G. P., PERKUS, M. E., DAVIS, S. W., WINSLOW, J. P. & PAOLETTI, E. (1990). The complete sequence of vaccinia virus. *Virology* **179**, 247–266.
- GORDON, J., KOVALA, T. & DALES, S. (1988). Molecular characterization of a prominent antigen of the vaccinia virus envelope. *Virology* **167**, 361–369.
- HILLER, G. & WEBER, K. (1982). A phosphorylated basic vaccinia virus virion polypeptide of molecular weight 11,000 is exposed on the surface of mature particles and interacts with actin-containing cytoskeletal elements. *Journal of Virology* **44**, 647–657.
- HILLER, G., WEBER, K., SCHNEIDER, L., PARAJSZ, C. & JUNGWIRTH, C. (1979). Interaction of assembled progeny pox viruses with the cellular cytoskeleton. *Virology* **98**, 142–153.
- HIRT, P., HILLER, G. & WITTEK, R. (1986). Localization and fine structure of a vaccinia virus gene encoding an envelope antigen. *Journal of Virology* **58**, 757–764.
- HOAGLAND, C. L., WARD, S. L., SMADEL, J. E. & RIVERS, T. M. (1941). Constituents of elementary bodies of vaccinia. IV. Demonstration of copper in the purified virus. *Journal of Experimental Medicine* **74**, 69–80.
- HOWARD, S. T. & SMITH, G. L. (1989). Two early vaccinia virus genes encode polypeptides related to protein kinases. *Journal of General Virology* **70**, 3187–3201.
- HOWARD, S. T., CHAN, Y. C. & SMITH, G. L. (1991). Vaccinia virus homologues of the Shope fibroma virus inverted terminal repeat proteins and a discontinuous ORF related to the tumour necrosis factor receptor family. *Virology* **180**, 633–647.
- ICHIHASHI, Y. (1977). Vaccinia-specific hemagglutinin. *Virology* **76**, 527–538.
- ICHINOSE, A., MCMULLEN, B. A., FUJIKAWA, K. & DAVIE, E. W. (1986). Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments. *Biochemistry* **25**, 4633–4638.
- IKUTA, K., TAKAMI, M., KIM, C. W., HONJO, T., MIYOSHI, T., TAGAYA, Y., KAWABE, T. & YODOI, J. (1987). Human lymphocyte Fc receptor for IgE: sequence homology of its cloned cDNA with animal lectins. *Proceedings of the National Academy of Sciences, U.S.A.* **84**, 819–823.
- INK, B. S. & PICKUP, D. J. (1990). Transcription of a poxvirus early gene is regulated both by a short promoter element and by a transcriptional termination signal controlling transcriptional interference. *Journal of Virology* **63**, 4632–4644.
- JIN, D., LI, Z., JIN, Q., YUWEN, H. & HOU, Y. (1989). Vaccinia virus hemagglutinin. A novel member of the immunoglobulin superfamily. *Journal of Experimental Medicine* **170**, 571–576.
- JOHNSON, D., LANAHAN, A., BUCK, C. R., SEHGAL, A., MORGAN, C., MERCER, E., BOTHWELL, M. & CHAO, M. (1986). Expression and structure of the human NGF receptor. *Cell* **47**, 545–554.
- JONG, A. Y., KUO, C.-I. & CAMPBELL, J. L. (1984). The *cdc8* gene of yeast encodes thymidylate kinase. *Journal of Biological Chemistry* **259**, 11052–11059.
- KERR, S. M. & SMITH, G. L. (1989). Vaccinia virus encodes a polypeptide with DNA ligase activity. *Nucleic Acids Research* **17**, 9039–9050.
- KERR, S. M. & SMITH, G. L. (1991). Vaccinia virus DNA ligase is non-essential for virus replication: recovery of plasmids from virus-infected cells. *Virology* **180**, 625–632.
- KOTWAL, G. J. & MOSS, B. (1988a). Analysis of a large cluster of non-essential genes deleted from a vaccinia virus terminal transposition mutant. *Virology* **167**, 524–537.
