Volume 13 Number 21 1985 Nucleic Acids Research

DNA sequence of the herpes simplex virus type 1 gene whose product is responsible for transcriptional activation of immediate early promoters

M.A.Dalrymple, D.J.McGeoch+, A.J.Davison+ and C.M.Preston+

Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, UK

Received 29 August 1985; Accepted 2 October 1985

ABSTRACT

Previous work has shown that transcriptional activation of herpes simplex virus type 1 (HSV-1) immediate early genes is mediated by a protein species (Vmw65) present in the tegument of infecting virions. This paper describes DNA sequence analysis and mRNA mapping of the Vmw65 gene in HSV-1 strain 17. The Vmw65 coding region was identified as a 490 codon sequence encoding a polypeptide of molecular weight 54,342 and characterised by a high proportion of charged amino acid residues. A homologue to Vmw65 was detected in the genome of varicella-zoster virus, another human herpesvirus. Apart from its role in trans-activation, Vmw65 is a major constituent of the virion. Its possible significance in virus structure is discussed.

INTRODUCTION

Herpes simplex virus type 1 (HSV-1) is a large, enveloped, icosahedral virus (1). It has a linear, double-stranded DNA genome of approximately 155 kilobase pairs (kbp) (2, 3) which is transcribed and replicated in the cell nucleus. At least three classes of genes have been distinguished in a temporal cascade, based on their requirements for de novo protein synthesis and viral DNA replication before transcriptional activation (4, 5, 6, 7). These classes have been named immediate early (IE), early (E) and late (L). The IE group consists of five genes named IE1(2 copies), 2, 3 (2 copies), 4 and 5. These genes are the first to be expressed after infection and do not require prior viral protein synthesis for their transcription (4, 5). Polypeptide Vmw175, the product of IE3, is required continuously for the expression of all E and L genes (8, 9). There is growing evidence that other IE proteins are involved in modulating transcription of the later temporal classes, probably in conjunction with Vmw175 (10, 11, 12).

The 5' flanking sequences which control expression of HSV-1

© IRL Press Limited, Oxford, England.

IE genes have been studied by several groups. In particular, IE3 and 4/5 (IE mRNAs 4 and 5 have identical 5' termini and upstream sequences) possess efficient promoters, including upstream regions with enhancer-like activity (13, 14, 15, 16, 17, 18, 19). Furthermore, transcription from all IE genes is strongly stimulated by a structural component of the herpes simplex virion (13, 15, 18, 20). Functional analysis of the IE gene regulatory regions (18, 21) located an AT rich, _cis_-acting, DNA sequence essential for the stimulatory response. This element is represented by a consensus sequence TAATGARATTC (R=purine) present in one or more copies far upstream of all IE genes, both in HSV-1 and HSV-2 (14, 20, 22, 23, 24). Thus, the available evidence suggests that stimulation of HSV IE gene transcription is mediated by a virion component which interacts, either directly or indirectly, with specific DNA sequences located far upstream from the mRNA 5' terminus.

Studies by Batterson and Roizman (25) have shown that the virion component is _trans_-acting and probably consists of one or more polypeptides located outside the nucleocapsid. Campbell, Palfreyman and Preston (26) applied a novel strategy to identify the HSV-1 gene or genes encoding the specific stimulatory factor. Cloned restriction fragments of the HSV-1 genome were transfected into tissue culture cells, together with a plasmid containing the HSV-1 thymidine kinase (TK) gene under the control of the IE3 promoter. Genomic fragments which stimulated the levels of TK activity were identified. Initial experiments located the IE specific _trans_-activator to the EcoRI I region (map coordinates 0.635-0.721). By a process of subcloning and insertion mutagenesis the authors were able to exclude all mRNAs known to map in EcoRI I except one (27). Hybrid arrested _in vitro_ translation and immunoprecipitation studies identified the gene responsible as that for Vmw65, a major virion tegument phosphoprotein (28, 29). The tegument is an amorphous layer of protein, which lies between the viral nucleocapsid and envelope, of unknown function.

We report here the complete nucleotide sequence of the gene encoding Vmw65. The gene maps to a 2.6kbp fragment located in the long unique region of the genome, between map coordinates

0.669 and 0.685. In this region there is only one potential open reading frame consistent with our mapping data and analysis of codon usage. Furthermore, the 5' and 3' flanking regions exhibit typical promoter and termination sequence elements.

MATERIALS AND METHODS

Plasmids

The sequence reported here is derived from the sequence of the BamHI F fragment of HSV-1 strain 17, cloned into the BamHI site of pAT153 (30) and named pGX158. The plasmid was propagated in *Escherichia coli* DH1 (31). Plasmid DNA was prepared by established techniques and banded by CsCl gradient ultracentrifugation (32).

DNA Sequence Analysis

The dideoxy sequencing methodology was used (33, 34). Plasmid pGX158 was digested with BamHI and the vector sequences separated by agarose gel electrophoresis. The insert fragment was electroeluted from the gel slice. Purified fragment was self-ligated,using T4 DNA ligase, and then sonicated using a Dawe probe sonicator (35). Sonicated DNA was end repaired by incubation with T4 DNA polymerase and all four deoxynucleoside triphosphates. The end repaired fragments were then size fractionated by a second round of gel electrophoresis and electroelution. Purified fragments in the size range 300-600 base pairs were cloned into SmaI cleaved and phosphatase-treated M13mp8 vector (36). Clones were transfected into *E.coli* JM101, suitable plaques picked and single stranded template isolated and sequenced.

