Characterization of the IE110 Gene of Herpes Simplex Virus Type 1

By LISE J. PERRY, FRAZER J. RIXON,† ROGER D. EVERETT,† MARGARET C. FRAME† AND DUNCAN J. McGEHOCH*†

Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, U.K.

(Accepted 18 July 1986)

SUMMARY

We have determined the DNA sequence of the herpes simplex virus type 1 (HSV-1) gene encoding the immediate early protein IE110, which is involved in transcriptional activation of later virus genes. The locations of the 5’ and 3’ termini of IE110 mRNA, together with the positions of two introns, were identified. Examination of the DNA sequence suggested that translation starts at the first ATG after the 5’ terminus of the mRNA, and that both introns occur in protein-coding sequence. The predicted IE110 polypeptide contains 775 amino acids, and has a molecular weight of 78,452. It contains a cysteine-rich region resembling regions found in several proteins which interact functionally with DNA. An antiserum was raised to the predicted C terminal amino acid sequence of the IE110 polypeptide and was shown to immunoprecipitate the native protein from HSV-1-infected cell extracts. The functional importance of regions of the protein was evaluated by construction of frameshift and deletion mutants of a plasmid-borne IE110 gene. The mutants were tested for IE110 function by short-term transfection assays, and the results were correlated with the DNA sequence and RNA mapping studies.

INTRODUCTION

After infection of tissue culture cells with herpes simplex virus type 1 (HSV-1), the first virus genes to be expressed are the five immediate early (IE) genes (Honess & Roizman, 1974; Watson et al., 1979; Anderson et al., 1980). It has long been recognized that the product of IE gene 3, IE175 (also known as ICP4), has a major role in activation of transcription of early and late genes (Preston, 1979; Watson & Clements, 1980; Dixon & Schaffer, 1980). The function of IE175 was demonstrated by the study of HSV-1 temperature-sensitive (ts) mutants with lesions in IE gene 3, but until recently lack of similar mutants in the other IE genes has hindered studies on their function. In the last 2 years, experimental schemes have been developed for assay of IE gene function using plasmid-cloned IE genes (Everett, 1983; Everett & Dunlop, 1984; Gelman & Silverstein, 1985; O'Hare & Hayward, 1985a, b). These can be introduced into culture cells to give expression of IE proteins, whose effect on transcription from HSV-1 promoters, introduced in a separate plasmid, can then be measured. Experiments of this type have shown that another IE protein, IE110 (also called ICP0), encoded by IE gene 1, can act as a transcriptional activator, either independently or in cooperation with IE175 (Everett, 1984; O'Hare & Hayward, 1985a, b; Quinlan & Knipe, 1985; Mavromara-Nazos et al., 1986).

For both IE175 and IE110, the mechanisms of transactivation remain obscure. It is not clear whether either protein acts directly (that is, by binding to DNA in the locality of the target transcription initiation site) or indirectly (for instance, by interaction with some other component, which ultimately affects transcription). Nor are relations between the actions of the two proteins clear. For instance, they could act at the same point of the activation process, or sequentially at different levels of the one process, or in parallel in different activation processes.

† Members of the MRC Virology Unit.
However, the existence of at least two HSV-specified transactivators of virus gene expression certainly supports the view that control of HSV transcription may be far from simple. Information on the IE gene 1 and its encoded protein is limited. The gene is located entirely within the long repeat element (R1; see Fig. 1) of the HSV-1 genome, and is thus diploid (Preston et al., 1978). However, the existence of deletion variants of HSV-1 shows that only one copy suffices for virus replication (Davison et al., 1981). IE110 is a phosphoprotein which accumulates in the nuclei of infected cells (Pereira et al., 1977). The protein has an estimated molecular weight of 110000 (Marsden et al., 1976) and has been shown to bind DNA (Hay & Hay, 1980).

In this paper, we report the DNA sequence of IE gene 1 and describe mRNA mapping experiments which show that the gene contains two introns, both of which are considered to be in protein-coding DNA. We give the deduced amino acid sequence of IE110, and show that an anti-peptide serum to the predicted C terminus of IE110 binds to the whole protein. Finally, we have constructed deletion and frameshift mutations within the gene and have evaluated the phenotype of the resulting mutated IE110 derivatives.

**METHODS**

**Plasmids.** Recombinant plasmids carrying HSV-1 strain 17 restriction fragments were used as follows. DNA sequence analysis used BamHI b cloned into the BamHI site of pAT153 (this plasmid was called pGX48), BamHI k, also in pAT153 (from A. J. Davison) and XhoI c in the XhoI site of pMK16 (pGX58, from N. D. Stow). In addition to pGX48 and pGX58, mRNA mapping experiments also used pGX53, which consists of the Sall–BamHI subfragment from the right end of the BamHI b fragment cloned into pAT153, and plasmid pJR3 consisting of the SsrI–PstI fragment from IR1 (including the complete IE110 gene) cloned into pUC9, obtained from J. Russell and C. M. Preston. Functional assays of the IE110 gene used pJR3 described above, and p110 consisting of the HindIII–SsrI fragment of pJR3 (containing all the HSV sequences of pJR3) cloned into the HindIII–SsrI sites of pBRSt (where pBRSt is a derivative of pBR322 in which the PvuII site has been converted to an SsrI site). pJR3 and p110 are similar except that p110 contains only a single BamHI site, while pJR3 has an additional BamHI site in the vector sequences. p175, used in later experiments, was derived from pGX58 and contained the coding sequences of the IE110 gene under the control of the simian virus 40 early promoter and enhancer, cloned into the EcoRI–PvuII fragment of pBR322, so that the upstream parts of IE genes 1 and 4 present in pGX58 were excluded. Finally, plasmid pgDCAT has the promoter for the glycoprotein D gene of HSV-1 linked to the chloramphenicol acetyltransferase (CAT) gene (McLauchlan et al., 1985). The CAT coding region is linked to an HSV-2 polyadenylation signal and the construct cloned into pUC9. Additional derivatives of pJR3 and p110 were constructed as described below.

