Volume 12 Number 5 1984 Nucleic Acids Research

A 3' co-terminal family of mRNAs from the herpes simplex virus type 1 short region: two overlapping reading frames encode unrelated polypeptides one of which has a highly reiterated amino acid sequence

F.J.Rixon and D.J.McGeoch

Medical Research Council Virology Unit, Department of Virology, Glasgow University, Glasgow G11 5JR, UK

Received 3 January 1984; Accepted 8 February 1984

ABSTRACT

We have used DNA sequencing, mRNA mapping and in vitro translation to characterise three partially overlapping genes in the genome of herpes simplex virus (HSV) type 1. These genes specify three mRNAs with distinct 5' termini but a common 3' terminus, the longest of which is immediate-early (IE) mRNA-5. The 12,000 MW (12K) IE polypeptide encoded by IEmRNA-5 is translated from an 88 codon open reading frame, leaving a 1200 base 3' non-translated region. The second mRNA (mRNA-B) is initiated within the coding sequence of IEmRNA-5, and encodes a 21K polypeptide. The 12K and 21K polypeptide coding regions do not overlap. The third mRNA (mRNA-C) is initiated within the coding region of mRNA-B, and encodes a 33K polypeptide. The reading frame for 33K has a 110 codon out-of-frame overlap with the 21K reading frame. This is the first instance of overlapping genes described for HSV. The 21K polypeptide is thought to be a DNA binding protein and is remarkable for an array of 24 tandem repeats of the sequence X/Pro/Arg (where X represents predominantly Glu, Asp, Thr, Ser or Val) in its C-terminal portion. This array, which occupies most of the region of overlap with 33K, can vary in repeat number between virus strains.

INTRODUCTION

Transcription in herpes simplex virus is temporally regulated and has been divided into three classes designated immediate-early, early and late (1-3). The majority of well characterised transcripts belonging to the early and late temporal classes are members of 3' co-terminal families comprising two or more mRNAs (4-7). Unlike the situation in adenoviruses and papovaviruses, few HSV mRNAs are spliced, so these overlapping mRNAs have unique 5' termini rather than common 5' leader sequences. No member of the immediate-early mRNA class (which is made in the absence of viral polypeptide synthesis) has previously been identified as belonging to a 3'

co-terminal family (8-11).

Five major IE mRNAs (designated IEmRNAs 1-5) have been described and mapped on the HSV genome (12,13)(Fig 1). IEmRNAs-4 and 5 (which do have spliced forms) are transcribed across the IR_S/U_S and TR_S/U_S junctions respectively (8,9,17)(Fig 1). IEmRNA-4 encodes a polypeptide with a molecular weight from polyacrylamide gel electrophoresis (MWapp) of 68K, while IEmRNA-5, which is of similar size, encodes an unrelated polypeptide (Vmw 12) of MWapp 12K (12,18). Only a fraction of the potential coding capacity of IEmRNA-5 would be required to encode Vmw 12, suggesting that this genome region might encode additional polypeptides. Two other polypeptides of MWapp 21K (Vmw 21) and 33K (Vmw 33) have been mapped to this approximate location (19,20). In order to clarify the situation we undertook a detailed investigation of the mRNAs mapping in this region.

MATERIALS AND METHODS

Preparation of cytoplasmic mRNA

Baby hamster kidney 21 (C13) cells were grown as monolayers in rotating 80 oz. bottles. All infections were carried out using HSV-1 (Glasgow strain 17) at 37°C. For the production of IE RNA, cell monolayers were infected at a multiplicity of infection of 50 p.f.u./cell. The cell monolayers were pre-treated and maintained in medium containing cycloheximide as previously described (2). For the production of 3 h and 6 h RNA, cell monolayers were infected at a multiplicity of infection of 10 p.f.u./cell and incubated for the appropriate time. Cytoplasmic RNA was prepared using the method of Kumar and Lindberg (21).

Cloning Procedures

Fragments of HSV-1 and HSV-2 DNA, generated by restriction endonuclease digestion, were cloned under Category I containment (U.K. Genetic Manipulation Advisory Group). The host bacterium was Escherichia coli K12 HB101 and the cloning vector pAT153 (16). Isolation of cloned DNA was as described previously (22). Clones ZF5, ZF15 and KR630, used in hybrid arrest translation, were HSV-1 DNA fragments cloned into M13mp8

for DNA sequencing (23).

Structural analysis of mRNAs

The DNA/RNA hybridisation procedures and the nuclease S1 and exonuclease VII digestion procedures were carried out as described previously (9).

Denaturing polyacrylamide gels containing 9M urea were run in 90 mM Tris, 90 mM boric acid, pH 8.3, 1 mM EDTA. Samples were dissolved in deionised formamide and denatured at 90°C for 2 min before loading. Electrophoresis was carried out at room temperature for 3-6h at 40W. The radio-labelled bands were detected by autoradiography.