Computing

The sequence data were managed with the database system of Staden (37), implemented by Dr.P.Taylor on a DEC PDP 11/44 machine running under the RSX-11M operating system.

Open reading frames were identified using the program of Blumenthal et al (38) and codon usage evaluated by the method of Staden and McLachlan (39). Amino acid sequence hydrophobicities were examined with the parameters of Kyte and Doolittle (40). Sequence homologies were assessed by use of a matrix comparison program (41) and an optimal alignment program (42).

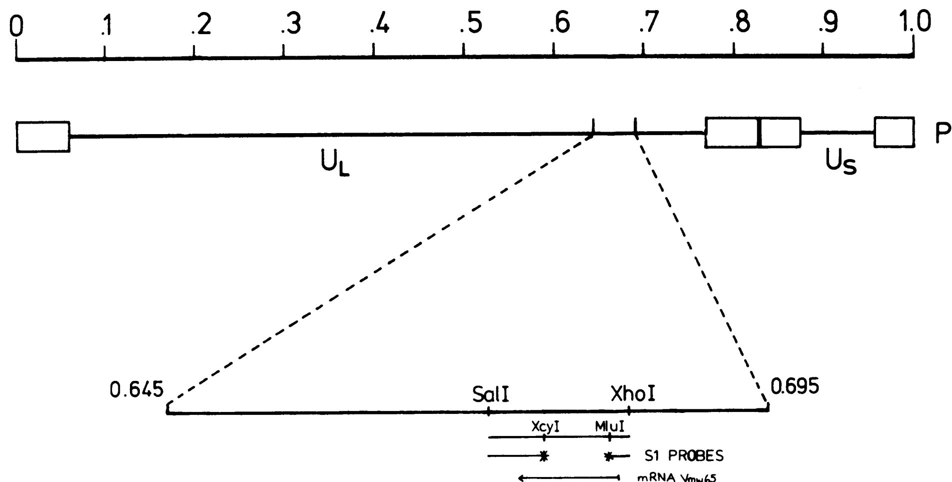

Figure 1: Location of the BamHI F fragment in the HSV-1 genome

A diagrammatic representation of the HSV-1 prototype genome. U_L and U_S are the long and short unique regions of the genome. The open boxes represent repeated sequences. BamHI F is located between 0.645 and 0.695 map units. The sequence presented is marked as a line between SalI and XhoI sites. Also shown are the positions of the end-labelled probes used to determine the 5' and 3' termini of the mRNA for Vmw65.

mRNA Mapping

Total cytoplasmic RNA was extracted from BSC-1 cells at 6h after infection with 20 plaque forming units of wild-type HSV-1 strain 17 per cell, or mock infected, as described by Preston (43).

To map the 5' terminus of the mRNA, 10ug of RNA was hybridised with a 5' end-labelled MluI or SalI fragment derived from the plasmid pMC1 (26), a 2.6kb sub-clone of pGX158 which contains the complete coding and flanking sequences of the Vmw65 gene and whose sequence is reported here. The 3' end of the mRNA was mapped by hybridisation to a 3' end labelled XcyI fragment. The respective probes are shown in Figure 1. The full procedure has been described previously (44).