**DNA sequence analysis.** The DNA sequence was determined by M13 chain termination reactions (Sanger et al., 1980), using M13 clones generated by ligation of sonicated DNA into the Smal site of M13mp8 (Messing & Vieira, 1982). The products of the sequencing reactions, labelled with [α-32P]dATP, were electrophoresed through 9 M-urea, 6% polyacrylamide, buffer gradient gels (Biggin et al., 1983). Where necessary, to resolve compressions arising from the high G + C composition of the DNA, 6% polyacrylamide gels were maintained at a high temperature during the run using a hot water jacket (80 to 85 °C).

**Computing.** Computing was performed with a DEC PDP 11/44 under RSX11M. Sequence analysis used the database system of Staden (1982). Codon usage evaluations used the program of Staden & McLachlan (1982). Sequence homologies were evaluated with a matrix comparison program (Pustell & Kafatos, 1982) and an alignment optimizing program (Taylor, 1984).

**mRNA mapping.** Cytoplasmic RNA was prepared by the method of Kumar & Lindberg (1972). For production of HSV-1 IE mRNA, cell monolayers were infected at a multiplicity of infection of 50 p.f.u./cell and the cell monolayers were pretreated and maintained in medium containing cycloheximide (Clements et al., 1976). Structural analysis was performed as described previously (Rixon & Clements, 1982). Briefly, either 3' or 5' labelled DNA was co-precipitated with cytoplasmic RNA from infected or mock-infected cells. The DNA/RNA mixture was denatured at 100 °C for 3 min and incubated for 16 h at 57.5 °C in 20 μl 90% (v/v) deionized formamide, 0.4 M-NaCl, 40 mM-PIPES pH 6.8, 1 mM-EDTA. Nuclease S1 and exonuclease VII digestions were performed at 37 °C.

**Production of anti-oligopeptide serum.** The peptide NH₂–Tyr–Glu–Gly–Ala–Ser–Thr–Arg–Asp–Glu–Gly–Lys–Gln–COOH (synthesized by Cambridge Research Biochemicals, Cambridge, U. K.) corresponds to the C terminal 11 amino acids of the predicted IE110 polypeptide linked to a Tyr residue. After coupling the peptide to bovine serum albumin (BSA), antisera were raised in rabbits as previously reported (Frame et al., 1986). Antibodies raised against BSA were removed by passing the immune serum through a BSA–Sepharose column.
Structure of the HSV-1 IE110 gene

RESULTS

Mapping of IE gene 1 mRNA

As shown in Fig. 1, IE gene 1 is located in the R_L element of the HSV-1 genome. We have determined the complete sequence of R_L, together with adjacent regions of U_L: the whole sequence will be presented elsewhere (L. J. Perry & D. J. McGeoch, unpublished). Residue numbering in this paper is based on the DNA sequence of R_L, starting with the residue in IR_L adjacent to the a' sequence (see legend to Fig. 1). Previous work had mapped the 5' terminus of mRNA-1 to residue 1714 (Mackem & Roizman, 1982) and the 3' terminus to residue 5301 (Rixon et al., 1984) on the leftward 5' to 3' strand of IR_L, giving a span of 3587 residues. Since the size of the mRNA, including the poly(A) tail, was estimated from agarose gel electrophoresis to be 3 kb (Watson et al., 1979), it seemed probable that the mRNA was spliced.

Fig. 2 gives a summary of the nuclease digestion mapping data described in the text and shows the structure of IE mRNA-1 as determined by these analyses. In the following descriptions, single positions are given for the mRNA 5' and 3' ends and for all splice sites. Where detailed analysis of small nuclease-resistant products revealed a number of closely spaced bands (probably due to imprecise nuclease cleavage at the hybrid ends), the size given is that which aligns the RNA with the appropriate splice junction recognition site on the DNA. The slightly larger size of the exonuclease VII-resistant bands compared to the nuclease S1-resistant bands is due to the processive nature of the exonuclease VII activity (Rixon & Clements, 1982). As described below, the DNA sequence encoding IE mRNA-1 has a high G + C composition. Furthermore, the distribution of G + C is not uniform throughout the gene, with certain regions having a particularly extreme G + C content. This causes difficulties in the analysis of RNA structure by nuclease digestion, since localized melting of double-stranded molecules in regions of lower G + C will occur and complex annealing of DNA and RNA involving more than two molecules may take place. In several of the examples described below more than one band was present following nuclease digestion. In each of these cases, only one of the bands coincided with a recognizable RNA processing signal. The other bands generally correspond to short A + T-rich stretches in an otherwise high G + C sequence. These bands were assumed to result from...
Fig. 1. Location and organization of the IE110 gene. The upper part of the figure shows the prototype arrangement of the HSV-1 genome. The long and short unique regions (UL and Us) are shown as solid lines, and the major repeat elements as open boxes. Terminal and internal long repeats are marked as TR_L and IRL, and the short repeat copies as TRs and IRS. The lower part of the figure shows an expansion of the IRL region. Numbering is based on our sequence data, starting with the residue adjacent to the α' sequence, as defined by Davison & Wilkie (1981). The region whose sequence is given in Fig. 4 is indicated by dotted lines. The location and structure of IE110 mRNA is shown, with introns. Proposed protein-coding regions are shown as open boxes. Locations of restriction fragments used for sequence analysis are shown (BamHI b, BamHI k and XhoI c).

