Volume 11 Number 18 1983 Nucleic Acids Research

Immediate-early mRNA-2 of herpes simplex viruses types 1 and 2 is unspliced: conserved sequences around the 5′ and 3′ termini correspond to transcription regulatory signals

J.Lindsay Whitton, Frazer J.Rixon, Andrew J.Easton+ and J.Barklie Clements*

MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK

Received 15 July 1983; Revised and Accepted 25 August 1983

SUMMARY

Nuclease S1 and exonuclease VII analyses of immediate-early (IE) mRNA-2 of herpes simplex viruses types 1 and 2 (HSV-1, HSV-2) show them to be unspliced and of similar length. The DNA sequences around the 5' and 3' termini have been determined. Comparison of the sequences around the 5' ends reveals several common features. (1) Four discrete blocks of upstream homology which are precisely colinear with respect to the 5' termini of the mRNAs; the blocks include the 'TATA' box, a G-C rich sequence and a sequence (AATTAAATACAT) which may be involved in the coordinate induction of the IE class of genes. (2) Several copies of the sequence CCCCGCCC, found in different upstream positions in HSV-1 and HSV-2, which may be important in the expression of a wide variety of eukaryotic genes. (3) Potential hairpin structures in the region of the 5' termini which are present at similar locations in HSV-1 and HSV-2. Sequence comparison around the 3' termini of IEmRNA-2 reveals high homology at the proposed C-terminus of the polypeptide.

INTRODUCTION

Herpes simplex viruses types 1 and 2 (HSV-1, HSV-2) although differing serologically (1,2) and exhibiting clinically distinct patterns of disease (3) are very similar in terms of both genome structure and genome organisation (4). The viruses have DNA genomes comprising two unique segments each of which is bounded by its own inverted repeats (Fig. 1). Although the overall homology between the two genomes as estimated by liquid hybridisation is only 50% (5), the distribution of homologous regions is apparently colinear (6). Furthermore, the analysis of HSV-1/HSV-2 intertypic recombinant viruses (7,8) both confirms the approximate colinearity of the genomes and indicates that certain encoded proteins, and presumably the DNA sequences controlling their expression, are interchangeable. This

conservation of genome organisation is reflected in the map locations and orientations of the immediate-early (IE) genes (Fig. 1; refs 9-11), a class of genes transcribed by host RNA polymerase II in the absence of de novo protein synthesis (12).

We have compared the equivalent regions of HSV-1 and HSV-2 which specify IEmRNA-2. This mRNA, which is located at approximate map coordinates 0.74 - 0.75, encodes a polypeptide of unknown function with a molecular weight of 63,000 in HSV-1 and of 65,000 in HSV-2 (7, 13). Gel electrophoretic analysis has shown that this mRNA is approximately 2Kb in length in HSV-1 and 1.75Kb in HSV-2 (14). We have determined the structure of IEmRNA-2 using nuclease digestion techniques and have compared the DNA sequences around the 5' and 3' ends in order to identify those sequences which are maintained by evolutionary constraints and which may therefore be of functional importance.

MATERIALS AND METHODS

Preparation of IE cytoplasmic mRNA

Baby hamster kidney 21 (C13) cells were grown as monolayers in rotating 80 oz. bottles. For the production of IE RNA, cell monolayers were infected with HSV-1 (Glasgow strain 17) or HSV-2 (strain HG 52) 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 (10). Cytoplasmic RNA was prepared using the method of Kumar and Lindberg (15).

Enzymes

All enzymes were obtained from Bethesda Research Laboratories or New England Biolabs with the exception of nuclease S1 (Boehringer Corporation Limited). They were used under the conditions outlined by the manufacturers.

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 E.Coli K12 HB101 and the cloning vector pAT153. Isolation of cloned DNA was as described previously (16).

Structural analysis of mRNAs

Structural analysis was performed using the nuclease digestion procedures of Berk and Sharp (17) which were modified by using either 5' or 3' end-labelled DNA probes (18).

Labelled DNA (less than 1 µg) was co-precipitated with known amounts of cytoplasmic mRNA from infected or mock-infected cells. The DNA/RNA pellet was resuspended in 20µl of 90% (v/v) formamide (deionised with Amberlite monobed resin MB-1), 0.4 M NaCl, 40mM PIPES, pH 6.8, 1mM EDTA. The mixture was heated to 90°C for 3 min, incubated at 57°C for either 5h or 16h, and then quenched on ice prior to nuclease digestion.

Nuclease S1 digestion was performed at 37°C for 1h in 200µl of 0.25M NaCl, 30mM NaOAc, pH 4.5, 1mM $ZnSO_4$ with 4000 units of nuclease S1. The nucleic acids were extracted with phenol/chloroform then precipitated with ethanol. The products were analysed by gel electrophoresis.

Exonuclease VII digestion was performed at 37°C for 1h in 200µl of 6.7mM KPO_4, pH 7.9, 8.3mM EDTA, 10mM 2-mercaptoethanol with 0.5 units of exonuclease VII. Thereafter, treatment was as for nuclease S1 digested samples.