RESULTS

Location of the Gene for Vmw65

The gene for Vmw65 lies between map coordinates 0.669 and

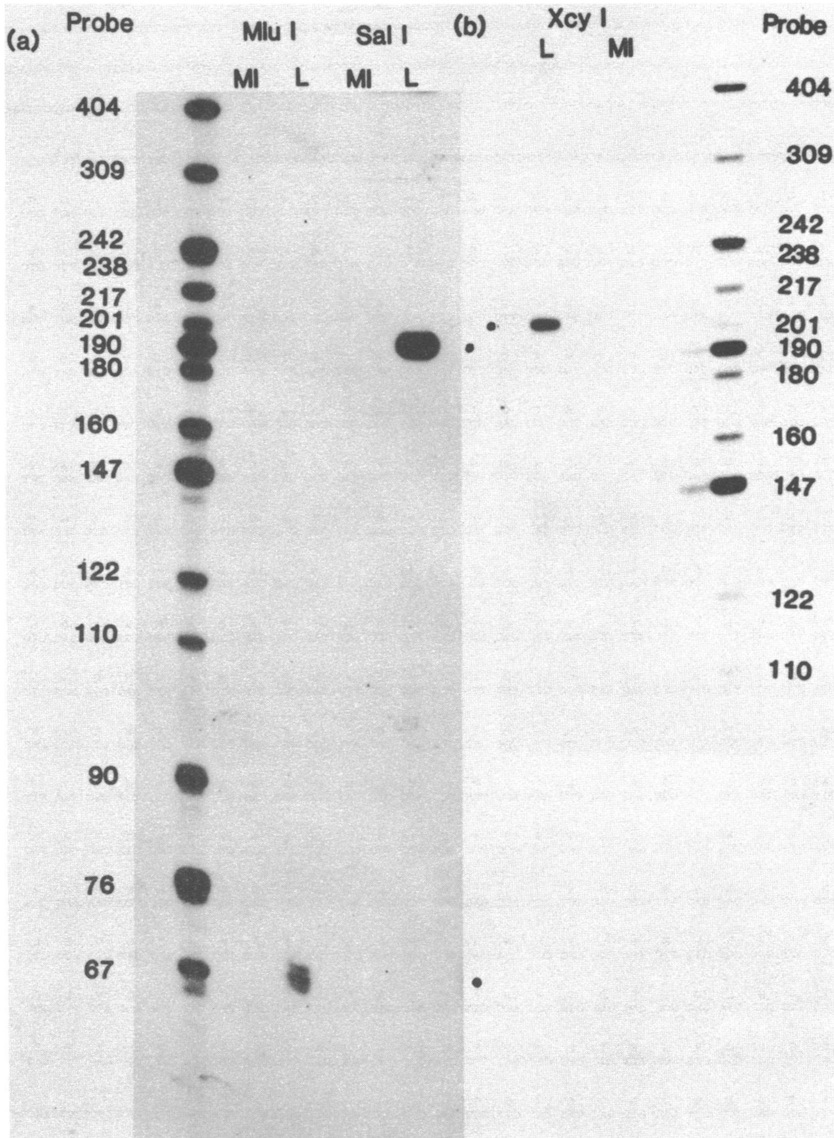

Figure 2: S1 mapping of the 5' and 3' termini of the mRNA for Vmw65

Panel (a) shows the results of using 5' end-labelled restriction fragments to map the mRNA initiation site. The probes used are indicated above the lanes. The hybrid bands obtained are discussed in the text.