Fig. 2. Structure of IE110 mRNA. In IRL, the direction of transcription of IE110 mRNA is from right to left. The gene contains two introns. Numbering is based on the sequence data, starting with the residue adjacent to the α' sequence. The lower part of the figure shows a summary of the nuclease mapping data described in the text. The horizontal lines indicate the size of each protected DNA fragment, from the restriction enzyme site used to map the mRNA to the mapped terminus or intron/exon boundary. The restriction enzyme site used is given by the following letters: Ba, BamHI; Hc, HincII; Hf, HinfI; Nc, NcoI; Sa, Sau3A1; Sl, SalI; Xh, XhoI. The vertical broken line indicates an XhoI site that was used more than once.

localized melting of DNA/DNA or DNA/RNA hybrids. However, it is possible that some of these bands may represent minor or aberrant splicing events.

Initial attempts to determine the structure of IE mRNA-I by nuclease digestion procedures were made using DNA fragments which had been labelled at the BamHI site at position 2508. No nuclease-resistant material was detected with either 5' labelled or 3' labelled DNA. In view of the discrepancy between the size of the mRNA and of the DNA sequence encoding it, mentioned above, it seemed probable that the BamHI site lay within an intron.

The 5' end of IE mRNA-I had previously been identified by Mackem & Roizman (1982). This was confirmed using pJR3 DNA which had been 5' labelled at the unique NcoI site at position 1866. This generated nuclease-resistant bands of around 151 bases length, placing the 5' end of
the IE mRNA-1 at position 1716. The 5' boundary of the putative intron was identified using a 3' labelled, 491 bp *HinI* subfragment of pJR3 (positions 1767 to 2256). This generated nuclease S1-resistant and exonuclease VII-resistant bands of around 154 bases (Fig. 3a) placing the splice donor site at position 1920. To locate the 3' boundary of this intron, the 435 bp *XhoI/BamHI* subfragment of pGX48 was 5' labelled at the *XhoI* site at position 2943. This generated two nuclease S1-resistant species, a major band of around 258 bases and a minor band of around 280 bases, and a single exonuclease VII-resistant band of around 258 bases (Fig. 3b). Examination of the DNA sequence encoding this region revealed that only the 258 base band coincides with a recognizable splice acceptor sequence at position 2686. The 280 base band coincides with the start of a relatively A + T-rich sequence which does not resemble a splice acceptor site. The data presented here do not exclude the possibility that small exons may exist between the mapped splice donor and acceptor sites. However, since the exonuclease VII-resistant and nuclease S1-resistant bands are of similar sizes, the introns must extend beyond the limits of the DNA fragments used for mapping (to within positions 2256 to 2508). Conclusive proof of the absence of more complicated splicing patterns will probably require cDNA cloning and sequence determination.
To determine whether the remaining 3' portion of IE mRNA-1 was unspliced, the 2561 bp XhoI/HpaI subfragment of pGX48 (positions 2943 to 5503) was 3' labelled at the XhoI site (position 2943). This generated three nuclease S1-resistant bands of 409, 427 and 470 bases (Fig. 3e). Examination of the DNA sequence of this region revealed that the 409 base band coincides with a potential splice donor sequence at position 3352. The two other nuclease S1-resistant products coincide with short A + T-rich regions which do not resemble any known splice donor signals. The position of this splice site was confirmed using a 3' labelled, 1063 bp HinfI subfragment (positions 2800 to 3861) of pGX53. This generated three nuclease S1-resistant bands of 552, 570 and 613 bases (data not shown), which correspond to those observed with the XhoI/HpaI fragment. It is clear from these results that the gene encoding IE mRNA-1 contains a second intron. Exonuclease VII digestion generated only full-length probe DNA, indicating that the intron lay entirely within this HinfI fragment. Therefore, the 3' boundary of this second intron was identified using the same 1063 bp HinfI subfragment of pGX53, which had been 5' labelled. This hybridization was performed at both 57.5 °C and at 60 °C and gave a complicated pattern of bands at both temperatures (Fig. 3d). However, there were differences in the abundance of the bands at the two temperatures, with a considerable increase in the relative intensity of a 375 base band at the higher temperature, suggesting that it represents a more stable hybrid. This band coincides with a possible splice acceptor signal at position 3488, which we propose represents the intron boundary. A similar pattern of bands was obtained (data not shown) with a 5' labelled, 549 bp HinclI/KpnI fragment (positions 3262 to 3811). In this case, the increase in relative intensity at the higher temperature was exhibited by a 322 base band which corresponds to the 375 base band produced with the HinfI fragment. We emphasize that the difficulties with these analyses are considered to result from the extreme base composition, and that finally definitive results will probably require cDNA cloning and sequence analysis.