Gel electrophoresis

Samples were electrophoresed on non-denaturing 1.5% (w/v) agarose gels in a buffer containing 90mM Tris, 90mM boric acid, pH 8.3, 1mM EDTA (TBE) or on alkaline 1.5% (w/v) agarose gels in 30mM NaOH, 2mM EDTA. Electrophoresis was carried out at room temperature for 16h at 50V and the gels dried down. Bands were visualised by autoradiography at -70°C using Kodak XS film. Denaturing polyacrylamide gels containing 9M urea (19) were run in TBE. 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 nucleotide sequences of DNA fragments, 3' or 5' end-labelled at a single site, were obtained by the chemical degradation procedure (19). All sequences presented were obtained from both DNA strands and/or in both directions.

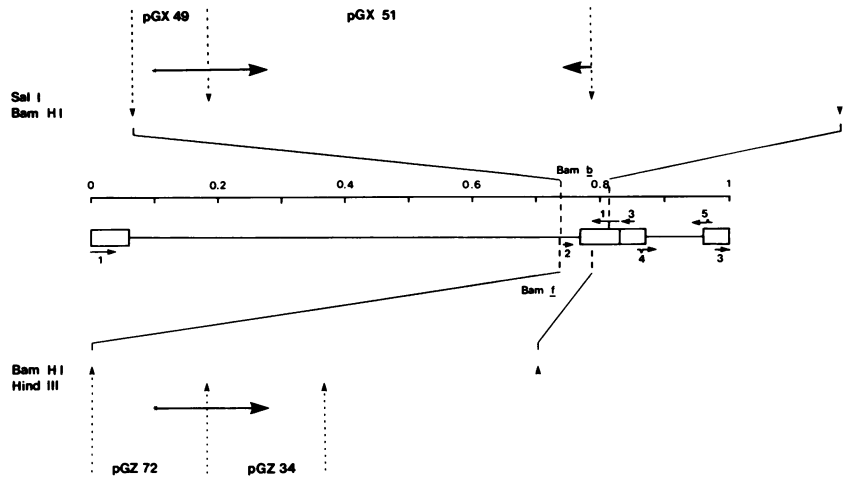

Fig. 1 Genomic map locations and orientations of the IEmRNAs (numbered 1-5) of HSV-1 and -2. The genome is shown in the P orientation (4): the open boxes represent the inverted repeats flanking the unique sequences. Fractional genomic lengths are indicated. The fragments Bam HI b of HSV-1 and Bam HI f of HSV-2 and their respective subclones (pGX49 & 51 and pGZ34 & 72) used as detailed in the text are shown on an expanded scale. The locations of IEmRNA-2 of HSV-1 and HSV-2 are shown with respect to these clones, as is the 3' terminus of IEmRNA-1 of HSV-1.

RESULTS AND DISCUSSION.

STRUCTURAL ANALYSES AND GENOME MAP LOCATIONS OF IEmRNA-2

Figure 1 shows the DNA clones used in the nuclease S1 and exonuclease VII analyses described below.

IEmRNA-2 of HSV-1.

The 3' end was located using pGX51 which was 3' end-labelled at the Sal I sites and hybridised to different amounts of HSV-1 cytoplasmic IE RNA. The resultant hybrids were digested with nuclease S1, and the digestion products analysed on neutral and alkaline agarose gels. A 790 bp nuclease S1-resistant band was present on both neutral and alkaline gels (Fig. 2A, tracks 2, 3, 7 & 8) indicating that the 3' portion is unspliced. The 240 base band also present (Fig. 2A, track 8) was generated by the 3' portion of IEmRNA-1 (Fig. 1), which was confirmed by separate hybridisations with each end of pGX51 (data not shown).