Panel (b) shows an equivalent experiment to map the 3' terminus of the message. In both panels MI stands for mock-infected RNA and L is late, infected-cell, mRNA.

```
GAGCCCCAGCCCGCTCCGCTTCTCGCCCCAGACGGCCCGTCGAGTGAAAACTTCCGTACCCAGACAATAAAGCACCAACAGGGGTTCATTCGGTGTTGGCGTTGCGTGCCTTTG  114

TTTCCCAATCCGACGGGGACCGGGACTGGGTGGCGGGGGGTGGGTTGGACAGCCGCCCTCGGTTCGCCTTCACGTGACAGGAGCCAATGTGGGGGGAAGTCACGAGGTACGGGG  228

                                                 O--------->
CGGCCCGTGCGGGTTGCTTAAATGCGTGGTGGCGACCACGGGCTGTCATTCCTCGGGAACGGACGGGGTTCCCGCTGCCCACTTCCCCCCATAAGGTCCGTCCGGTCCTCTAAC  342

                                                                                                                     0
GCGTTTGGGGGTTTTCTCTTCCCGCGCCGTCGGGCGTCCCACACTCTCTGGGCGGGCGGGGACGATCGCATCAAAAGCCCGATATCGTCTTTCCCGTATCAACCCCACCCA  453

 M   D   L   L   V   D   E   L   F   A   D   M   N   A   D   G   A   S   P   P   P   P   R   P   A   G   G   P     28
ATG GAC CTC TTG GTC GAC GAG CTG TTT GCC GAC ATG AAC GCG GAC GGC GCT TCG CCA CCG CCC CCC CGC CCG GCC GGG GGT CCC    537

 K   N   T   P   A   A   P   P   L   Y   A   T   G   R   L   S   Q   A   Q   L   M   P   S   P   P   M   P   V     56
AAA AAC ACC CCG GCG GCC CCC CCG CTG TAC GCA ACG GGG CGC CTG AGC CAG GCC CAG CTC ATG CCC TCC CCA CCC ATG CCC GTC    621

 P   P   A   A   L   F   N   R   L   L   D   D   L   G   F   S   A   G   P   A   L   C   T   M   L   D   T   W     84
CCC CCC GCC GCC CTC TTT AAC CGT CTC CTC GAC GAC TTG GGC TTT AGC GCG GGC CCC GCG CTA TGT ACC ATG CTC GAT ACC TGG    705

 N   E   D   L   F   S   A   L   P   T   N   A   D   L   Y   R   E   C   K   F   L   S   T   L   P   S   D   V    112
AAC GAG GAT CTG TTT TCG GCG CTA CCG ACC AAC GCC GAC CTG TAC CGG GAG TGT AAA TTC CTA TCA ACG CTG CCC AGC GAT GTG    789

 V   E   W   G   D   A   Y   V   P   E   R   T   Q   I   D   I   R   A   H   G   D   V   A   F   P   T   L   P    140
GTG GAA TGG GGG GAC GCG TAC GTC CCC GAA CGC ACC CAA ATC GAC ATT CGC GCC CAC GGC GAC GTG GCC TTC CCT ACG CTT CCG    873

 A   T   R   D   G   L   G   L   Y   Y   E   A   L   S   R   F   F   H   A   E   L   R   A   R   E   E   S   Y    168
GCC ACC CGC GAC GGC CTC GGG CTC TAC TAC GAA GCG CTC TCT CGT TTC TTC CAC GCC GAG CTA CGG GCG CGG GAG GAG AGC TAT    957

 R   T   V   L   A   N   F   C   S   A   L   Y   R   Y   L   R   A   S   V   R   Q   L   H   R   Q   A   H   M    196
CGA ACC GTG TTG GCC AAC TTC TGC TCG GCC CTG TAC CGG TAC CTG CGC GCC AGC GTC CGG CAG CTG CAC CGC CAG GCG CAC ATG   1041

 R   G   R   D   R   D   L   G   E   M   L   R   A   T   I   A   D   R   Y   Y   R   E   T   A   R   L   A   R    224
CGC GGA CGC GAT CGC GAC CTG GGA GAA ATG CTG CGC GCC ACG ATC GCG GAC AGG TAC TAC CGA GAG ACC GCT CGT CTG GCG CGT   1125

 V   L   F   L   H   L   Y   L   F   L   T   R   E   I   L   W   A   A   Y   A   E   Q   M   M   R   P   D   L    252
GTT TTG TTT TTG CAT TTG TAT CTA TTT TTG ACC CGC GAG ATC CTA TGG GCC GCG TAC GCC GAG CAG ATG ATG CGG CCC GAC CTG   1209

 F   D   C   L   C   C   D   L   E   S   W   R   Q   L   A   G   L   F   Q   P   F   M   F   V   N   G   A   L    280
TTT GAC TGC CTC TGT TGC GAC CTG GAG AGC TGG CGT CAG TTG GCG GGT CTG TTC CAG CCC TTC ATG TTC GTC AAC GGA GCG CTC   1293

 T   V   R   G   V   P   I   E   A   R   R   L   R   E   L   N   H   I   R   E   H   L   N   L   P   L   V   R    308
ACC GTC CGG GGA GTG CCA ATC GAG GCC CGC CGG CTG CGG GAG CTA AAC CAC ATT CGC GAG CAC CTT AAC CTC CCG CTG GTG CGC   1377

 S   A   A   T   E   E   P   G   A   P   L   T   T   P   P   T   L   H   G   N   Q   A   R   A   S   G   Y   F    336
AGC GCG GCT ACG GAG GAG CCA GGG GCG CCG TTG ACG ACC CCT CCC ACC CTG CAT GGC AAC CAG GCC CGC GCC TCT GGG TAC TTT   1461

 M   V   L   I   R   A   K   L   D   S   Y   S   S   F   T   T   S   P   S   E   A   V   M   R   E   H   A   Y    364
ATG GTG TTG ATT CGG GCG AAG TTG GAC TCG TAT TCC AGC TTC ACG ACC TCG CCC TCC GAG GCG GTC ATG CGG GAA CAC GCG TAC   1545

 S   R   A   F   T   K   N   N   Y   G   S   T   I   E   G   L   L   D   L   P   D   D   D   A   P   E   E   A    392
AGC CGC GCG CGT ACG AAA AAC AAT TAC GGG TCT ACC ATC GAG GGC CTG CTC GAT CTC CCG GAC GAC GAC GCC CCC GAA GAG GCG   1629

 G   L   A   A   P   R   L   S   F   L   P   A   G   H   T   R   R   L   S   T   A   P   P   T   D   V   S   L    420
GGG CTG GCG GCT CCG CGC CTG TCC TTT CTC CCC GCG GGA CAC ACG CGC AGA CTG TCG ACG GCC CCC CCG ACC GAT GTC AGC CTG   1713

 G   D   E   L   H   L   D   G   E   D   V   A   M   A   H   A   D   A   L   D   D   F   D   L   D   M   L   G    448
GGG GAC GAG CTC CAC TTA GAC GGC GAG GAC GTG GCG ATG GCG CAT GCC GAC GCG CTA GAC GAT TTC GAT CTG GAC ATG TTG GGG   1797

 D   G   D   S   P   G   P   G   F   T   P   H   D   S   A   P   Y   G   A   L   D   M   A   D   F   E   F   E    476
GAC GGG GAT TCC CCG GGG CCG GGA TTT ACC CCC CAC GAC TCC GCC CCC TAC GGC GCT CTG GAT ATG GCC GAC TTC GAG TTT GAG   1881

 Q   M   F   T   D   A   L   G   I   D   E   Y   G   G   -                                                         491
CAG ATG TTT ACC GAT GCC CTT GGA ATT GAC GAG TAC GGT GGG TAG GGGGCGCGACCGGACCCGCATCCCCCGTCTGGGTTTTCCCCTCCCGTCACCGGT  1980

                                     --------X
TCGTATCCACAATAAACACGAGCACATACATTACAAAACCTGCGGTTGTCGTCTGATTATTTGGTGGTGGGGGAAAGAACTAGCCAGGAGACGGGACCGCGCAACCAACCCACT  2094

GGGGTCTGGGTTGCCGGCGTGTGTGTTAGCCGCGTCTGCGGGCCTGTCGTGTAGATTCGAAACCACGGACGGGTGATTGTGTCGCAGGGCGGCCCGCGTATAAAGGCGAGAGCG  2208

CGGGACCGTTTCCGCATTTGGCCGGGGGCTGGGGCGGCGGGTAGCCTTCGCGGGAGATACTGCGTTTTTTTGCGCCGGCCCCGTCGCTCCCGTCCATTCCCATCGCGAGGGGGT  2322

CCGGCGGCACCTACCCCGGCCTCCATCCGCGCTGTGGGGTCTTTTTCTTTTTTGGGGGGTAGCGGACATCCGATAACCCGCGTCTATCGCCACCATGTCGGCTCGCGAACCCGC  2436

GGGGCGCAGGAGGCGCGCATCCACCCGCCCCCGCGCCTCGCCCGTGGCGGACGAGCCAGCGGGCGATGGGGTGGGGTTCATGGGGTACCTGCGTGCGGTGTTCCGCGGGGATGA  2550

CGACAGCGAGCTAGAGGCTCTGGAGGAGATGGCGGGCGACGAGCCGCCCGTGCGCCGTC                                                        2609
```

0.685 on the prototype HSV-1 genome, in a region bounded by SalI and XhoI sites (see figure 1 and reference 26). The length of this fragment of the genome is 2609 base pairs. Approximately 30,000 characters were present in the database and greater than 98.5% of the sequence was determined on both strands. The base composition is 66.0% G+C.