The remainder of the 3' portion of IE mRNA-1 was analysed using 3' labelled 1470 bp HinfI (positions 4376 to 5843) and 1685 bp Sau3A1 (positions 3855 to 5540) subfragments of pJR3. These generated nuclease S1-resistant bands of 925 bases (HinfI) and 1446 bases (Sau3A1) respectively (data not shown), placing the 3' end at the previously mapped location, position 5301 (Rixon et al., 1984).

From the above data, the structure of the gene for IE mRNA-1 is shown to consist of three exons of 205, 667 and 1812 bases, separated by two introns of 767 and 136 bases.

**Sequence of IE gene 1**

Fig. 4 represents the DNA sequence of residues 1201 to 5520 of IR1. This was determined by the M13/chain terminator method, using plasmid-cloned copies of restriction fragments BamHI b, BamHI k and XhoI c, as shown in Fig. 1. The positions of the 5' and 3' termini of mRNA-1 are marked at residues 1716 and 5301 respectively. There is a potential TATA sequence at 1689 to 1693. The promoter and activator regions 5' to the 5' terminus of the mRNA have been examined by Mackem & Roizman (1982). At the downstream end of the gene, adjacent to the 3' terminus of the mRNA, there are two copies of the polyadenylation-associated sequence AATAAA, at residues 5239 and 5279. Downstream of the 3' terminus lies a set of nine tandem copies of a 16 bp sequence (Rixon et al., 1984). Also present downstream of the 3' terminus of mRNA-1 is the sequence TGTGTTGG, at 5312 to 5319, which resembles the consensus YGTGTTYY identified by McLauchlan et al. (1985) as being required for efficient formation of mRNA 3' termini.

There are consensus splice donor and acceptor sequences close to the mapped boundaries of the two introns (Mount, 1982). We thus consider that intron 1 runs from 1921 to 2685 and intron 2 from 3353 to 3488. The sequence of intron 1 contains three imperfect tandem copies of a 54 bp sequence and also a number of other imperfect repeat elements. Starting at residue 2647, in intron 1, Fig. 4 shows a run of eight C residues, as found in the BamHI b clone analysed, while the XhoI c clone contains nine residues here. The sequence of intron 2 is particularly purine-rich.

We have located the polypeptide-coding sequence for IE110 by examination of open reading frames and use of the codon usage evaluation program of Staden & McLachlan (1982). We conclude that translation begins with the first ATG after the 5' terminus of the mRNA, at 1864,
Table 1. Codon usage catalogue of the IE110 gene

<table>
<thead>
<tr>
<th>Codon</th>
<th>Ser</th>
<th>Tyr</th>
<th>Cys</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT Phe</td>
<td>2</td>
<td>TAT</td>
<td>TGT</td>
</tr>
<tr>
<td>TTC Phe</td>
<td>7</td>
<td>TAC</td>
<td>TGC</td>
</tr>
<tr>
<td>TTA Leu</td>
<td>1</td>
<td>TAA</td>
<td>TGA</td>
</tr>
<tr>
<td>TTA Leu</td>
<td>1</td>
<td>TAC</td>
<td>TGC</td>
</tr>
<tr>
<td>CTT Leu</td>
<td>2</td>
<td>CAT</td>
<td>CGT</td>
</tr>
<tr>
<td>CTC Leu</td>
<td>6</td>
<td>CAC</td>
<td>CGC</td>
</tr>
<tr>
<td>CTA Leu</td>
<td>1</td>
<td>CAA</td>
<td>CGA</td>
</tr>
<tr>
<td>CTT Leu</td>
<td>26</td>
<td>CAG</td>
<td>CGG</td>
</tr>
<tr>
<td>ATT Ile</td>
<td>0</td>
<td>AAT</td>
<td>AGT</td>
</tr>
<tr>
<td>ATC Ile</td>
<td>1</td>
<td>AAG</td>
<td>AGG</td>
</tr>
<tr>
<td>ATA Ile</td>
<td>2</td>
<td>AAA</td>
<td>AGA</td>
</tr>
<tr>
<td>ATG Met</td>
<td>11</td>
<td>AAG</td>
<td>AGG</td>
</tr>
<tr>
<td>GTT Val</td>
<td>5</td>
<td>GAT</td>
<td>GGT</td>
</tr>
<tr>
<td>GTC Val</td>
<td>1</td>
<td>GAA</td>
<td>GGG</td>
</tr>
<tr>
<td>GTA Val</td>
<td>1</td>
<td>GAC</td>
<td>GGC</td>
</tr>
<tr>
<td>GTG Val</td>
<td>26</td>
<td>GAG</td>
<td>GGG</td>
</tr>
</tbody>
</table>