DNA sequencing

The sequence of BamHI x was determined previously (24). The sequence of residues 1345 to 2400 (Fig 4) was determined by cloning sonicated fragments of HSV-1 KpnI h DNA into M13mp8 (23) and sequencing the inserts by the chain terminator method (25).

Hybrid arrest

Hybrid arrested translation was performed essentially as described by Preston and McGeoch (26). 10 ug of plasmid DNA (either pAT 153 or M13 was used as a cloning vector) was denatured in 85 ul of 95% formamide containing 60 mM PIPES (pH 7.4) at 95°C for 5 min. 6 ul of cytoplasmic RNA from HSV-1 infected cells and 8 ul of 5 M NaCl were added and the mixture was incubated for 1 h at 58°C. Incubation was stopped by addition of 1 ml of ice cold HSB (500 mM NaCl, 20 mM Tris-HCl [pH 7.5], 1 mM EDTA). Poly(A)+ RNA was selected on an oligo (dT) cellulose column and ethanol precipitated after addition of *E. coli* rRNA. The RNA was pelleted, dissolved in deionized water and used for *in vitro* translation.

In vitro translation

RNA samples were translated in a micrococcal nuclease-treated fractionated reticulocyte system using the procedure described by Preston (27).

RESULTS

Analysis of transcription at IEmRNA-5 locus.

Northern blotting analysis using the plasmids pGX55 and pGX57 as probes (Fig 1), indicated the presence of 2 smaller

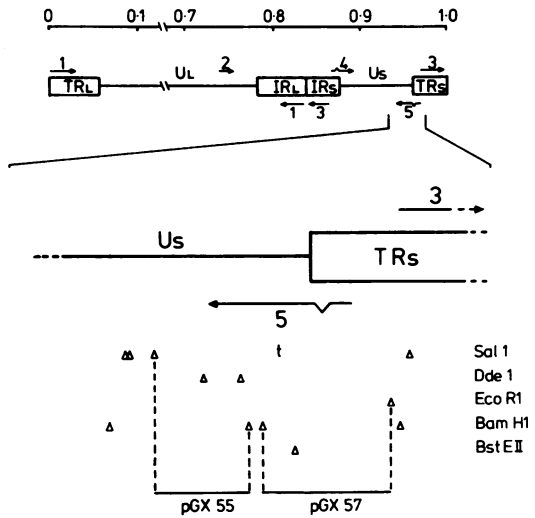

FIG. 1 Diagram of the HSV-1 genome arranged in the prototype orientation (14), showing the map locations and orientations of the five IEmRNAs. Fractional genomic length is given at the top. The long and short regions of the genome consist of unique regions (U_L and U_S) flanked and joined by inverted, repeated regions (TR_L/IR_L and IR_S/TR_S) (15). The 5' portions of IEmRNAs-4 and 5, which map within IRs and TRs, are of identical sequence while the 3' portions extending into opposite sides of Us are unrelated (8,9). Both unspliced and spliced forms of these 2 mRNAs occur (9), the latter having a single common intron mapping in IRs and TRs. The expansion shows the restriction enzyme sites and cloned DNA fragments (SalI t, pGX55 and pGX57) which were used for Northern blot analysis and nuclease digestion analysis of the mRNAs mapping in the vicinity of IEmRNA-5, as detailed in the text. The DNA fragments were cloned into pAT 153 (16).

mRNAs in addition to IEmRNA-5 at late times post-infection (PI)(data not shown).

Nuclease analysis was performed on mRNA made under IE conditions and at 3 h and 6 h PI. pGX57 5' labelled at the BamHI site (Fig 1) generated 4 nuclease S1-resistant bands (Fig 2A, lanes 1,2, & 3). The 1,227 and 770 base bands representing the previously described unspliced and spliced forms of IEmRNA-5 (9) were formed with all 3 mRNA preparations. However, 2 additional bands, with lengths of 581 and 287 bases were present only in the 3 h and 6 h samples. Both were present in equal amounts in nuclease S1 (Fig 2A, lanes 2 & 3) and exonuclease VII