The 5' end was located using pGX49 which was 5' end-

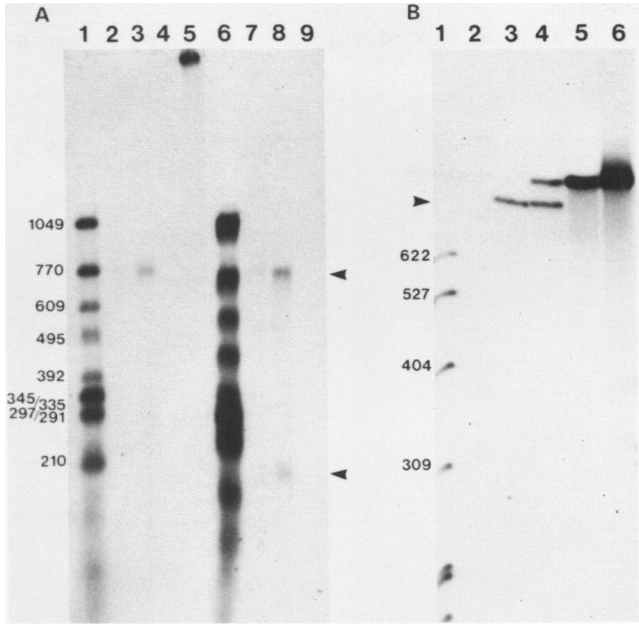

Fig. 2 Analysis of 3' and 5' portions of IEmRNA-2 of HSV-1. (A) pGX51 3' end-labelled at the Sal I sites was hybridised to: tracks 2 & 7, 5µg of IE cytoplasmic RNA; tracks 3 & 8, 15µg of IE cytoplasmic RNA; tracks 4 & 9, 20µg of mock-infected cytoplasmic RNA. Samples were digested with nuclease S1 as detailed in the text. Track 5 shows the intact DNA probe. End-labelled Hinc II digested øX174 RF DNA was run as size standard (tracks 1 & 6). The samples were electrophoresed on a 1.5% (w/v) neutral agarose gel (tracks 1 to 5) or on a 1.5% (w/v) alkaline agarose gel (tracks 6 to 9). (B) pGX49 5' end-labelled at the Sal I site was hybridised to : tracks 3 & 4, 15µg of IE cytoplasmic RNA; tracks 2 & 5, 20µg of mock-infected cytoplasmic RNA. Samples 2 & 3 were digested with nuclease S1 and samples 4 & 5 with exonuclease VII. Track 6 shows the intact DNA probe. End-labelled Hpa II digested pBR322 DNA was run as size standard (track 1). All samples were electrophoresed on an 8% denaturing polyacrylamide gel.

labelled at the Sal 1 site. Following hybridisation to HSV-1 IE RNA, a single nuclease S1-resistant band of 860 bases was seen on denaturing gels (Fig. 2B, track 3); exonuclease VII digestion also generated an 860 base band (Fig. 2B, track 4) indicating that the 5' portion of this mRNA also is unspliced. Thus the length of this unspliced mRNA is 1750 bases.

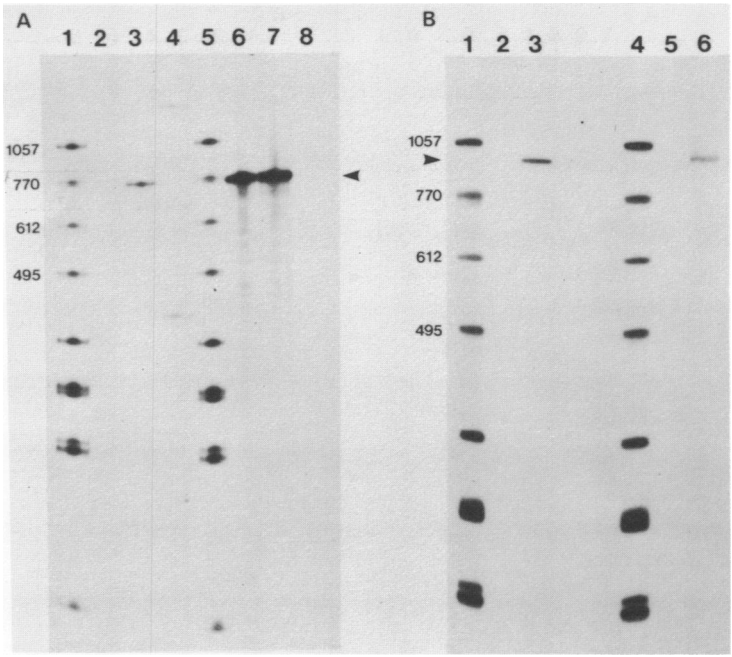

Fig. 3 Analysis of 3' and 5' portions of IEmRNA-2 of HSV-2. (A) pGZ34 was 3' end-labelled at the Hind III sites and hybridised to: tracks 2 & 6, 5µg of IE cytoplasmic RNA; tracks 3 & 7, 15µg of IE cytoplasmic RNA; tracks 4 & 8, 20µg of mock-infected cytoplasmic RNA. Samples 2-4 were digested with nuclease S1, and samples 6-8 with exonuclease VII. End-labelled Hinc II digested øX174 RF DNA was run as size standard (tracks 1 & 5). All samples were electrophoresed on a 6% denaturing polyacrylamide gel. (B) pGZ72 was 5' end-labelled at the Hind III site and hybridised to: tracks 2 & 5, 20µg of mock-infected cytoplasmic RNA; tracks 3 & 6, 15µg of IE cytoplasmic RNA. Samples 2 & 3 were digested with nuclease S1, and samples 5 & 6 with exonuclease VII. End-labelled Hinc II digested øX174 RF DNA was run as size standard (tracks 1 & 4). All samples were electrophoresed on a 6% denaturing polyacrylamide gel.

IEmRNA-2 of HSV-2.

The 3' end was located using pGZ34 which was 3' end-labelled at the Hind III sites. The probe was hybridised to HSV-2 cytoplasmic IE RNA and then digested with either nuclease S1 or exonuclease VII; the nuclease-resistant products were run on a denaturing polyacrylamide gel. A 765 base band was present in both the nuclease S1 and the exonuclease VII treated samples (Fig. 3A, tracks 2, 3, 6 & 7), showing this portion of the mRNA