We have accurately located the 5' and 3' termini of the single, 1.7kb, mRNA which is known to map entirely within this region (27). The results are presented in figures 2 and 3. The 5' end of the mRNA was mapped using two independent probes end-labelled at SalI and MluI sites respectively. The MluI probe gave a hybrid band of about 67bp in size, consistent with the hybrid of approximately 200bp obtained using the SalI probe. These data position the mRNA start site at approximate position 274. The mRNA 3' terminus was mapped using a 3' end-labelled XcyI probe which gave an S1 hybrid band of about 205bp, positioning the end of the mRNA at approximately 2020. Note that the sequence presented is numbered right to left (5' to 3') with respect to the genomic orientation.

In the 5' flanking regions it is possible to discern various sequence motifs which are similar to sequences implicated in control of transcription. These are as follows:

(a) TTAAAT, at position 246, which is a reasonable candidate for the "TATA box" (45), a sequence important for efficient and accurate initiation of RNA polymerase II transcription. It is in approximately the correct position, about -30 with respect to the cap site.

(b) AGCCAATGT, at position 196, an excellent match to a sequence, GGCCAATCT, implicated in the transcriptional activity of globin promoters (46, 47) but also found in many others (48).

Figure 3: The sequence of the gene for Vmw65

We present here 2609 base pairs of DNA sequence including the gene for Vmw65. The initiation site of the transcript for this gene is shown as O-----. The terminus of the mRNA is marked as -----X. Canonical sequences implicated in the control of transcription in other systems, and referred to individually in the text, are underlined. The single letter amino acid code is used to indicate the extent of the open reading frame for Vmw65.

(c) GC rich motifs similar to those found in both the SV40 21bp repeats and the HSV-1 TK gene (49, 50, 51). Specifically, CCGCCC at position 167 and GGGCGG at position 226, are in approximately similar places relative to the "TATA box" as those elements found in the TK promoter.

On the 3' side of the coding region lies a typical polyadenylation signal sequence AATAAA (52), at position 1991, followed about 30bp downstream by a sequence TGCGCTTG. This is consistent with a consensus (YGTGTTYY, where Y=pyrimidine) found in this position in many genes and shown to be functionally important in the termination of HSV mRNA (53,54).

The first initiating ATG codon to be found after the mRNA start was at position 454 and conforms satisfactorily with the rules of Kozak (55). This codon initates an open reading frame of 490 amino acids whose codon usage was consistent with that of other, known, HSV-1 proteins as determined by the program of Staden and McLachlan (39).

The DNA sequence of Fig.3 also contains what we consider to be the termini of the genes flanking the gene for Vmw65. At position 66 is a polyadenylation signal, presumably that for two, 3' co-terminal, messages coding for a 42,000 molecular weight polypeptide located upstream of Vmw65 (27). Hall and co-workers also indicate that a 70,000 molecular weight polypeptide should be translated from a mRNA initiating soon after the end of the mRNA for Vmw65. We believe that this protein begins at position 2417 and there is a reasonable "TATA" box at 2193.

The Vmw65 Protein

From the above predicted open reading frame, Vmw65 is composed of 490 amino acid residues and has a molecular weight of 54,342. The amino acid composition is presented in Table 1. The protein is highly charged, with the acidic residues, aspartic and glutamic acid, accounting for 14.3% of the total amino acids. Those amino acid residues with strongly basic side chains, arginine and lysine, make up a further 9%. The predicted net acidic character is consistent with two-dimensional gel electrophoresis of HSV-1 proteins, which has shown Vmw65 to be amongst the most acidic proteins in infected cells (56, 57).

Table 1: Predicted amino acid composition of Vmw65

Residue	No.	%	Residue	No.	%	Residue	No.	%
Ala	58	11.8	Arg	40	8.2	Asn	12	2.4
Asp	41	8.4	Cys	6	1.2	Gln	10	2.0
Glu	29	5.9	Gly	32	6.5	His	13	2.7
Ile	9	1.8	Leu	65	13.3	Lys	4	0.8
Met	16	3.3	Phe	23	4.7	Pro	40	8.2
Ser	26	5.3	Thr	27	5.5	Trp	4	0.8
			Val	17	3.5			

Analysis of hydropathicity shows that there are no extensive regions of an extreme hydrophobic or hydrophilic nature (data not shown). However, a histogram (figure 4) of the numbers of basic or acidic residues every 30 amino acids reveals that there is a distinct polarity along the peptide backbone, with acidic residues prevalent in the carboxy-terminal 80 amino acids and