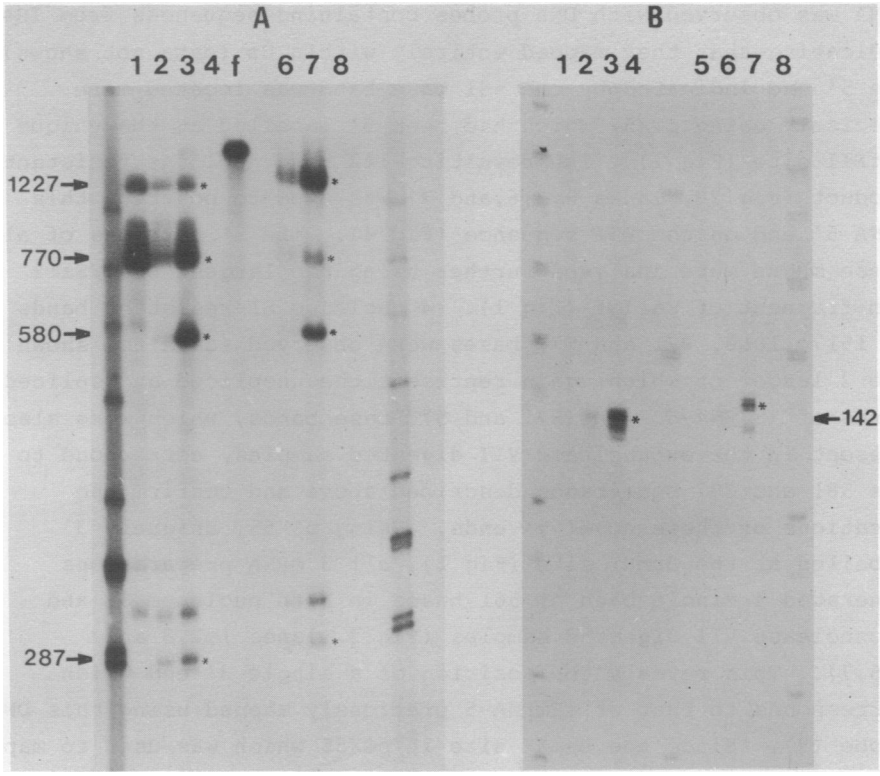

FIG. 2 Nuclease S1 and exonuclease VII analysis of the IEmRNA-5 mRNA family. Probes complementary to the 5' portions of IEmRNA-5 were hybridised to the following RNA samples: 1 and 5. 15 ug of cytoplasmic RNA made under IE conditions. 2 and 6. 15 ug of cytoplasmic RNA prepared at 3 h PI. 3 and 7. 15 ug of cytoplasmic RNA prepared at 6 h PI. 4 and 8. 20 ug of mock-infected cytoplasmic RNA. Track f in panel A is unhybridised DNA probe. Samples 1-4 were digested with nuclease S1 and samples 5-8 were digested with exonuclease VII. The DNA probes used were pGX57 5' labelled at the BamHI site (panel A) and pGX57 5' labelled at the BstEII site (panel B)(Fig. 1). All samples were separated on a 6% denaturing polyacrylamide gel. The sizes of the protected fragments in bases are given alongside the bands. The additional 310 base band in panel A coincides with an AT rich sequence and is probably due to localised melting of the DNA/RNA hybrids at this point. 5' labelled HincII digested øX 174 DNA and 3' labelled HpaII digested pBR 322 DNA were used as size standards.

(Fig 2A, lanes 6 & 7) digested material, suggesting that they represent the 5' portions of unspliced transcripts. This conclusion is supported by Northern blotting data, since neither

mRNA was observed with DNA probes containing sequences from IRs, indicating that they mapped entirely within Us (data not shown). The 5' end indicated by the 581 base band was located more precisely using pGX57 which had been 5' labelled at the unique BstEII site (Fig 1). The resulting 142 base nuclease resistant product (Fig 2B, lanes 2,3,6,and 7) was used to position this mRNA 5' end on the DNA sequence (Fig 4). The 5' portions of all these mRNAs were analysed further using the largest DdeI/SalI sub-fragment of SalI t (Fig 1). 4 nuclease S1-resistant bands of 1517, 1060, 871 and 577 bases were observed (data not shown), the 2 larger of which again represent the unspliced and spliced forms of IEmRNA-5. The 871 and 577 base bands, which were also present in the exonuclease VII digested samples, correspond to the 581 and 287 base bands described above and confirm the locations of these novel 5' ends. Using pGX55, uniquely 3' labelled at the BamHI site (Fig 1), all 3 mRNA preparations generated a single band of 561 bases in both nuclease S1 and exonuclease VII digested samples (Fig 3, lanes 1,2,3 and 5,6,7). This reveals the position of a single 3' end which corresponds to that of IEmRNA-5 previously mapped using this DNA probe (9). Since the BamHI site in pGX55 which was used to map the 3' end lies upstream from the DdeI site in SalI t which was used to map the 5' ends, we conclude from these experiments that IEmRNA-5 is a component of a 3' co-terminal family with 3 members; namely, the unspliced (1990 base) and spliced (1780 base) forms of IEmRNA-5, a 1341 base mRNA (henceforth designated mRNA-B) and a 1047 base mRNA (designated mRNA-C).

The DNA sequence at the IEmRNA-5 locus.

The sequence of the DNA region encoding the mRNAs identified above, was determined (Fig 4). The salient features of this sequence are detailed below.