Structure of the HSV-1 IE110 gene

Table 2. Predicted amino acid composition of IE110

<table>
<thead>
<tr>
<th>Residue</th>
<th>No.</th>
<th>%</th>
<th>Residue</th>
<th>No.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>110</td>
<td>14.2</td>
<td>Leu</td>
<td>36</td>
<td>4.6</td>
</tr>
<tr>
<td>Arg</td>
<td>66</td>
<td>8.5</td>
<td>Lys</td>
<td>8</td>
<td>1.0</td>
</tr>
<tr>
<td>Asn</td>
<td>18</td>
<td>2.3</td>
<td>Met</td>
<td>11</td>
<td>1.4</td>
</tr>
<tr>
<td>Asp</td>
<td>39</td>
<td>5.0</td>
<td>Phe</td>
<td>9</td>
<td>1.2</td>
</tr>
<tr>
<td>Cys</td>
<td>14</td>
<td>1.8</td>
<td>Pro</td>
<td>103</td>
<td>13.3</td>
</tr>
<tr>
<td>Gln</td>
<td>20</td>
<td>2.6</td>
<td>Ser</td>
<td>78</td>
<td>10.1</td>
</tr>
<tr>
<td>Glu</td>
<td>36</td>
<td>4.6</td>
<td>Thr</td>
<td>48</td>
<td>6.2</td>
</tr>
<tr>
<td>Gly</td>
<td>94</td>
<td>12.1</td>
<td>Trp</td>
<td>7</td>
<td>0.9</td>
</tr>
<tr>
<td>His</td>
<td>15</td>
<td>1.9</td>
<td>Tyr</td>
<td>7</td>
<td>0.9</td>
</tr>
<tr>
<td>Ile</td>
<td>13</td>
<td>1.7</td>
<td>Val</td>
<td>43</td>
<td>5.5</td>
</tr>
</tbody>
</table>

giving a 5' non-coding region of 148 residues. This ATG, in exon 1, conforms to the initiation codon consensus of Kozak (1984), and initiates an open reading frame (allowing for introns) of 775 codons, terminating with TAA at 5090. This leaves a 3' non-coding region of 209 residues. The polypeptide coding region of IE110 has a base composition of 75.4% G + C. This is a very high value, even for an HSV gene (see Discussion).

Amino acid sequence of IE110

The 775 amino acid polypeptide predicted by the above reading frame has a molecular weight of 78452. This is considerably lower than the previous estimate of 110000 obtained from gel electrophoretic mobility (Marsden et al., 1976). However, such revisions of molecular weights have occurred frequently in HSV sequence analysis (see, for example, McGeoch et al., 1985). The IE110 sequence contains a small excess of basic over acidic residues (Table 2). The five most common amino acid types are Ala, Pro, Gly, Ser and Arg. Ala, Pro, Gly and Arg are the amino acids which possess codons containing only G and C residues (see Table 1), so that their prominence correlates with the extreme base composition of the gene. The distribution of these residues is not uniform; they are most abundant in a region beginning with the start of exon 3 (amino acid residue 242) and ending with residue 560. In addition, amino acid residues 554 to 594 consist mostly of Ser and Ala residues.

Recently, it has been found that many DNA-binding proteins contain a Cys-rich locality, which is thought to be involved in coordination of metal ions (Miller et al., 1985; Berg, 1986). The most characteristic sequence element within these regions comprises two Cys residues separated by two other amino acids. We have now found that IE110 also belongs to this class: as shown in Fig. 5, nine of the 14 Cys residues in IE110 are located in a 58 residue region between amino acids 99 and 156, and these include three C--C pairs.
CGGGCCGGCCGCTTCCGGGAAATTTCGCGGCCTATCCGGCCTCCGGGCTCCGGCGGGCGTGATACACTCTGCCCTGTGGGTTCGGTTATGCATAATGGGCAACCCCGGTATTCCCCGCCTCCCGCGCCGCGCGTAACCACTCCcCTGGGGTTCCGGGTTATGCTAATT

GCTTTTTTGGCGGAACACACGGCCCCTCGCGCATTGGCCCGCGGGTCGCTCAATGAACCCGCATTGGTCCCCTGGGGTTCCGGGTATGGTAATGAGTTTCTTCGGGAAGGCGGG~GCCC

CGGGGCACCGACGCAGGCCAAGCCCCTGTTGCGTCGGCGGGAGGGGCATGCTAATGGGGTTCTTTGGGGGACACCGGGTTGGGCCCCCA~TCGGGGGCCGGGCCGTGCATGCT~TGAT

ATTCTTTGGGGGCCGGCCTTGGTGTCCCCCGGGGACACGGGCGGCGCGCGCGGTTGCGCCCTCCCGGCCGCGCCTCCGGGACGCGCGCCGGGCATTTGGGGAAGGGGAGGGGAAAAG

GCGTGGGGTATATACTGCCCTGCACGCGATTGGGCACTTGGGGGGCGCCATATTGGGGGGGCCATGTTGGGGGACCCCCGACCCTTACACTGGAACCGGCCGCCATGTTGGGGGACCCCCAC