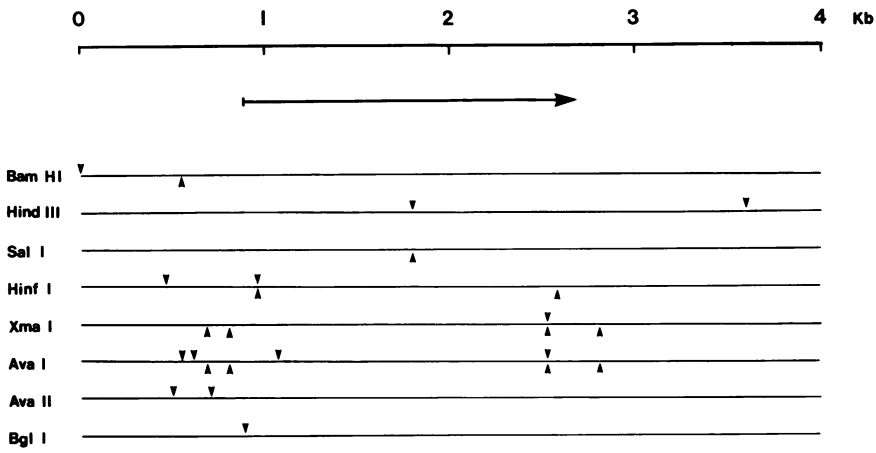

Fig. 4 Restriction endonuclease cleavage sites used in nuclease digestion analyses and in DNA sequencing. Sites present in HSV-2 are shown above the lines and those in HSV-1 below the lines. These maps are incomplete and show only those sites used in the work described in this paper. The arrow represents IEmRNA-2 and is drawn to scale.

to be unspliced.

The 5' end was positioned using pGZ72, 5' end-labelled at the Hind III site. A band of approximately 960 bases was generated by both nuclease S1 and exonuclease VII digestion (Fig. 3B, tracks 3 & 6), showing that this part of the mRNA is unspliced. The total length of this unspliced mRNA is therefore 1725 bases.

DNA SEQUENCE COMPARISON IN THE REGIONS OF THE mRNA 5' TERMINI.

The restriction endonuclease cleavage sites used in DNA sequence analysis and in the precise positioning of the 5' termini are shown in Figure 4. Figure 5 shows a comparison of the HSV-1 and HSV-2 DNA sequences around the 5' termini of the mRNAs; the sequences are aligned to give maximum homology. A Hinf 1 site common to both sequences (Fig. 5, position +50) was 5' end-labelled and used to locate exactly the 5' termini of the HSV-1 and HSV-2 mRNAs by co-electrophoresis of nuclease S1-resistant bands alongside the DNA sequence (data not shown).

Blocks of homology are present in the 5' flanking sequences.

Comparison of the 5' flanking sequences of numerous

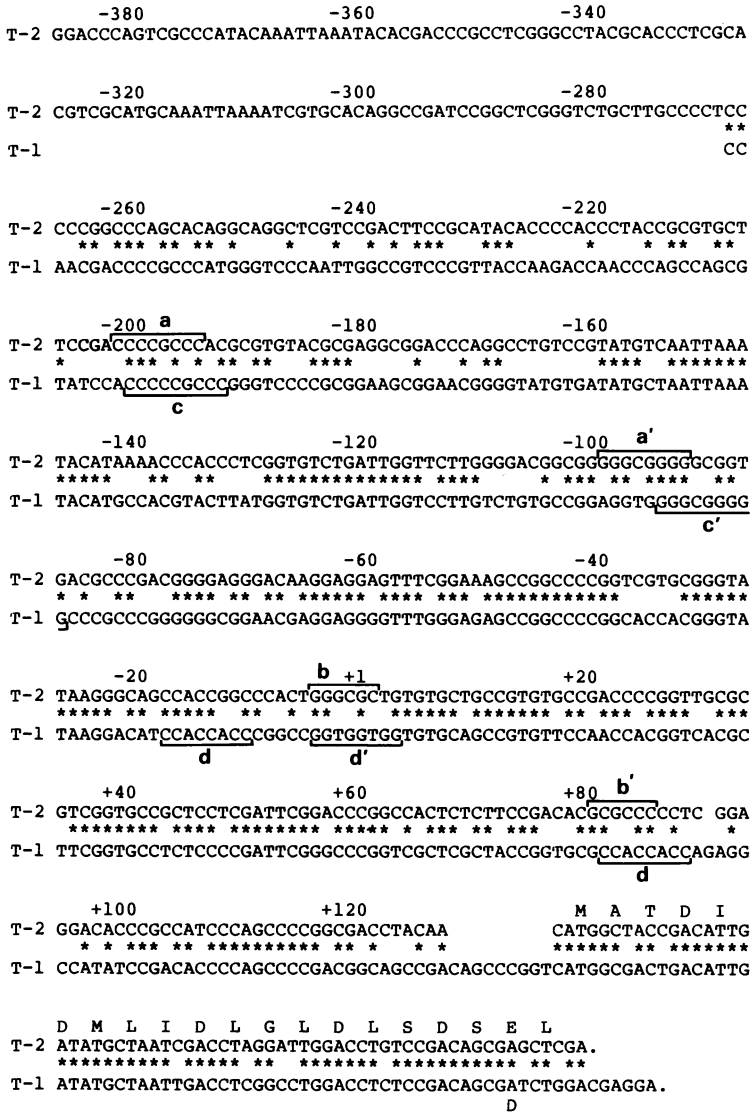

Fig. 5 Nucleotide sequence encoding the 5' portion of IEmRNA-2 of HSV-2 (T-2) and of HSV-1 (T-1). The sequences are aligned to give maximum homology and identical bases are indicated by an asterisk. The 5' termini of both mRNAs are at position +1. Inverted repeated sequences (a/a' & b/b' in HSV-2; c/c' & d/d'/d in HSV-1) are indicated. The amino acids encoded by the open reading frame beginning with the first ATG are shown in single letter code above the HSV-2 sequence. The HSV-1 amino acid sequence is identical with the single exception shown below the HSV-1 DNA sequence.

eukaryotic genes transcribed by RNA polymerase II has identified several blocks of conserved sequences only one of which, the 'TATA' box, is present in the sequence data shown here. Consensus sequences absent are (a) the cap site (consensus YCATTR; ref 20) although the mRNAs do follow the general pattern of initiating at a purine bracketed by pyrimidines; (b) the pentameric sequence (consensus GATCC) found in some eukaryotic genes at -35 (20); (c) the 'Chambon box' (consensus GGYCAATCT) located around -70 (21) which has been shown to be important in the regulation of β-globin gene expression (22-25).