(a) The 5' end of IEmRNA-5 at position 136 has been mapped previously and the appropriate transcriptional regulation signals and splice points identified (8,9). The 264 base open reading frame commencing at position 667 would specify a 9.8K polypeptide and was thought to encode Vmw 12. Generation of a monospecific antibody against Vmw 12, using a synthetic oligopeptide predicted from this sequence, confirmed that this

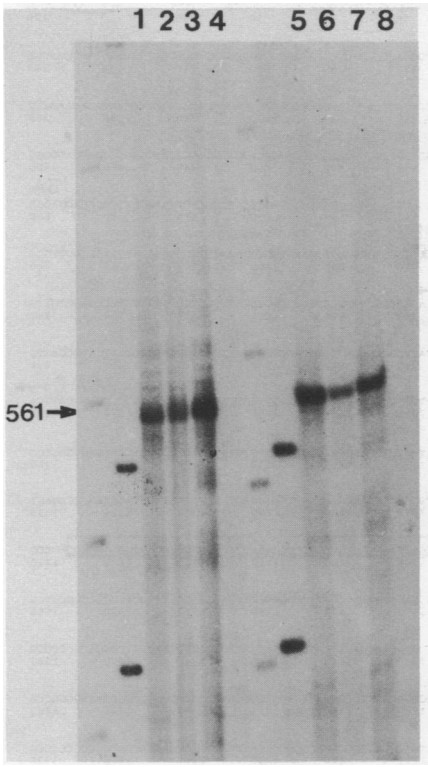

FIG. 3 Analysis of the IEmRNA-5 3' co-terminus. pGX55 3' labelled at the BamHI site was hybridised to the following RNA samples: 1 and 5. 15 ug of cytoplasmic RNA made under IE conditions. 2 and 6. 15 ug of cytoplasmic RNA prepared at 3 h PI. 3 and 7. 15 ug of cytoplasmic RNA prepared at 6 h PI. 4 and 8. 20 ug of mock-infected cytoplasmic RNA. Samples 1-4 were digested with nuclease S1 and samples 5-8 were digested with exonuclease VII. All samples were separated on a 6% denaturing polyacrylamide gel. 5' labelled HincII digested øX 174 DNA and 3' labelled HinfI digested pAT 153 DNA were used as size standards.

was indeed the case (J. Palfreyman, man. submitted).

(b) The 5' end of mRNA-B at position 784 lies 23 bp downstream from a sequence TAAAA which resembles a 'TATA" box (consensus TATAA (29)). The first ATG (position 998) in mRNA-B initiates a 483 base open reading frame (frame B) which would specify a 17.8K polypeptide.

(c) The 5' end of mRNA-C at position 1078 lies 20 bp downstream from a potential 'TATA' box (sequence TAAAA). The first ATG (position 1118) begins a 363 base open reading frame which constitutes the 3' portion of the reading frame described for mRNA-B. However, sequences flanking this ATG place it in a class which is invariably inactive as an initiating codon (31). The second ATG (position 1149), which does belong to an active class, initiates a 936 base open reading frame (frame C) which would specify a 34K polypeptide. The 5' portion of frame C and the 3' portion of frame B overlap for 332 bp and the