- KOTWAL, G. J. & MOSS, B. (1988b). Vaccinia virus encodes a secretory polypeptide structurally related to complement control proteins. *Nature, London* **335**, 176–178.
- KOTWAL, G. J. & MOSS, B. (1989). Vaccinia virus encodes two proteins that are structurally related to the plasma serine protease inhibitor superfamily. *Journal of Virology* **63**, 600–606.
- KOTWAL, G. J., ISAACS, S. N., MCKENZIE, R., FRANK, M. M. & MOSS, B. (1990). Inhibition of the complement cascade by the major secretory protein of vaccinia virus. *Science* **250**, 827–830.
- KWIATKOWSKI, D. J. & BRUNS, G. A. P. (1988). Human profilin: molecular cloning, sequence comparison, and chromosomal analysis. *Journal of Biological Chemistry* **263**, 5910–5915.
- LAKE, J. R. & COOPER, P. D. (1980). Deletions of the terminal sequences in the genomes of the white pox (*u*) and host-restricted (*p*) mutants of rabbitpox virus. *Journal of General Virology* **48**, 135–147.
- LEE-CHEN, G.-J. & NILES, E. G. (1988). Transcriptional and translational mapping of the 13 genes in the vaccinia virus *HindIII* D fragment. *Virology* **163**, 52–63.
- LUX, S. E., JOHN, K. M. & BENNET, V. (1990). Analysis of the cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature, London* **344**, 36–42.
- MCCORD, J. M. & FRIDOVICH, I. (1969). Superoxide dismutase. An enzyme function for erthrocyte (hemocuprein). *Journal of Biological Chemistry* **244**, 6049–6055.
- MCGEOCH, D. J. (1990). Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Research* **18**, 4105–4110.
- MCGEOCH, D. J., DALRYMPLE, M. A., DAVISON, A. J., DOLAN, A., FRAME, M. C., McNAB, D., PERRY, L. J., SCOTT, J. E. & TAYLOR, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *Journal of General Virology* **69**, 1531–1574.
- MACKETT, M. & ARCHARD, L. C. (1979). Conservation and variation in orthopoxvirus genome structure. *Journal of General Virology* **45**, 683–701.
- MALLETT, S., FOSSUM, S. & BARCLAY, A. N. (1990). Characteristics of the MRC OX40 antigen of activated CD4 positive T lymphocytes — a molecule related to nerve growth factor receptor. *EMBO Journal* **9**, 1063–1068.
- MEYUHAS, O. & KLEIN, A. (1990). The ribosomal protein L7 gene. Its primary structure and functional analysis of the promoter region. *Journal of Biological Chemistry* **265**, 11465–11473.
- MOSS, B. (1990). Poxviridae and their replication. In *Virology*, pp. 2079–2111. Edited by B. N. Fields. New York: Raven Press.
- MOYER, R. W., GRAVES, R. L. & ROTHE, C. T. (1980). The white pox (*u*) mutants of rabbit poxvirus. III. Terminal DNA sequence duplication and transposition in rabbit poxvirus. *Cell* **22**, 545–553.
- NG, N. F., TRINH, K.-Y. & HEW, C. L. (1986). Structure of an antifreeze polypeptide precursor from the sea raven, *Hemitriperus americanus*. *Journal of Biological Chemistry* **261**, 15690–15695.
- NILES, E. G., CONDIT, R. C., CARO, P., DAVIDSON, K., MATUSICK, L. & SETO, J. (1986). Nucleotide sequence and genetic map of the 16-kb vaccinia virus *HindIII* D fragment. *Virology* **153**, 96–112.
- NODA, M., IKEDA, T., KAYANO, T., SUZUKI, H., TAKESHIMA, H., KURASAKI, M., TAKAHASHI, H. & NUMA, S. (1986). Existence of sodium channel messenger RNAs in rat brain. *Nature, London* **320**, 188–192.
- NYSTROM, L.-E., LINDBERG, U., KENDRICK-JONES, J. & JAKES, R. (1979). The amino acid sequence of profilin from calf spleen. *FEBS Letters* **101**, 161–165.