When the HSV-1 and HSV-2 DNA sequences in the region of the 5' termini of IEmRNA-2 are aligned to give maximum homology (Fig. 5), the 5' termini of the two mRNAs are coincident on the DNA sequence, and four regions of high homology are revealed upstream of the 5' ends. Moreover, the distances of these conserved regions from the mRNA 5' termini are precisely maintained in both HSV-1 and HSV-2. The four blocks of homology are:

(a) CGGGTATAAGG (Fig. 5, positions -22 to -32) which corresponds to the 'TATA' box (consensus sequence $G\frac{T}{C}ATA\frac{T}{A}AAG$;refs 20, 26, 27) shown in various systems to be necessary for the precise positioning of 5' termini (22, 28-30), and which is present 5' to all HSV genes so far examined. In vitro assays using a region upstream of HSV-1 IEmRNA-3 from which the 'TATA' box had been deleted showed that transcription still took place, but the 5' end of the mRNA was no longer correctly positioned (31).

(b) a 39 bp stretch of 84% homology (Fig. 5, positions -38 to -76). Within these sequences is a 12 bp perfectly maintained sequence (AGCCGGCCCCGG). Part of this (CCGGCCCC) is found in direct or inverted orientation upstream of HSV-1 IEmRNAs -1 (32), -3 -4 and -5 (33) and this sequence is present in three copies upstream of the 5' terminus of HSV-2 IE mRNA-3. The locations of this sequence in HSV-2 do not match those of HSV-1 (34).

(c) a stretch GGTGTCTGATTGGT$\frac{T}{C}$CTTG (Fig. 5, positions -110 to -128) in which 18 out of 19 bases are conserved. No similar sequence has been identified upstream of other HSV-1 IE or HSV-2

IE genes.

(d) the sequence AATTAAATACAT (Fig. 5, positions -142 to -153) of which 11 out of 12 bases are present as a direct repeat in HSV-2 from -357 to -367; the HSV-1 sequence has not been extended to include this region. Sequences similar to this have been found 5' to all of the IEmRNAs examined to date, and a lengthy consensus suggested G[C/T]ATGNTAATGA[G/A]ATTC[C/T]TTGNGGG (32). HSV-1 IEmRNA-2 contains two sequences fitting this consensus (32); one is that identified in this paper at -142 to -153 and the second maps in the protein coding region, at position +156. The sequence at position -142 identified by us both by intertypic comparison and by its direct repetition at position -356 in HSV-2, is incompletely represented at position +156.

Non-colinear homologies.

A computer-assisted search of the sequences revealed one non-colinear homology of note. This sequence (CCCCGCCC) is present in HSV-1 at positions -254 and -192 and also as an inverted repeat at position -87. The sequence is present in HSV-2 in the direct orientation at position -194 and as an inverted repeat at position -92. This sequence is present in several copies in both orientations in the upstream region of IEmRNA-3 of HSV-1 (33) and HSV-2 (34).

Locations of the major inverted repeats.

Several inverted repeats upstream of HSV-1 IEmRNA-2 have been noted and a possible hairpin structure around the 5' end proposed (35). A computer-assisted search of our HSV-1 and HSV-2 DNA sequences shown in Figure 5 identified numerous inverted repeats of 6 bases or more (21 in HSV-1; 10 in HSV-2). The positions of the two longest HSV-2 inverted repeats (a/a' and b/b') together with the two longest of HSV-1 , the second of which is present as three copies (c/c' and d/d'/d), are shown in Figure 5. The a/a' and c/c' repeats represent the sequence CCCCGCCC and its inverted form discussed above. The potential HSV-1 hairpin (d/d') noted previously (35) has no counterpart at the equivalent location in HSV-2. However the possible hairpin (d'/d) does have a corresponding structure in HSV-2 (b/b') in a similar position relative to the mRNA 5' terminus, even though

the inverted repeats are dissimilar in sequence. This observation, and the similar positions of the HSV-2 a/a' and the HSV-1 c/c' inverted repeats, may indicate a role for these putative hairpin structures.

Homologies within the untranslated leader and in polypeptide coding sequences.

The untranslated leader of IEmRNA-2 is 10 bases shorter in HSV-2 than in HSV-1. The sequences shown in Figure 5 are aligned to give maximum homology across the region encoding this mRNA. Within the untranslated leader the overall homology is high, and the difference in length results largely from an insertion/deletion of 9 bases just prior to the first ATG. Comparison of untranslated leaders of IE mRNAs -3, -4 and -5 shows that such high homology is unusual within the IE class (34).