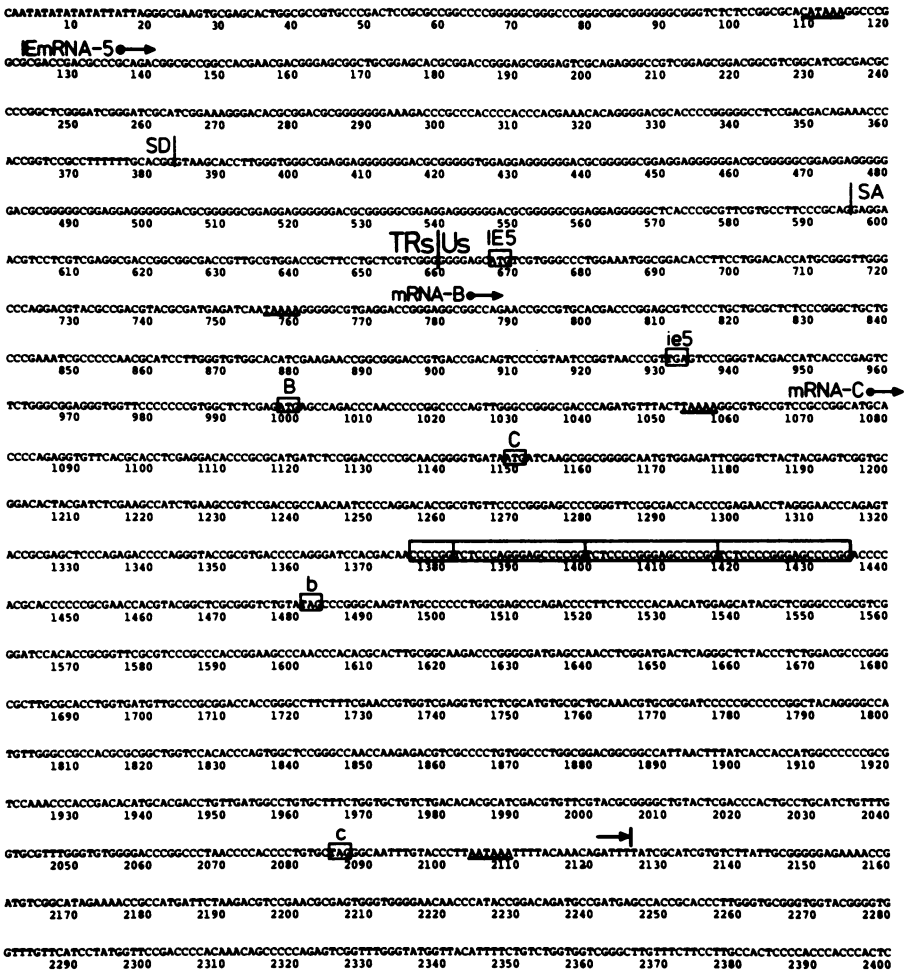

FIG. 4 Nucleotide sequence of the genome region specifying IEmRNA-5. The DNA sequence of 2400 residues of the mRNA-sense strand is shown, starting upstream from IEmRNA-5. Residues 1-1368 are from Murchie and McGeoch (24). Residues 1345-2400 are from our sequence for the HSV-1 strain 17 KpnI h fragment (McGeoch et al. in preparation). A version of the sequence extending to residue 2188 for HSV-1 Patton, which lacks one copy of the 18 bp repeat described below, has been published by Watson and Vande Woude (28). The junction between Us and TRs is indicated. Arrows above the sequence define the unique 5' ends of the 3 mRNAs (●—▶) and the 3' co-terminus (—▶|). The 5' ends are designated IEmRNA-5, mRNA-B and mRNA-C as appropriate. The 'TATA' (29) and AATAAA (30) transcriptional regulation signals are underlined. The splice donor (SD) and splice acceptor (SA) sites are shown for the spliced form of IEmRNA-5. The ATG initiation codons and the termination codons for each of

the predicted open reading frames are enclosed in boxes and they are designated IE5, B and C (initiation codons) and ie5, b and c (termination codons), to correspond with their parental mRNAs. Reading frames B and C overlap between positions 1149 and 1480. The open boxes enclose an 18 bp DNA sequence which is present as $3\frac{1}{3}$ direct repeats.

polypeptides they encode would be specified by different reading frames across this region.

Polypeptides encoded by the IEmRNA-5 3' co-terminal mRNAs.

Examination of the DNA sequence suggests a model for this 3' co-terminal family in which the 3 overlapping mRNAs encode unrelated polypeptides of 9.8K (MWapp 12K), 17.8K and 34K. As described above polypeptides of MWapp 21K and 33K have been mapped to this region. To determine whether these are equivalent to the predicted 17.8K and 34K polypeptides, cloned DNA fragments were used to arrest *in vitro* translation of mRNA prepared 6 h PI (Fig 5). Hybridisation of 6 h mRNA to pGX57, which specifies the 5' portions of all 3 mRNAs, arrested translation of 33K and 21K polypeptides (presumed to be Vmw 33 and Vmw 21). The other 3 DNA probes were cloned in a single-stranded M13 vector which allowed simultaneous determination of coding limits and orientations of the coding strands (Fig 5). ZF15 did not arrest either Vmw 33 or Vmw 21, ZF5 arrested Vmw 33 but not Vmw 21, and KR630 arrested both Vmw 33 and Vmw 21. These overlapping DNA fragments delineate the possible sequences encoding these polypeptides. Thus the C terminus of Vmw 33 lies between positions 1650 and 2184, while the C terminus of Vmw 21 lies between 1345 and 1650. In addition the 5' portions of the mRNAs specifying these polypeptides must extend to within pGX57. Within these limits the only suitable open reading frames are those identified above. Therefore we conclude that the 3 mRNAs comprising this 3' co-terminal family encode polypeptides with MWapp of 12K (IEmRNA-5), 21K (mRNA-B) and 33K (mRNA-C).

DISCUSSION

mRNA families with common 3' and unique 5' termini are known from several regions of the HSV-1 genome (4-7). However this particular example has a number of interesting features.