TCATACAGGGAGCCGCGCCCCGACATTGGGCGCCATGGTATAGGCGGCGCTAGTGGGAGGCGGAGAGCGAGACGCAGCAGCCAGGCAGACTCGGGCCGCCCCCTCTCCGCATCACCACAGAAGC

..... O ..... > 5' Terminus of IEII0 mRNA

MEPRPGASTRRPEGRPQRE 19

CCCGCCTACGTTGCGACCACCCCAACAGGACCCTTCGTCGCCGACCTCCACCCGGCGACATCGACACGGACGCTGGGACACCCGCAGCTGGGTATATGGGTCGAGGAGGGGAGGGGAGGAG

N Terminus of IEII0 End of exon 1 /

GTGAGGGGCCGGGCGCCATGTCTGGGGCGcCATATTGGGGGGCGCCATATTGGGGGGGCCATGTTGGGGGACCCCCGACCCTTACACTGGAACCGGCCGCCATGTTGGGGGACCCCCAC

TCATACAGGGAGCCGCGCCCCGACATTGGGCGCCATGGTATAGGCGGCGCTAGTGGGAGGCGGAGAGCGAGACGCAGCAGCCAGGCAGACTCGGGCCGCCCCCTCTCCGCATCACCACAGAAGC

TATATACAGGGACCGGGGTCGCGCCCTGTGGGGGTGCACTGCACAGCAACACAGAAGGCGGTTGCCGCTGTTGGGCGTGGTCCCCCTCCACCTCTCTCCCGGACAGCGAGCGAGTACCCGCCGGCCTGAGGGCCGCCCCAGCGCGAG