In both HSV-1 and HSV-2, the nucleotides flanking the first ATG correspond to the consensus sequence for an 'active' initiation codon (36). The homology within the proposed coding regions is higher than that in the leader; of the 8 base changes, 7 occur in the 3rd position of the triplets, and of the 20 codons compared, 19 specify identical amino acids. This supports the view that this ATG encodes the N-terminal amino acid of the respective HSV-1 and HSV-2 polypeptides.

DNA SEQUENCE COMPARISON IN THE REGIONS OF THE mRNA 3' TERMINI.

Figure 6 compares sequences around the 3' termini of the two mRNAs. The exact positions of the termini were determined by co-electrophoresis of nuclease S1-resistant bands alongside the DNA sequences (data not shown). The fragments used in sequencing and as probes were 3' end-labelled at the Xma 1 site (Fig. 6, position 1) for HSV-2 and at the Hinf I site (Fig. 6, position 59) for HSV-1.

The 3' terminal regions contain polyadenylation and other putative regulatory signals.

In common with all other polyadenylated HSV cytoplasmic mRNAs, both HSV IEmRNA-2 species contain the canonical AATAAA polyadenylation signal (37) upstream of their 3' termini. In addition, within a background of low homology a short sequence

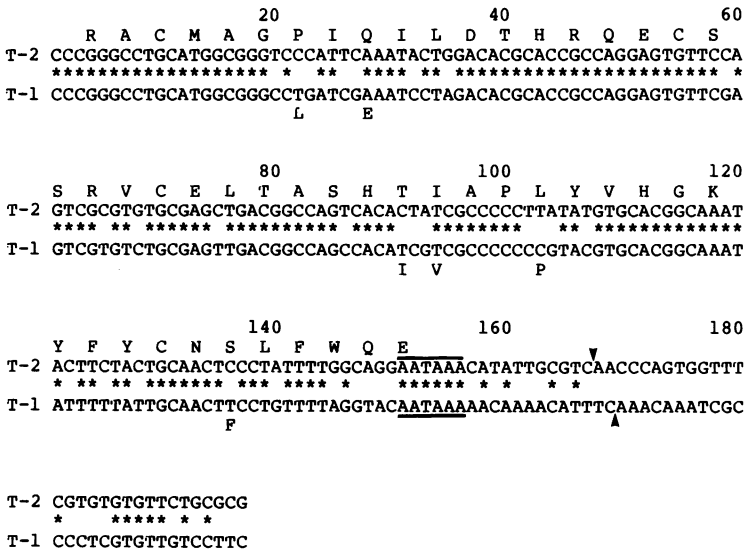

Fig. 6 Nucleotide sequences encoding the 3' portions of IEmRNA-2 of HSV-2 (T-2) and of HSV-1 (T-1). Identical bases are indicated by an asterisk. The AATAAA box is indicated and the 3' termini marked by arrows. The proposed HSV-2 amino acid sequence (see text) is shown in single letter code above the HSV-2 DNA sequence, with non-conserved amino acids shown below the HSV-1 DNA sequence.

(GTGTT; Fig. 6 position 186) is conserved 30 bases downstream of the AATAAA signal. This sequence corresponds to the core of the consensus YGTGTTYY found at this location downstream of many eukaryotic mRNAs (38).

The polypeptide coding region may extend into the AATAAA box in HSV-2.

There is a high degree of homology until position 143 (Fig. 6) and we suggest that both mRNAs are polypeptide coding to this point. Four pieces of evidence support this proposition. Firstly, the position at which homology decreases corresponds to a translational termination triplet (TAG) in the HSV-1 sequence. Secondly, the reading frame containing the stop codon is open until this point in both HSV-1 and HSV-2. Thirdly, of the 24 base mismatches in this sequence, 16 fall in the 3rd position in the codons of this reading frame. Fourthly, of the 47 amino acids that would be encoded, 41 are common to both sequences.

Assuming that these are the reading frames used, the HSV-2 polypeptide would be two amino acids longer at the C-terminal end. Translation would terminate at the TAA stop codon within the AATAAA box; a similar situation has been described for polypeptide IX of adenovirus type 12 (39).

GENERAL DISCUSSION

Our examination of the DNA sequences flanking IEmRNA-2 of HSV-1 and HSV-2 has identified conserved regions which we suggest are transcriptional regulatory signals. Sequence comparison of regions flanking other genes transcribed by RNA polymerase II has identified a number of consensus sequences, certain of which have been shown by functional assays to be important in the control of transcription; for example the 'TATA' homology, initially identified by comparison of 5' flanking regions, was subsequently found to be important in the positioning of the mRNA 5' terminus (22, 28-30).

We believe that comparison of the 5' flanking regions of HSV genes may be of particular interest as their expression is coordinately regulated and controlled by viral polypeptides. The IE class of genes is transcribed by RNA polymerase II in the absence of *de novo* protein synthesis and the 5' flanking regions of these genes appear to contain positive regulatory elements which are activated by a component of the viral inoculum, presumed to be a virion polypeptide (40). An HSV-2 viral inoculum can activate the upstream regions of HSV-1 IE genes (31), suggesting that the sequences of the regulatory elements involved in the coordinate induction are maintained and should therefore be identifiable by HSV-1/HSV-2 sequence comparison. In HSV-1 this element, along with the other signals necessary for active transcription, appears to lie within a fragment extending 270 bases upstream of the 5' terminus of IEmRNA-2 (35). The entire sequence of this region is presented in Figure 5, where it is compared to the equivalent HSV-2 tract. Comparison of these sequences should therefore reveal signals involved in the control of transcription including that required for coordinate induction. Such a comparison has revealed four discrete regions conserved in terms of both DNA sequence and position with

respect to the mRNA 5' termini, and a further conserved block located in different positions in HSV-1 and HSV-2. High overall homology between HSV-1 and HSV-2 cannot be invoked to explain the homologous regions, as upstream of position -159 (Fig. 5) the homology is low. One of the conserved sequences is the 'TATA' box located at position -22.