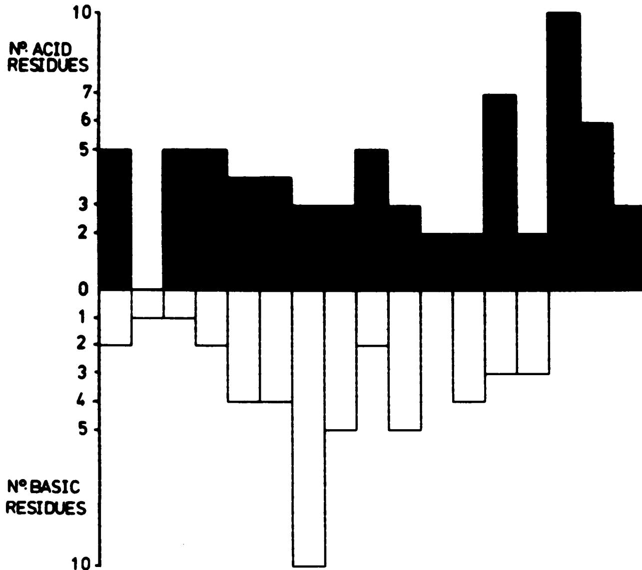

Figure 4: Histogram of acid/basic residue distribution
The distribution of acidic and basic residues along the polypeptide backbone is plotted. The interval is 30 amino acids.

HOMOLOGY MATRIX PLOT
X AXIS=HSV1BAMF.54K   from base no. 1 to base no. 490
Y AXIS=VZVCCC.45   from base no. 1 to base no. 410
Range=15 Scale= 0.98 Minimum value plotted= 28 Compressed 4 times

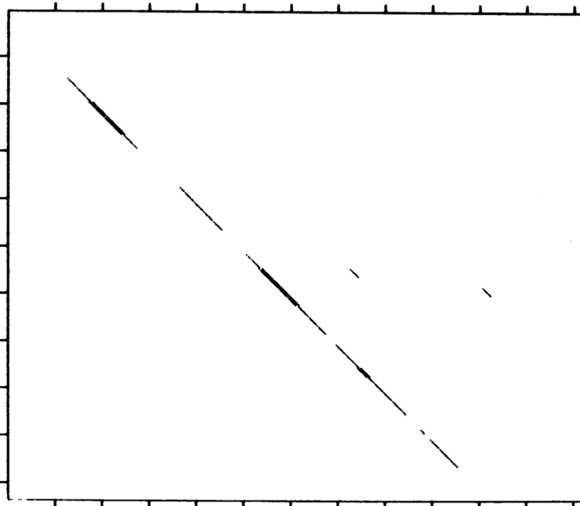

Figure 5: A dot matrix plot of homology between Vmw65 and VZV 45K

This plot was obtained by use of the program of Pustell and Kafatos (41) with the parameters indicated in the figure. Graphical output was programmed by Philip Taylor. It clearly shows colinearity between the two amino acid sequences except for the carboxy terminal 80 amino acid residues of Vmw65.

basic residues concentrated near the amino-terminal region. The relatively high leucine content may indicate that Vmw65 has a well defined secondary structure.

Homology to other proteins

We have detected a clear homologue to Vmw65 in the genome of varicella-zoster virus (VZV), another human herpesvirus (A.Davison, manuscript in preparation). The VZV protein has a molecular weight of 45,000 (45K). Figure 5 shows a dot matrix homology plot between the two proteins. It can be seen that they are approximately colinear except that the HSV-1 protein has 80 amino acid residues at the carboxy-terminus which the VZV 45K lacks. Figure 6 is an optimal alignment of the two amino acid sequences. This indicates that approximately 35% of amino acid residues are conserved between the two proteins. Since the VZV

```
  1 MD   LLVDELFADMNADGASPPPPRPA   GGPKNTPAAPPLYATGRLSQA QLMPSPPMPVPPAALFNRLLDDL
    *    *        *          *         *       *    * *   ****    *  *   *  **
  1 MECNLGTEHPSTDTWNRSKTEQAVVDAFDESLFGDVASDIGFETSLYSHAVKTAPSPPWVASPKILYQQLIRDL

 70 GFSAGPALCTMLDTWNEDLFSALPTNADLYRECKFLSTLPSDV    VEWGDAYVPERTQIDIRAHGDVAFPTLP
     ** ** *   * ********* * * ***     ** * **   *   *    *     *  * *  *  *
 75 DFSEGPRLLSCLETWNEDLFSCFPINEDLYSDMMVLSPDPDDVISTVSTKD  HVEMFNLTTR  GSVRLPSPP

141 ATRDGLGLYYEALSRFFHAELRAREESYRTVLANFCSALYRYLRASVRQLHRQAHMRGRDRDLGEMLRATIADR
       ** *     * ******* *  *  *  ***                 *   **  * *
145 KQPTGLPAYVQEVQDSFTVELRAREEAYTKLLVTYCKSIIRYLQGTAKRTTIGLNIQNPDQKAYTQLRQSILLR

215 YYRETARLARVLFLHLYLFLTREILWAAYAEQMMRPDLFDCLCCDLESWRQLAGLFQPFMFVNGALTVRGVPIE
    **** * *** * *****  ***  *  ** *  ** *  *      **   * *    *    * *
219 YYREVASLARLLYLHLYLTVTREFSWRLYASQSAHPDVFAALKFTWTERRQFTCAFHPVLCNHGIVLLEGKPLT

289 ARRLRELNHIREHLNLPLVRSAATEEPGAPLTTPPTLHGNQARASGYFMVLIRAKLDSYSSFTTSPSEAVMREH
    *  *** *  *  * *****    **   **   *     *  *    * *** *           *  *
293 ASALREINYRRRELGLPLVRCGLVEENKSPLVQQPSFSVHLPRSVGFLTHHIKRKLDAYAVKHPQEPRHVRADH

363 AYSRARTKNNYGSTIEGLLDLPDDDAPEEAGLAAPRLSFLPAGHTRRLSTAPPTDVSLGDELHLDGEDVAMAHA
     *       **** *           *  ** *  *    *   *   ***   *
367 PYAKVVENRNYGSSI             EAMILAPPSPSEILPGDPPR    PPTCGFLTR

437 DALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG

409
```