VGGGRGDADHHDDDSASEADSTDTEFLGGLGPQGVDGG

GTGGGGGGGGCCCGGGGGGACGCCGACCACCATGACGACGACTCCGCCTCCGAGGCGCAGCACGGAGCACGGGACAGTGCGGCTGCTGGGGCCGGCCGACGGTCGATGGGGGGCG

NQRFAPRYLTLGGHTVRALSP. THPEPTTDEDDDDLDD 241

~TCAGCGGTTCGCCCCGCGGTACCTGACCCTGGGGGGGCACACGGTGAGGGCCCTGTCGCCCAC~CAC~GGAGCCCACCACGGA~GAGGATGACGACGA~TGGACGACGGTGAGGcG

End of exon 2 /

GGGGGCGGC~GGAC~TGGGGGAGGAGGAGGAGGAGGGGGGGGGAGGGAGGAATAGG~GGGCGGGCGAGGAAAGGGCGGG~cGGGGAGGGGGcGTAACcTGAT~G~GCcCcCCGTTGT~

ADYVPPAPRRTPRAPPRRGAAAPPVTGGASHAAPQPAA279

TcTTGCAGCAGATACGTACcGCcGcCCcCCGGGACGcccCGCGCccCCACGCAGAGGCGCGcGCGCcCCGTGACGGGCGGGGCGTcTACGAGCcCcAGCGGCCGC

End of exon 2 /

GGGGGCGGC~GGAC~TGGGGGAGGAGGAGGAGGAGGGGGGGGGAGGGAGGAATAGG~GGGCGGGCGAGGAAAGGGCGGG~cGGGGAGGGGGcGTAACcTGAT~G~GCcCcCCGTTGT~

ADYVPPAPRRTPRAPPRRGAAAPPVTGGASHAAPQPAA279

TcTTGCAGCAGATACGTACcGCcGcCCcCCGGGACGcccCGCGCccCCACGCAGAGGCGCGcGCGCcCCGTGACGGGCGGGGCGTcTACGAGCcCcAGCGGCCGC

End of exon 2 /

GGGGGCGGC~GGAC~TGGGGGAGGAGGAGGAGGAGGGGGGGGGAGGGAGGAATAGG~GGGCGGGCGAGGAAAGGGCGGG~cGGGGAGGGGGcGTAACcTGAT~G~GCcCcCCGTTGT~

ADYVPPAPRRTPRAPPRRGAAAPPVTGGASHAAPQPAA279

TcTTGCAGCAGATACGTACcGCcGcCCcCCGGGACGcccCGCGCccCCACGCAGAGGCGCGcGCGCcCCGTGACGGGCGGGGCGTcTACGAGCcCcAGCGGCCGC

End of exon 2 /

GGGGGCGGC~GGAC~TGGGGGAGGAGGAGGAGGAGGGGGGGGGAGGGAGGAATAGG~GGGCGGGCGAGGAAAGGGCGGG~cGGGGAGGGGGcGTAACcTGAT~G~GCcCcCCGTTGT~

ADYVPPAPRRTPRAPPRRGAAAPPVTGGASHAAPQPAA279

TcTTGCAGCAGATACGTACcGCcGcCCcCCGGGACGcccCGCGCccCCACGCAGAGGCGCGcGCGCcCCGTGACGGGCGGGGCGTcTACGAGCcCcAGCGGCCGC

End of exon 2 /
Fig. 4. DNA sequence of the IE110 gene and its surroundings. The DNA sequence is shown for residues 1201 to 5520 of IRL, as the leftward 5' to 3' strand only. The mapped termini and intron/exon boundaries of IE110 mRNA are indicated, and the predicted amino acid sequence of IE110 is given in single letter code. Proposed TATA box and AATAAA polyadenylation-associated sequences are underlined. Sets of tandem reiterations are marked as \[ . . . \].
We investigated the possibility that the IE110 protein might be related to the IE175 protein, which is also involved in transactivation of later virus genes, but could not detect any homology, global or local, between the IE110 and IE175 amino acid sequences, beyond noting that both contain a very Ser-rich region (amino acids 554 to 594 of IE110, and 177 to 199 of IE175; McGeoch et al., 1986). Conceivably, these residues could be involved in the phosphorylation of the proteins. In addition, no homology of IE110 sequences was detected with other HSV IE proteins (sequences from Murchie & McGeoch, 1982; McGeoch et al., 1985; L. J. Perry & D. J. McGeoch, unpublished data), or with the major IE protein of the betaherpesvirus, human cytomegalovirus (Stenberg et al., 1984) or with any polypeptides predicted from the complete genome sequence of the gammaherpesvirus, Epstein–Barr virus (Baer et al., 1984). A probable homologue to IE gene 1 was detected in the genome of the related alphaherpesvirus, varicella-zoster virus (VZV). This is gene 61 of VZV, which is proposed to lack introns and to encode a 467 amino acid polypeptide (Davison & Scott, 1986). Similarity between the two amino acid sequences is extremely limited [so that the correspondence was not detected in tests carried out by Davison & Scott (1986)] and is essentially limited to the Cys-rich region, where all three C--C pairs are conserved (Fig. 5). We consider that this limited similarity, taken together with the corresponding genomic locations of the two genes, constitutes a reasonable case for them being related and having similar functions. Confirmation of this assignment, however, must await experimental demonstration of transcriptional activation by the VZV gene 61 product.

Detection of IE110 using anti-oligopeptide serum

An antiserum raised to the C terminal 11 amino acids of the predicted sequence of IE110 immunoprecipitates the native protein from extracts of cells infected with HSV-1 and labelled with [35S]methionine under IE conditions (Fig. 6, lane 2). Specificity was shown by inhibition of the immunoprecipitation by inclusion of an excess of the synthetic oligopeptide in the reaction mix (Fig. 6, lane 6). This confirms our reading frame assignment at the C terminus. In addition, no reaction with extracts of mock-infected cells was observed. Also specifically immunoprecipitated is a lower molecular weight species (apparent mol. wt. of 40000) which may be a breakdown product of IE110 or a species associated with it in the infected cell.

Mutational analysis of the IE110 gene

Recent work from several laboratories has shown that IE110 can activate transcription from HSV promoters in transient transfection assays (Everett, 1984; O’Hare & Hayward, 1985a, b; Quinlan & Knipe, 1985; Gelman & Silverstein, 1985). The promoter of the HSV-1 glycoprotein D gene was found to be slightly activated by co-transfection of plasmids encoding IE175, while inclusion of plasmids expressing both IE175 and IE110 resulted in very substantial activation (Everett, 1984). This observation provides the basis for an assay of the function of IE110 in this experimental system. This assay measures the stimulatory activity of IE110 in the presence of IE175 and thus may not exactly duplicate the effects on stimulation by IE110 alone. The assay was used to determine if the effect of frameshift and deletion mutants within IE gene 1 could be correlated with the mRNA mapping and DNA sequence data presented above. In the experiments reported here, plasmids pJR3 and p110, which contain HSV-1 R_L sequences between the SsrI site at 895 and the Psrl site at 7308, were used as a source of functional IE110...
Structure of the HSV-1 IE110 gene

Fig. 6. Immunoprecipitation of HSV-1 IE110 polypeptide by antiserum directed against the C terminus of IE110. An autoradiograph is shown of 35S-labelled proteins extracted from HSV-1-infected cells and fractionated by PAGE. Lanes 1, 2 and 3 show 12 h, IE and mock-infected whole cell extracts, respectively. The numbers at the left indicate estimated $M_r$ values ($\times 10^3$) for major species in lane 1. Lanes 4, 5 and 6 show material immunoprecipitated from IE extracts, using preimmune rabbit serum, anti-IE110 peptide serum and anti-IE110 peptide serum preincubated with synthetic peptide, respectively. ■, Species specifically precipitated by the anti-peptide serum.

The IE110 gene mutations made are shown in Fig. 7 and the results obtained are summarized in Table 3. Deletion of sequences downstream of the HpaI site at 5501 (plasmid p110del7) resulted in an increase in transactivation. While the reason for the increase is not resolved, the important conclusion for this plasmid is that the functionally complete IE110 gene lies upstream of residue 5501. On the other hand, deletion between the XhoI site at 2939 and the SalI site at 5065 (plasmid p110del4) eliminated IE110 activity. Thus, these transactivation data are consistent with the DNA sequence interpretation. Frameshift mutations, p110del6 and
Fig. 7. Structure of the p110 series plasmids. (a) Parental plasmid, pJR3. HSV-1 sequences are shown as an open box, and restriction sites are indicated with their positions from the sequence listing (Fig. 4). The location of the IE110 gene is shown in (b). (c to k) Structures of plasmids pl1Odel4, pl1Odel6, pl1Odel7, pl1Odel8, pl1Odel10, pl1Odel11, pl1Odel1, pl1Odel14 and p110El IF, respectively.