The A-T rich sequence at position -142 in both HSV-1 and HSV-2 (repeated at position -356 in HSV-2) is a good candidate for the IEmRNA-2 positive regulatory element. At least three regions containing a positive regulatory function have been identified upstream of HSV-1 IEmRNA-3 (31,41). Each of these regions (located at -110 to -160; -160 to -330; upstream of -330) contains at least part of an A-T rich sequence similar to that conserved upstream of HSV-1 and HSV-2 IEmRNA-2 and directly repeated upstream of the HSV-2 mRNA. We suggest that the sequence AATTAAATACAT is involved in the coordinate induction of HSV-1 and HSV-2 IEmRNA-2 by a virion polypeptide, in agreement with a suggestion previously made on the basis of comparison of the upstream regions of HSV-1 IE genes (32, 41). A-T rich sequences have been implicated in the coordinate induction of several classes of eukaryotic genes (42). For example, in the chick oviduct genes the sequence (AAAATGGGC) lies at position -140 (43); in the yeast his4 gene the sequence (consensus AAGTGACTC; ref 42) is represented as three direct repeats at positions -139, -183 and -195 (44), only one of which is required to allow the gene to be positively regulated (45).

The sequence CCGGCCCC is present upstream of all HSV IE genes examined to date. It is present in both direct and inverted orientations upstream of HSV-1 IEmRNA-3 (33) and deletion of all of one copy and part of the second copy abolishes transcription from this region (31); however as other sequences of possible significance are included in this deletion (eg. GGGCGGGG, discussed below), it is impossible to draw definitive conclusions from this result. Nevertheless, its conservation in terms of both sequence and position upstream of IEmRNA-2 supports a possible role in transcriptional regulation.

The stretch of 18 out of 19 matched bases from position -115 (Fig. 5) has no counterpart upstream of other HSV IE genes

so far examined and hence is unlikely to represent an IE class-specific regulatory element. However, the high degree of sequence conservation suggests some functional role, perhaps in gene-specific control.

The sequences CCCCGCCC and its inverted form GGGCGGGG are represented by three copies in HSV-1 and two in HSV-2 (Fig. 5). Unlike the other homologies identified by intertypic comparison, the precise positions of these sequences from their respective mRNA 5' termini are not conserved. This sequence is present at position -98 upstream of the HSV-1 thymidine kinase gene and base changes within the sequence lead to a twenty-fold reduction in transcription levels of this early gene (46). The presence of this sequence upstream of both IE and early HSV genes argues against a role as a class-specific regulatory sequence. Similar sequences have been identified upstream of several other eukaryotic genes. For example, the sequence YYCCGCCC is present in 6 non-tandem repeats upstream of the SV40 early promoter and the production of T antigen requires the presence of at least some of these repeats (28, 47, 48). Furthermore, the amount produced is proportional to the number of repeats present (47, 48) and they appear to function in either orientation (48). The sequence CCNCACCC (where N is A, T or C) is present as direct repeats at positions -97 and -111 upstream of the rabbit β-globin gene; these repeats are important in transcriptional control of this gene and the same sequence is present upstream of the β-globin genes of other species (24). These observations suggest that this family of related sequences may be important in the regulation of a wide variety of eukaryotic genes.

The comparative approach to the identification of putative regulatory signals adopted in this paper has allowed the precise delimitation of conserved sequences certain of which have previously been shown to be involved in transcriptional regulation. Additional conserved elements have been identified also, and their function may be assayed by sequence modification or by their insertion as separate modules into a eukaryotic expression vector. These studies are currently in progress.

Acknowledgments
We thank Professor J. H. Subak-Sharpe for continued interest and helpful discussion, and Moyra E. Campbell for her excellent technical assistance.
JLW was the recipient of a Medical Research Council Training Fellowship number G84/785. AJE was the recipient of a Medical Research Council grant for training in research techniques. The work was supported by a grant from MRC G978/709/5B.

+Current address: Searle Research and Development, Lane End Road, High Wycombe, Bucks, UK