- OHNO, H., TAKIMOTO, G. & MCKEITHAN, T. W. (1990). The candidate proto-oncogene *bcl-3* is related to genes implicated in cell lineage determination and cell cycle control. *Cell* **60**, 991–997.
- ONO, Y., FUJII, T., OGITA, K., KIKKAWA, U., IGARASHI, K. & NISHIZUKA, Y. (1988). The structure, expression, and properties of additional members of the protein kinase C family. *Journal of Biological Chemistry* **263**, 6927–6932.
- PALUMBO, G. J., PICKUP, D. J., FREDRICKSON, T. N., MCINTYRE, L. J. & BULLER, R. M. L. (1989). Inhibition of an inflammatory response is mediated by a 38-kDa protein of cowpox virus. *Virology* **172**, 262–273.
- PAYNE, L. G. & NORRBY, E. (1976). Presence of haemagglutinin in the envelope of extracellular vaccinia virus particles. *Journal of General Virology* **32**, 63–72.
- PEARSON, W. R. & LIPMAN, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, U.S.A.* **85**, 2444–2448.
- PICKUP, D. J., INK, B. S., PARSONS, B. L., HU, W. & JOKLIK, W. K. (1984). Spontaneous deletions and duplications of sequences in the

- genome of cowpox virus. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 6817-6821.
- PICKUP, D. J., INK, B. S., HU, W., RAY, C. A. & JOKLIK, W. K. (1986). Hemorrhage in lesions caused by cowpox virus is induced by a viral protein that is related to plasma protein inhibitors of serine proteases. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 7698-7702.
- PLUCIENNICZAK, A., SCHROEDER, E., ZETTLMEISS, G. & STREECK, R. E. (1985). Nucleotide sequence of a cluster of early and late genes in a conserved segment of the vaccinia virus genome. *Nucleic Acids Research* **13**, 985-998.
- QUERY, C. C., BENTLEY, R. C. & KEENE, J. D. (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**, 89-101.
- REMPEL, R. E., ANDERSON, M. K., EVANS, E. & TRAKTMAN, P. (1990). Temperature-sensitive vaccinia virus mutants identify a gene with an essential role in viral replication. *Journal of Virology* **64**, 574-583.
- RODRIGUEZ, J. F. & ESTEBAN, M. (1987). Mapping and nucleotide sequence of the vaccinia virus gene that encodes a 14-kilodalton fusion protein. *Journal of Virology* **61**, 3550-3554.
- RODRIGUEZ, J. F., KAHN, J. S. & ESTEBAN, M. (1986). Molecular cloning, encoding sequence, and expression of vaccinia virus nucleic acid-dependent nucleoside triphosphatase gene. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 9566-9570.
- ROSEL, J. & MOSS, B. (1985). Transcriptional and translational mapping of a vaccinia virus gene encoding the precursor of the major core antigen 4b. *Journal of Virology* **56**, 830-838.
- ROSEL, J. L., EARL, P. L., WEIR, J. P. & MOSS, B. (1986). Conserved TAAATG sequence at the transcriptional and translational initiation sites of vaccinia virus late genes deduced by structural and functional analysis of the HindIII H fragment. *Journal of Virology* **60**, 436-449.
- ROSEMAN, N. A. & HRUBY, D. E. (1987). Nucleotide sequence and transcript organization of a region of the vaccinia virus genome which encodes a constitutively expressed gene required for DNA replication. *Journal of Virology* **61**, 1398-1406.
- ROSEMAN, N. A. & SLABAUGH, M. B. (1990). The vaccinia virus HindIII F fragment: nucleotide sequence of the left 6.2 kb. *Virology* **178**, 410-418.
- SAI, S., TANAKA, T., KOSHER, R. A. & TANZER, M. L. (1986). Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 5081-5085.
- SCHININA, M. E., BARRA, D., GENTILOMO, S., BOSSA, F., CAPO, C., ROTILIO, G. & CALABRESE, L. (1986). Primary structure of a cationic Cu,Zn superoxide dismutase. The sheep enzyme. *FEBS Letters* **207**, 7-10.