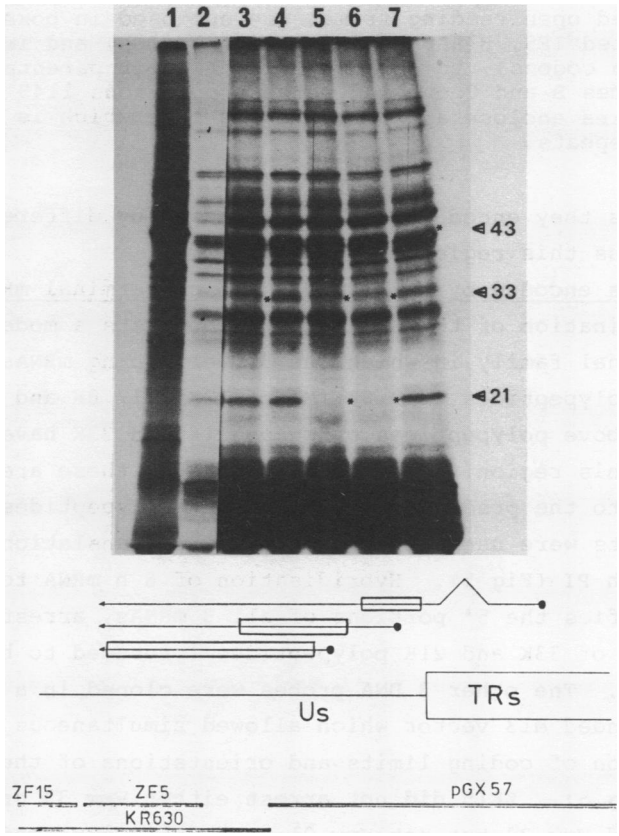

FIG. 5 Hybrid arrested translation of HSV-1 mRNA. Hybridisation and in vitro translation were performed as described in methods. The locations and structures of the mRNAs forming the IEmRNA-5 3' co-terminal family are shown. The open boxes indicate the predicted open reading frames for each mRNA. The DNA hybridisation probes used are also shown. pGX57 was cloned into pAT 153; ZF15, ZF5 and KR630 were cloned into a single-stranded M13 vector. Using the base numbering system from Fig 4 the probes mapped between -587 (ie. 587 bp upstream of the first base shown in Fig 4) and 1362 (pGX57), 1345 and 3940 (KR630), 1650 and 2184 (ZF5) and 2184 and 2322 (ZF15). ZF15, KR630 and pGX57 extend beyond the limits of this figure. Lane 1 shows the translation profile of mock-infected mRNA. Lanes 2-7 show the translation profiles of mRNA prepared from cells 6 h PI with 10 pfu/cell of HSV-1 and arrested with the following DNA probes: 2. control track, no DNA added. 3. ZF5. 4. ZF15. 5. KR630. 6. pGX57. 7. HSV-1 BamHI p (included as a control which was known to arrest the 43K polypeptide (26)). Asterisks to the right of each track indicate the expected position of the arrested bands. The samples were analysed on a 12.5% polyacrylamide gel.

(1) It is the only one known to include an IEmRNA.
(2) The 5' end and 'TATA' box for mRNA-B map within the DNA sequences encoding Vmw 12, and the 5' end and 'TATA' box for mRNA-C map within the DNA sequences encoding Vmw 21. A similar situation has been described previously for HSV-1 (4).
(3) The reading frames encoding Vmw 21 and Vmw 33 overlap between positions 1149 and 1480 in which region the two polypeptides are specified by different reading frames (Fig 4). This is the first demonstration of overlapping, functional reading frames for herpesviruses although examples are known for other viruses (32).
(4) The DNA sequence encoding the overlapping portions of Vmw 21 and Vmw 33 contains $3\frac{1}{3}$ copies of a tandemly reiterated 18 bp sequence, which must specify reiterated amino acid sequences. Thus Vmw 33 will contain 3 tandem repeats of the amino acid sequence LPGSPG while Vmw 21 will contain 3 tandem repeats of the hexapeptide SPREPR. Short, tandemly reiterated DNA sequences are a feature of the HSV-1 genome (22,24); however, this is the first example where the reiteration is thought to be coding. The number of copies of this DNA sequence varies between different strains of HSV-1 (this paper and Ref. 33) which should result in size variation of Vmw 21 and Vmw 33. Variation of Vmw 33 has not been reported. Vmw 21 however, is known to exhibit inter-strain variability, which may be attributable to this phenomenon (34).
(5) The C terminal portion of Vmw 21 contains 24 tandem repeats of the tripeptide:- X/Pro/Arg (Fig 6). This repeated amino acid sequence occupies most of the region of overlap between the coding sequences of Vmw 33 and Vmw 21. Each 18 bp DNA reiteration encodes 2 of these tripeptides but the repeated amino acid sequence extends well beyond the reiterated DNA sequence. There is no equivalent tandemly reiterated component in Vmw 33 apart from that specified by the 18 bp reiterated nucleotide sequence, and the remainder of the nucleotide sequence has no reiterated component. That this non-reiterated nucleotide sequence can encode a reiterated amino acid sequence is possible because all the potential codons for proline and arginine are used. A similar type of repeating amino acid

(A)

Met Ser Gln Thr Gln Pro Pro Ala Pro Val Gly Pro Gly Asp Pro Asp Val Tyr Leu Lys

Gly Val Pro Ser Ala Gly Met His Pro Arg Gly Val His Ala Pro Arg Gly His Pro Arg

Met Ile Ser Gly Pro Pro Gln Arg Gly Asp Asn Asp Gln Ala Ala Gly Gln Cys Gly Asp

Ser Gly Leu Leu Arg Val Gly Ala Asp Thr Thr Ile Ser Lys Pro Ser Glu Ala Val Arg

Pro Pro Thr Ile Pro Arg▲Thr Pro Arg▲Val Pro Arg▲Glu Pro Arg▲Val Pro Arg▲Pro Pro

Arg▲Glu Pro Arg▲Glu Pro Arg▲Val Pro Arg▲Ala Pro Arg▲Asp Pro Arg▲Val Pro Arg▲Asp

Pro Arg▲Asp Pro Arg▲Gln Pro Arg Ser Pro Arg▲Glu Pro Arg Ser Pro Arg▲Glu Pro Arg

Ser Pro Arg▲Glu Pro Arg Thr Pro Arg▲Thr Pro Arg▲Glu Pro Arg▲Thr Ala Arg Gly Ser

Val

(B)