*To whom correspondence should be addressed

REFERENCES

1. Pauls, F.P. and Dowdle, W.R. (1967) J. Immun., 98, 941-947.
2. Plummer, G. (1964) Brit. J. Exp. Pathol., 45, 135-141.
3. Dowdle, W.R., Nahmias, A.J., Harwell, R.W. & Pauls, F.P. (1967) J. Immun., 99, 974-980.
4. Roizman, B., Carmichael, S., de The, G.B., Masic, M., Nahmias, A., Plowright, W., Rapp, F., Sheldrick, P., Takashi, M., Terni, M. and Wolfe, K. (1978) in Oncogenesis and Herpesviruses III vol 2, pp 1079-1082. Eds., de The, G., Rapp, F. & Henle, W. IARC Scientific Publications, No. 24, Lyon.
5. Kieff, E.D., Hoyer, B., Bachenheimer, S.L. & Roizman, B. (1972) J. Virol., 9, 738-745.
6. Davison, A.J. & Wilkie, N.M. (1983) in press.
7. Preston, V.G., Davison, A.J., Marsden, H.S., Timbury, M.C., Subak-Sharpe, J.H. and Wilkie, N.M. (1978) J. Virol., 28, 449-517.
8. Morse, L.S., Pereira, L., Roizman, B. & Schaffer, P.A. (1978) J. Virol., 26, 389-410.
9. Clements, J.B., McLauchlan, J. & McGeoch, D.J. (1979) Nucleic Acids Res., 7, 77-91.
10. Clements, J.B., Watson, R.J. & Wilkie, N.M. (1977) Cell, 12, 275-285.
11 Easton, A.J. & Clements, J.B. (1982) Nucleic Acids Res., 8, 2627-2645.
12. Costanzo, F., Campadelli-Fiume, G., Foa-Tomasi, L. & Cassai, E. (1977) J. Virol., 21, 996-1001.
13. Watson, R.J., Preston, C.M. & Clements, J.B. (1979) J. Virol., 31, 42-52.
14. Clements, J.B., Easton, A.J. & Rixon, F.J. (1981) in Herpes simplex virus DNA: recent studies on the internal organisation and replication of the viral genome. Developments in Molecular Virology, vol. 1, pp27-42, Ed Y. Becker, Martinus Nijhoff, B.V., The Hague.
15. Kumar, A. & Lindberg, U. (1972) Proc. Natl. Acad. Sci. USA, 69, 681-685.
16. Davison, A.J. & Wilkie, N.M. (1981) J. Gen. Virol., 55, 315-331.
17. Berk, A.J., & Sharp P.A. (1978) Proc. Natl. Acad. Sci. USA, 75, 1274-1278.

18. Weaver, R.F. & Weismann, C. (1979) Nucleic Acids Res., 7, 1175-1193.
19. Maxam, A.M. & Gilbert, W. (1980) Methods in Enzymology 65, 499-560.
20. Busslinger, M., Portmann, R., Irminger, J.C. & Birnsteil, M.L. (1980) Nucleic Acids Res., 8, 957-978.
21. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) Nucleic Acids Res., 8, 127-142.
22. Grosveld, G.C., de Boer, E., Shewmaker, C.K. & Flavell, R.A. (1982) Nature, 295, 120-126.
23. Dierks, P., van Ooyen, A., Mantei, N. & Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA, 78, 1411-1415.
24. Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. & Weissmann, C. (1983) Cell, 32, 695-706.
25. Grosveld, G.C., Rosenthal, A. & Flavell, R.A. (1982) Nucleic Acids Res., 10, 4951-4971.
26. Proudfoot, N.J. Nature, 279, 376.
27. Gannon, F., O'Hare, K., Perrin, F,. LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. & Chambon, P. (1979) Nature, 278, 428-434.
28. Benoist, C. & Chambon, P. (1981) Nature, 290, 304-310.
29. Mathis, D.J. & Chambon, P. (1981) Nature, 290, 310-315.
30. Grosschedl, R. & Birnsteil, M.L. (1980) Proc. Natl. Acad. Sci. USA, 77, 1432-1436.
31. Cordingley, M.G., Campbell, M.E. & Preston, C.M. (1983) Nucleic Acids Res., 11, 2347-2365.
32. Mackem, S. & Roizman, B. (1982) J. Virol., 44, 939-949.
33. Murchie, M-J. & M^{c}Geoch, D.J. (1982) J. Gen. Virol., 62, 1-15.
34. Whitton, J.L. & Clements, J.B., in preparation.
35. Mackem, S. & Roizman, B. (1982) J. Virol., 43, 1015-1023.
36. Kozak, M. (1981) Nucleic Acids Res., 9, 5233-5252.
37. Proudfoot, N.J. & Brownlee, G.G. (1976) Nature, 263, 211-214.
38. M^{c}Lauchlan, J. & Clements, J.B., manuscript submitted.
39. Bos, J.C., Polder, L.J., Schrier, P.I., van den Elsen, P.J., van der Eb, A.J. & van Ormondt, H. (1981) Cell, 27, 121-131.
40. Post, L.E., Mackem, S. & Roizman, B. (1981) Cell, 24, 555-566.
41. Mackem, S. & Roizman, B. (1982) Proc. Natl. Acad. Sci. USA, 79, 4917-4921.
42. Davidson, E.H., Jacobs, H.T. & Britten, R.J. (1983) Nature, 301, 468-470.
43. Grez, M., Land, H., Giesecke, K. & Schutz, G. (1981) Cell, 25, 743-752.
44. Donahue, T.F., Farabaugh, P.J & Fink, G.R. (1982) Gene, 18, 47-52.
45. Donahue, T.F., Davis, R.S., Lucchini, G. & Fink, G.R. (1983) Cell, 32, 89-98.
46. M^{c}Knight, S.L. & Kingsbury, R. (1982) Science, 217, 316-324.
47. Fromm, M. & Berg, P. (1982) J. Mol. Appl. Genet., 1, 457-481.
48. Everett, R., Baty, D. & Chambon, P. (1983) Nucleic Acids Res., 11, 2447-2464.