- SCHMITT, J. F. C. & STUNNENBERG, H. G. (1989). Sequence and transcriptional analysis of the vaccinia virus HindIII I fragment. *Journal of Virology* **62**, 1889-1897.
- SCHULZ, T. F., SCHWAEBLE, W., STANLEY, K. K., WEISS, E. & DIERICH, M. P. (1986). Human complement factor H: isolation of cDNA clones and partial cDNA sequence of the 38-kDa tryptic fragment containing the binding site for C3b. *European Journal of Immunology* **16**, 1351-1355.
- SHIDA, H. (1986). Nucleotide sequence of the vaccinia virus hemagglutinin gene. *Virology* **150**, 451-462.
- SHUMAN, S., GOLDER, M. & MOSS, B. (1989). Insertional mutagenesis of the vaccinia virus gene encoding a type I DNA topoisomerase: evidence that the gene is essential for virus growth. *Virology* **170**, 302-306.
- SIMS, J. E., MARCH, C. J., COSMAN, D., WIDMER, M. B., MACDONALD, H. R., MCMAHAN, C. J., GRUBIN, C. E., WIGNALL, J. M., JACKSON, J. L., CALL, S. M., FRIEND, D., ALPERT, A. R., GILLIS, S., URDAL, D. L. & DOWER, S. K. (1988). cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily. *Science* **241**, 585-589.
- SIMS, J. E., ACRES, R. B., GRUBIN, C. E., MCMAHAN, C. J., WIGNALL, J. M., MARCH, C. J. & DOWER, S. K. (1989). Cloning the interleukin 1 receptor from human T cells. *Proceedings of the National Academy of Sciences, U.S.A.* **86**, 8946-8950.
- SLABAUGH, M. & ROSEMAN, N. (1989). Retroviral protease-like gene in the vaccinia virus genome. *Proceedings of the National Academy of Sciences, U.S.A.* **86**, 4152-4155.
- SLABAUGH, M. B., JOHNSON, T. L. & MATHEWS, C. K. (1984). Vaccinia virus induces ribonucleotide reductase in primate cells. *Journal of Virology* **52**, 507-514.
- SLABAUGH, M., ROSEMAN, N., DAVIS, R. & MATHEWS, C. (1988). Vaccinia virus-encoded ribonucleotide reductase: sequence conservation of the gene for the small subunit and its amplification in hydroxyurea-resistant mutants. *Journal of Virology* **62**, 519-527.
- SMITH, C. A., DAVIS, T., ANDERSON, D., SOLAM, L., BECKMANN, M. P., JERZY, R., DOWER, S. K., COSMAN, D. & GOODWIN, R. G. (1990). A receptor for tumor necrosis factor defines an unusual family of cellular and viral proteins. *Science* **248**, 1019-1023.
- SMITH, G. L. & CHAN, Y. C. (1991). Two vaccinia virus proteins structurally related to the interleukin-1 receptor and the immunoglobulin superfamily. *Journal of General Virology* **72**, 511-518.
- SMITH, G. L., HOWARD, S. T. & CHAN, Y. C. (1989a). Vaccinia virus encodes a family of genes with homology to serine protease inhibitors. *Journal of General Virology* **70**, 2333-2343.
- SMITH, G. L., CHAN, Y. S. & KERR, S. M. (1989b). Transcriptional mapping and nucleotide sequence of a vaccinia virus gene with extensive homology to yeast DNA ligase. *Nucleic Acids Research* **17**, 9050-9061.
- SMITH, G. L., DE CARLOS, A. & CHAN, Y. S. (1989c). Vaccinia virus encodes a thymidylate kinase gene: sequence and transcriptional mapping. *Nucleic Acids Research* **17**, 7581-7590.
- SPEHNER, D., GILLARD, S., DRILLIEN, R. & KIRN, A. (1988). A cowpox virus gene required for multiplication in Chinese hamster ovary cells. *Journal of Virology* **62**, 1297-1304.