<u>Figure 6</u>: <u>An optimal alignment of the amino acid sequences of Vmw65 and VZV 45K</u>

Obtained using the optimal alignment program of Taylor (42) and using the default parameters recommended by the author. The top line is the amino acid sequence of Vmw65.

and HSV-1 genomes have diverged considerably, to the extent that their base compositions differ by 21 percentage points, we regard this homology as a good confirmation of the correctness of our Vmw65 reading frame assignment. We have been unable to detect a homologue to Vmw65 in another herpesvirus, Epstein Barr virus, whose genome has been completely sequenced (58).Similarly, a search of the National Biological Research Foundation protein database produced no matches of obvious significance.

DISCUSSION

The results of Campbell and co-workers (26), using transfection of plasmid clones, preclude the possibilities either that the tegument's structural integrity is involved in

IE gene stimulation or that other viral proteins play an essential part. Since no other open reading frames of significant length, with acceptable codon usage, were detected on either strand we conclude that the polypeptide described here is alone responsible for specific trans-activation of HSV-1 IE promoters. The difference between the molecular weight determined by SDS-polyacrylamide gel electrophoresis and that obtained by sequence analysis cannnot be due to phosphorylation since the in vitro translation product is of similar electrophoretic mobility (26). We believe that it is more likely that the large number of charged residues result in aberrant electrophoretic mobility. The first AUG codon may not be the sole initiating codon in vivo, since experiments involving hybrid-arrest translation, in vitro, followed by two-dimensional gel electrophoresis have detected a minor polypeptide which probably results from initiation of translation at the second AUG codon, 33bp further downstream (C.M.Preston and L.Haarr, unpublished observations). This smaller polypeptide is capable of trans-activation, since plasmids possessing an 8bp oligonucleotide insert at the SalI site between the two ATG codons are active in the co-transfection assay (26). Polypeptides resulting from translation initiation at downstream, in phase, AUG codons of the HSV thymidine kinase mRNA have been previously described (57,59).

The VZV 45K polypeptide is homologous to HSV Vmw65 but lacks the carboxy-terminal 80 amino acids, which are relatively rich in acidic residues. It will be interesting to determine whether the VZV homologue and Vmw65 can stimulate expression of the heterologous IE genes. The VZV gene homologous to HSV-1 IE3 does not exhibit sequences corresponding to the TAATGARATTC consensus in its 5'-flanking region (60). Therefore, either the VZV 45K protein recognises a different upstream sequence or it may lack the ability to stimulate transcription.

The function of the viral tegument is unknown. Vmw65 comprises a significant proportion of the total protein in the tegument and is likely to play an important structural role. The tegument may be instrumental in the maturation of virus particles by directing the envelopment of nucleocapsids. Two

possible mechanisms by which this could be brought about are either by direct interaction with the lipid component of the envelope or, indirectly, by complexing with virally encoded envelope proteins. There is no obvious sequence evidence, in the form of a large hydrophobic domain, for direct membrane interaction. It is known that Vmw65 is tightly complexed with glycoprotein B and possibly other viral glycoproteins which reside in the envelope (M.Campbell, personal communication). Another possible role for the protein may be in maintaining the structural integrity of the tegument. Our analysis of the distribution of acidic/basic residues indicates a potential polarity to the protein molecule. Such a polarity might allow Vmw65 to be arranged, in the tegument, in an ordered fashion and result in a structure held in place by hydrogen bonding interactions. It is thought that the tegument is especially sensitive to dehydration (61) which is compatible with such a mechanism. Other functions we can postulate for Vmw65 would be important in the early stages of infection, for example the tegument might provide a protective layer around the nucleocapsid or it may play a role in the virion's nuclear tropism.

In conclusion, we present here the DNA and implied protein sequences of a gene from HSV-1 whose product specifically stimulates transcription from HSV-1 IE promoters. The predicted protein is notably acidic. It shows homology to an open reading frame in the VZV genome, but no significant homology to any other known amino acid sequence. The sequence will allow detailed molecular genetic analyses to define functionally important regions of the polypeptide. Further immunological and biochemical experiments may provide insights into how this protein specifically stimulates HSV IE transcription and into the mechanisms of promoter recognition.

ACKNOWLEDGEMENTS

We thank Professor J. H. Subak-Sharpe for reading the manuscript and Dr. D. Bzik for his help and interest in the work. Thanks are also due to Dr. A. Coulson and Dr. P. Taylor for their help in setting up protein database searches. M.D. was a recipient of a Medical Research Council Research Training Award.

+ Members of the MRC Virology Unit.