Met Ile Lys Arg Arg Gly Asn Val Glu Ile Arg Val Tyr Tyr Glu Ser Val Arg Thr Leu

Arg Ser Arg Ser His Leu Lys Pro Ser Asp Arg Gln Gln Ser Pro Gly His Arg Val Phe

Pro Gly Ser Pro Gly Phe Arg Asp His Pro Glu Asn Leu Gly Asn Pro Glu Tyr Arg Glu

Leu Pro Glu Thr Pro Gly Tyr Arg Val Thr Pro Gly Ile His Asp Asn Pro Gly Leu Pro

Gly Ser Pro Gly Leu Pro Gly Ser Pro Gly Leu Pro Gly Ser Pro Gly Pro His Ala Pro

Pro Ala Asn His Val Arg Leu Ala Gly Leu Tyr Ser Pro Gly Lys Tyr Ala Pro Leu Ala

Ser Pro Asp Pro Phe Ser Pro Gln His Gly Ala Tyr Ala Arg Ala Arg Val Gly Ile His

Thr Ala Val Arg Val Pro Pro Thr Gly Ser Pro Thr His Thr His Leu Arg Gln Asp Pro

Gly Asp Glu Pro Thr Ser Asp Asp Ser Gly Leu Tyr Pro Leu Asp Ala Arg Ala Leu Ala

His Leu Val Met Leu Pro Ala Asp His Arg Ala Phe Phe Arg Thr Val Val Glu Val Ser

Arg Met Cys Ala Ala Asn Val Arg Asp Pro Pro Pro Pro Ala Thr Gly Ala Met Leu Gly

Arg His Ala Arg Leu Val His Thr Gln Trp Leu Arg Ala Asn Gln Glu Thr Ser Pro Leu

Trp Pro Trp Arg Thr Ala Ala Ile Asn Phe Ile Thr Thr Met Ala Pro Arg Val Gln Thr

His Arg His Met His Asp Leu Leu Met Ala Cys Ala Phe Trp Cys Cys Leu Thr His Ala

Ser Thr Cys Ser Tyr Ala Gly Leu Tyr Ser Thr His Cys Leu His Leu Phe Gly Ala Phe

Gly Cys Gly Asp Pro Ala Leu Thr Pro Pro Leu Cys

FIG. 6 The predicted amino acid sequences of the 21K (panel A) and 33K (panel B) polypeptides. In each case the underlined portion of sequence indicates the region of overlap between the reading frames encoding these two polypeptides. The boxes enclosing the repeating 6 amino acid sequences correspond to the boxes enclosing the 18 bp DNA repeats shown in Fig 4. The additional repeating 3 amino acid sequence (X/Pro/Arg), present in the 21K polypeptide, is indicated by the triangles.

sequence is present in collagen where the repeating triplet Gly/X/X (where X represents any amino acid) is specified by a non-repeating nucleotide sequence (35). Some other examples of highly repetitive amino acid sequences have been reported but these are generally encoded by reiterated nucleotide sequences (36,37).

It is interesting to speculate how such a system might have originated in HSV-1. The regularity of the highly reiterated amino acid component of the 21K polypeptide suggests that it is the amino acid sequence which is under selective pressure rather than the DNA sequence. Given this assumption there are two ways this system might have evolved. (a) The DNA sequence throughout the region was reiterated and has evolved into the present largely non-reiterated form, either through random mutation or as a result of selection acting upon the 33K polypeptide. (b) A pre-existing reiterated amino acid sequence has predisposed the DNA to the formation of tandem duplications which can arise without perturbing this amino acid sequence.

Since the repeating amino acid sequence in Vmw 21 is not simply a product of repetitive DNA its high degree of regularity suggests that this is directly related to its function. A 21K polypeptide (presumed to be Vmw 21) has been identified as a DNA binding protein with a high affinity for double-stranded DNA (38). We speculate therefore that Vmw 21 interacts with DNA through this basic repeating amino acid component. Recent work (R. Dalziel and H. Marsden, man. submitted) has shown that a 21K protein has a preferential affinity for the 'a' sequence which is present at each terminus of HSV genomic DNA and at the 'joint' between the long and short genome regions (39). It is not yet known if this corresponds to the 21K described here but if it does it may indicate a role for this polypeptide in circularisation of HSV-1 DNA, or in excision of unit length DNA from replicative concatemers.

No properties are known at present for Vmw 33, which is a minor component of infected cells and has only been reliably identified by *in vitro* translation.

ACKNOWLEDGEMENTS

We thank S. Donald and A. Dolan for their assistance.