- SPEHNER, D., DRILLIEN, R. & LECOCQ, J. P. (1990). Construction of fowlpox virus vectors with intergenic insertions: expression of the beta-galactosidase gene and the measles virus fusion gene. *Journal of Virology* **64**, 527-533.
- SPIESS, M. & LODISH, H. F. (1985). Sequence of a second human asialoglycoprotein receptor: conservation of two receptor genes during evolution. *Proceedings of the National Academy of Sciences, U.S.A.* **82**, 6465-6469.
- SPYROPOULOS, D. D., ROBERT, B. E., PANICALI, D. L. & COHEN, L. K. (1988). Delineation of the viral products of recombination in vaccinia virus-infected cells. *Journal of Virology* **62**, 1046-1054.
- SRI WIDADA, J., FERRAZ, C. & LIAUTARD, J. P. (1989). Total coding sequence of profilin from *Mus musculus* macrophage. *Nucleic Acids Research* **17**, 2855.
- STADEN, R. (1982). Automation of computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731-4751.
- STEHLE, T. & SCHULZ, G. E. (1990). Three-dimensional structure of the complex of guanylate kinase from yeast with its substrate GMP. *Journal of Molecular Biology* **211**, 249-254.
- STEINMAN, H. M., NAIK, V. R., ABERNETHY, J. L. & HILL, R. L. (1974). Bovine erythrocyte superoxide dismutase. Complete amino acid sequence. *Journal of Biological Chemistry* **249**, 7326-7338.
- TAINER, J. A., GETZOFF, E. D., RICHARDSON, J. S. & RICHARDSON, D. C. (1983). Structure and mechanism of copper, zinc superoxide dismutase. *Nature, London* **306**, 284-287.
- TAKIO, K., BLUMENTHAL, D. K., WALSH, K. A., TITANI, K. & KREBS, E. G. (1986). Amino acid sequence of rabbit skeletal muscle myosin light chain kinase. *Biochemistry* **25**, 8049-8057.
- TAMIN, A., VILLARREAL, E. C., WEINRICH, S. L. & HRUBY, D. E. (1988). Nucleotide sequence and molecular genetic analysis of the vaccinia virus HindIII N/M region encoding the genes responsible for resistance to alpha-amanitin. *Virology* **165**, 141-150.
- TENGELSEN, L. A., SLABAUGH, M. B., BIBLER, J. K. & HRUBY, D. E. (1988). Nucleotide sequence and molecular genetic analysis of ribonucleotide reductase encoded by vaccinia virus. *Virology* **164**, 121-131.
- THE, V. L., LACHANCE, Y., LABRIE, C., LEBLANC, G., THOMAS, J., STRICKLER, R. C. & LABRIE, F. (1989). Full length cDNA structure and deduced amino acid sequence of human  $3\beta$ -hydroxy-5-ene steroid dehydrogenase. *Molecular Endocrinology* **3**, 1310-1312.

- TOMLEY, F., BINNS, M., CAMPBELL, J. & BOURSNELL, M. (1988). Sequence analysis of an 11·2 kilobase, near-terminal, *Bam*HI fragment of fowlpox virus. *Journal of General Virology* **69**, 1025–1040.
- TOWNSEND, A. R. M., BASTIN, J., GOULD, K., BROWNLEE, G. G., ANDREW, A., BOYLE, D. B., CHAN, Y. S. & SMITH, G. L. (1988). Defective presentation to class I restricted CTL in vaccinia infected cells is overcome by enhanced degradation of antigen. *Journal of Experimental Medicine* **168**, 1211–1224.
- TRAKTMAN, P., ANDERSON, M. K. & REMPEL, R. E. (1989). Vaccinia virus encodes an essential gene with strong homology to protein kinases. *Journal of Biological Chemistry* **264**, 21458–21461.
- TSOA, H., REN, G. F. & CHU, C. M. (1986). Gene coding for the 11,000-Dalton polypeptide of the Tian Tan strain of vaccinia virus and its 5' flanking region: nucleotide sequence. *Journal of Virology* **57**, 693–696.