LITERATURE CITED

1. Wildy, P., Russell, W.C. and Horne, R.W. (1960) Virology 12, 204-224.
2. Becker, Y., Dym, H. and Sarov, I. (1968) Virology 36, 184-192.
3. Graham, B.J., Ludwig, H., Bronson, D.L., Benyesh-Melnick, M. and Biswal, N. (1972) Biochem. Biophys. Acta 259, 13-23.
4. Honess, R.W. and Roizman, B. (1974) J. Virol. 14, 8-19.
5. Swanstrom, R.I., Piro, K. and Wagner, E.K. (1975) Virology 66, 140-150.
6. Clements, J.B., Watson, R.J. and Wilkie, N.M. (1977) Cell 12, 275-285.
7. Jones, P.C. and Roizman, B. (1979) J. Virol. 31, 299-314.
8. Preston, C.M. (1979) J. Virol. 29, 275-285.
9. Watson, R.J. and Clements, J.B. (1980) Nature 285, 329-330.
10. Everett, R.D. (1984) EMBO J. 3, 3135-3141.
11. O'Hare, P. and Hayward, G.S. (1985) J. Virol. 53, 751-760.
12. Quinlan, M.P. and Knipe, D.M. (1985) Mol. Cell. Biol. 5, 957-963.
13. Post, L.E., Mackem, S. and Roizman, B. (1981) Cell 24, 555-565.
14. Mackem, S. and Roizman, B. (1982) Proc. Natl. Acad. Sci. USA 79, 4917-4921.
15. Cordingley, M.G., Campbell, M.E.M. and Preston, C.M. (1983) Nucleic Acids Res. 11, 2347-2365.
16. Lang, J.C., Spandidos, D. and Wilkie, N.M. (1984) EMBO J. 3, 389-395.
17. Preston, C.M. and Tannahill, D. (1984) Virology 137,439-444.
18. Preston, C.M., Cordingley, M.G. and Stow, N.D. (1984) Virology 50, 708-716.
19. Puga, A., Gomez-Marquez, J., Brayton, P.R., Cantin, E.M., Long, L.K., Barbacid, M. and Notkins, A.L. (1985) J. Virol. 54, 879-881.
20. Mackem, S. and Roizman, B. (1982) J. Virol. 44, 939-949.
21. Kristie, T.M. and Roizman, B. (1984). Proc. Natl. Acad. Sci. USA 81, 4065-4069.
22. Murchie, M.J. and McGeoch, D.J. (1982) J. Gen. Virol. 62, 1-15.
23. Whitton, J.L., Rixon, F.J., Easton, A.E. and Clements, J.B. (1983) Nucleic Acids Res. 11, 6271-6287.
24. Whitton, J.L. and Clements, J.B. (1984) Nucleic Acids Res. 12, 2061-2079.
25. Batterson, W. and Roizman, B. (1983) J. Virol. 46, 371-377.
26. Campbell, M.E.M., Palfreyman, J.W. and Preston, C.M. (1984) J. Mol. Biol. 180, 1-19.
27. Hall, L.M., Draper, K.G., Frink, R.J., Costa, R.H. and Wagner, E.K. (1982) J. Virol. 43, 594-607.
28. Marsden, H.S., Stow, N.D., Preston, V.G., Timbury, M.C. and Wilkie, N.M. (1978) J. Virol. 28, 624-642.
29. Lemaster, S. and Roizman, B. (1979) J. Virol. 35, 798-811.
30. Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216-218.
31. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.

32. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory CSH.
33. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
34. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
35. Deininger, P.L. (1983) Anal. Biochem. 129, 216-223.
36. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
37. Staden, R. (1982) Nucleic Acids Res. 10, 4731-4751.
38. Blumenthal, R.M., Rice, P.J. and Roberts, R.J. (1982) Nucleic Acids Res. 10, 91-101.
39. Staden, R. and McLachlan, A.D. (1982) Nucleic Acids Res. 10, 141-156.
40. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
41. Pustell, J. and Kafatos, F.C. (1982) Nucleic Acids Res. 10, 4765-4782.
42. Taylor, P. (1984) Nucleic Acids Res. 12, 447-456.
43. Preston, C.M. (1977) Virology 78, 349-353.
44. Rixon, F.J. and Clements, J.B. (1982) Nucleic Acids Res. 10, 2241-2256.
45. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science (Wash.) 209, 1409-1414.
46. Dierks, P., Van Ooyen, A., Mantel, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1411-1415.
47. Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. (1982) Nature 295, 120-126.
48. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142.
49. Everett, R.D., Baty, D. and Chambon, P. (1983) Nucleic Acids Res. 11, 2447-2464.
50. McKnight, S.L., Gavis, E.R., Kingsbury, R. and Axel, R. (1981) Cell 25, 385-398.
51. McKnight, S.L., Kingsbury, K.C., Spence, A. and Smith, M. (1984) Cell 37, 253-262.
52. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
53. Taya, Y., Devos, R., Tavernier, J., Cheroutre, H., Engler, G. and Fiers, W. (1982) EMBO J. 1, 953-958.
54. McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985) Nucleic Acids Res. 13, 1347-1368.
55. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
56. Haarr, L. and Marsden, H.S. (1981) J. Gen. Virol. 52, 77-92.
57. Marsden, H.S., Haarr, L. and Preston, C.M. (1983) J. Virol. 46, 434-445.
58. Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., Tuffnell, P.S. and Barrell, B.G. (1984) Nature 310, 207-211.
59. Preston, C.M. and McGeoch, D.J. (1981) J. Virol. 38, 593-605.
60. Davison, A.J. and Scott, J.E. (1985) J. Gen. Virol. 66, 207-220.
61. Vernon, S.K., Lawrence, W.C., Long, C.A., Rubin, B.A. and Sheffield, J.B. (1982) J. Ultrastruct. Res. 81, 163-171.