Dr. C. Preston provided materials for *in vitro* translation and gave advice on this technique. SalI *t* was provided by Dr M.-J. Davison. We also thank Professor J. H. Subak-Sharpe for helpful discussion and comment on this manuscript.

REFERENCES

1. Swanstrom, R.I. & Wagner, E.K. (1974) Virology 60, 522-533.
2. Clements, J.B., Watson, R.J. & Wilkie, N.M. (1977) Cell 12, 275-285.
3. Jones, P.C. & Roizman, B. (1979) J. Virol. 31, 299-314.
4. McLauchlan, J. & Clements, J.B. (1983) J. Gen. Virol. 64, 997-1006.
5. Hall, L.M., Draper, K.G., Frink, R.J., Costa, R.H. & Wagner, E.K. (1982) J. Virol. 43, 594-607.
6. Watson, R.J., Colberg-Poley, A.M., Marcus-Sekura, C.J., Carter, B.J. & Enquist, L.W. (1983) Nucleic Acids Res. 11, 1507-1522.
7. Costa, R.H., Draper, K.G., Banks, L., Powell, K.L., Cohen, G., Eisenberg, R. & Wagner, E.K. (1983) J. Virol. 48, 591-603.
8. Watson, R.J., Sullivan, M. & Vande Woude, G.F. (1981) J. Virol. 37, 431-444.
9. Rixon, F.J. & Clements, J.B. (1982) Nucleic Acids Res. 10, 2241-2255.
10. Whitton, J.L., Rixon, F.J., Easton, A.J. & Clements, J.B. (1983) Nucleic Acids Res. 11, 6271-6287.
11. Rixon, F.J., Campbell, M.E. & Clements, J.B. (1982) The EMBO Journal 1, 1273-1277.
12. Watson, R.J., Preston, C.M. & Clements, J.B. (1979) J. Virol. 31, 42-52.
13. Anderson, K.P., Costa, R.H., Holland, L.E. & Wagner, E.K. (1980) J. Virol. 34, 9-27.
14. Morse, L.S., Buchman, T.G., Roizman, B. & Schaffer, P.A. (1977) J. Virol. 24, 231-248.
15. Sheldrick, P. & Berthelot, N. (1974) Cold Spring Harbor Symp. Quant. Biol. U.S.A. 39, 667-678.
16. Twigg, A.J. & Sherratt, D. (1980) Nature 283, 216-218.
17. Clements, J.B., McLauchlan, J. & McGeoch, D.J. (1979) Nucleic Acids Res. 7, 77-91.
18. Marsden, H.S., Lang, J., Davison, A.J., Hope, R.G. & MacDonald, D.M. (1982) J. Gen. Virol. 62, 17-27.
19. Marsden, H.S., Stow, N.D., Preston, V.G., Timbury, M.C. & Wilkie, N.M. (1978) J. Virol. 28, 624-642.
20. Lee, G.T.Y., Para, M.F. & Spear, P.G. (1982) J. Virol. 43, 41-49.
21. Kumar, A. & Lindberg, U. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 681-685.
22. Davison, A.J. & Wilkie, N.M. (1981) J. Gen. Virol. 55, 315-331.
23. Messing, J. & Vieira, J. (1983) Gene 19, 269-276.
24. Murchie, M.-J. & McGeoch, D.J. (1982) J. Gen. Virol. 62, 1-15.
25. Sanger,F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B. (1980) J. Molec. Biol. 143, 161-178.
26. Preston, C.M. & McGeoch, D,J. (1981) J. Virol. 38, 593-605.

27. Preston, C.M. (197J. Virol. 29, 275-284.
28. Watson, R,J. & Vande Woude, G.F. (1982) Nucleic Acids Res. 10, 979-991.
29. Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. & Chambon, P. (1979) Nature 278, 428-434.
30. Fitzgerald, M. & Shenk, T. (1981) Cell 24, 251-260.
31. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
32. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E. & Griffin, B.E. (1980) Nature 283, 445-453.
33. Watson, R.J., Umene, K. & Enquist, L.W. (1981) Nucleic Acids Res. 9, 4189-4199.
34. Lonsdale, D.M., Brown, S.M., Subak-Sharpe, J.H., Warren, K.G. & Koprowski, H. (1979) J. Gen. Virol. 43, 151-171.
35. Wozney, J., Hanahan, D., Tate, V., Boedtker, H. & Doty, P. (1981) Nature 294, 129-135.
36. Garfinkle, M.D., Pruitt, R.E. & Meyerowitz, E.M. (1983) J. Molec. Biol. 168, 765-789.
37. Godson, G.N., Ellis, J., Svec, P., Schlesinger, D.H. & Nussenzweig, V. (1983) Nature 305, 29-33.
38. Bayliss, G.J., Marsden, H.S. & Hay, (1975) J. Virology 68, 124-134.
39. Wadsworth, S., Hayward, G.S. & Roizman, B. (1976) J. Virol. 17, 503-512.