- UEDA, Y., MORIKAWA, S. & MATSUURA, Y. (1990). Identification and nucleotide sequence of the gene encoding a surface antigen induced by vaccinia virus. *Virology* **177**, 588–594.
- UPTON, C., DELANGE, A. M. & MCFADDEN, G. (1987). Tumorigenic poxviruses: genomic organization and DNA sequence of the terminal inverted repeats of Shope fibroma virus. *Virology* **160**, 20–30.
- UPTON, C., MACEN, J. L., MARANCHUK, R. A., DELANGE, A. M. & MCFADDEN, G. (1988). Tumorigenic poxviruses: fine analysis of the recombination junctions in malignant rabbit fibroma virus, a recombinant between Shope fibroma virus and myxoma virus. *Virology* **166**, 229–239.
- VAN MEIR, E. & WITTEK, R. (1988). Fine structure of the vaccinia virus gene encoding the precursor of the major core protein 4a. *Archives of Virology* **102**, 19–27.
- VENKATESAN, S., BAROUDY, B. & MOSS, B. (1981). Distinctive nucleotide sequences adjacent to multiple initiation and termination sites within an early vaccinia virus gene. *Cell* **25**, 805–813.
- VENKATESAN, S., GERSHOWITZ, A. & MOSS, B. (1982). Complete nucleotide sequences of two adjacent early vaccinia virus genes located within the inverted terminal repetition. *Journal of Virology* **44**, 637–646.
- WEINRICH, S. L. & HRUBY, D. E. (1986). A tandemly-oriented late gene cluster within the vaccinia virus genome. *Nucleic Acids Research* **14**, 3003–3016.
- WEIR, J. P. & MOSS, B. (1983). Nucleotide sequence of the vaccinia virus thymidine kinase gene and the nature of spontaneous frameshift mutations. *Journal of Virology* **46**, 530–537.
- WEIR, J. P. & MOSS, B. (1984). Regulation of expression and nucleotide sequence of a late vaccinia virus gene. *Journal of Virology* **51**, 662–669.
- WITTEK, R., MENNA, A., SCHÜMPERLI, D., STOFFEL, S., MÜLLER, H. K. & WYLER, R. (1977). *Hind*III and *Sst*I restriction sites mapped on rabbit poxvirus and vaccinia virus DNA. *Journal of Virology* **23**, 669–678.
- WITTEK, R., BARBOSA, E., COOPER, J. A., GARON, C. F., CHAN, H. & MOSS, B. (1980). Inverted terminal repetition in vaccinia virus DNA encodes early mRNAs. *Nature, London* **285**, 21–25.
- YAMASAKI, K., TAGA, T., HIRAT, Y., YAWATA, H., KAWANISHI, Y., SEED, B., TANIGUCHI, T., HIRANO, T. & KISHIMOTO, T. (1988). Molecular structure of interleukin 6 receptor. *Proceedings of the Japanese Academy* **64**, 209–211.
- YUEN, L. & MOSS, B. (1987). Oligonucleotide sequence signalling transcriptional termination of vaccinia virus early genes. *Proceedings of the National Academy of Sciences, U.S.A.* **84**, 6417–6421.
- ZHAO, H. F., SIMARD, J., LABRIE, C., BRETON, N., RHEAUME, E., LUU-THIE, V. L. & LABRIE, F. (1989). Molecular cloning, cDNA structure and predicted amino acid sequence of bovine 3 beta-hydroxy-5-ene steroid dehydrogenase/delta 5-delta 4 isomerase. *FEBS Letters* **259**, 153–157.
- ZHOU, J., CRAWFORD, L., MCLEAN, L., SUN, X.-Y., STANLEY, M., ALMOND, N. & SMITH, G. L. (1990). Increased antibody responses to human papillomavirus type 16 L1 protein expressed from recombinant vaccinia virus lacking serine protease inhibitor genes. *Journal of General Virology* **71**, 2185–2190.

---

(Received 8 February 1991; Accepted 1 March 1991)