Table 3. Activation of the gD promoter by plasmids expressing IE175 and IE110

<table>
<thead>
<tr>
<th>Activating plasmids</th>
<th>Activity*</th>
<th>Expts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGX58 + pJR3</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>pGX58 + pBR322</td>
<td>3.7 ± 0.9</td>
<td>10</td>
</tr>
<tr>
<td>pJR3 + pBR322</td>
<td>3.4 ± 0.8</td>
<td>9</td>
</tr>
<tr>
<td>pBR322 (8ug)</td>
<td>2.4 ± 0.6</td>
<td>5</td>
</tr>
<tr>
<td>pGX58 + pl10del1</td>
<td>18.6 ± 4.5</td>
<td>7</td>
</tr>
<tr>
<td>pGX58 + pl10del4</td>
<td>3.0 ± 0.8</td>
<td>3</td>
</tr>
<tr>
<td>pGX58 + pl10del6</td>
<td>7.8 ± 1.8</td>
<td>4</td>
</tr>
<tr>
<td>pGX58 + pl10del7</td>
<td>31.1 ± 5.8</td>
<td>3</td>
</tr>
<tr>
<td>pGX58 + pl10del8</td>
<td>8.0 ± 1.2</td>
<td>4</td>
</tr>
<tr>
<td>pGX58 + pl10del10</td>
<td>53.7 ± 13.0</td>
<td>7</td>
</tr>
<tr>
<td>pGX58 + pl110</td>
<td>167.3 ± 39.9</td>
<td>3</td>
</tr>
<tr>
<td>pGX58 + pl110del1</td>
<td>167.8 ± 46.8</td>
<td>5</td>
</tr>
<tr>
<td>pGX58 + pl110del14</td>
<td>18.7 ± 4.9</td>
<td>6</td>
</tr>
<tr>
<td>p175 + pl11†</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>p175 + pl10E11†</td>
<td>104 ± 25.6</td>
<td>5</td>
</tr>
<tr>
<td>p175 + pl10E11F†</td>
<td>80 ± 1.0</td>
<td>2</td>
</tr>
</tbody>
</table>

* Activities are expressed as a percentage of the value obtained for the parental plasmid, ± the standard error.
† Plasmids p111, pl10E11 and p110E11F were tested in a later series of experiments than the others, and used p175 as a source of IE175.
pl10del8, both mapping in exon 2, gave very low levels of activation. We do not know whether the apparent, small effects with these plasmids are actually due to a residual activity of the N terminal fragments of IE110 expressed. Plasmid pl10del10, containing a frameshift mutation at the SalI site at 5065, near the assigned translational termination site at 5095, gave an intermediate activation. We interpret this as indicating that the immediate C terminal residues are not crucial for IE110 activity. We also constructed two in-frame deletions within predicted coding sequence. Plasmid pl10del has lost 354 bp (118 codons) from exon 2 (including almost all of the Cys-rich region), and pl10del14 has lost 483 bp (181 codons) from exon 3. Rather surprisingly, both retain low levels of transactivating activity. Evidently, the IE110 protein can still function to some extent despite substantial structural alterations in two different regions. Finally, frameshift mutations were constructed at the BamHI site, residue 2508 (pl10del11), and at the HaeIII site at residue 3440 (pl10E11F). These plasmids gave similar activation levels to those obtained with their parent. These results are consistent with the mRNA mapping data, which placed the BamHI site at position 2508 and the HaeIII site at position 3440 within introns 1 and 2 respectively.

DISCUSSION

This paper describes the DNA sequence, mRNA mapping and preliminary mutational analysis of IE gene 1, encoding IE110. In addition, we have demonstrated a specific interaction of anti-C terminal oligopeptide serum with the native protein extracted from HSV-1-infected cells. The HSV genome contains five separate IE genes (Watson et al., 1979; Anderson et al., 1980). Of these, IE genes 2 and 3 possess no introns (Whitton et al., 1983; Rixon et al., 1982), while genes 4 and 5 each contain a single intron in the 5' non-coding region (Watson et al., 1981; Rixon & Clements, 1982; Murchie & McGeoch, 1982). IE gene 1 is the only IE gene with multiple introns, and the only one containing introns in polypeptide-coding sequences. Costa et al. (1985) have described a late gene of HSV-1 containing an intron, which lies within polypeptide-coding sequences. Fontichiaro et al. (1985) have claimed to detect a second intron in IE genes 4 and 5, but we consider that their interpretation is erroneous and results from a secondary structure artefact in high G + C DNA (Rixon & Clements, 1982).

HSV-1 DNA has an estimated overall base composition of 67% G + C (Kieff et al., 1971) and most genes so far sequenced have coding regions around 65% G + C (see McGeoch et al., 1985). The coding region of IE gene 1 is strikingly higher, at 75.4% G + C. This is the second highest value in any determined protein-coding sequence, so far as we are aware, and is exceeded only by that of another HSV-1 gene, the IE gene 3, which is situated in the R5 region of the genome and has a coding region composition of 81.5% (McGeoch et al., 1986). Inspection of the IE gene 1 codon usage catalogue (Table 1) shows that the base composition is achieved partly by encoding high levels of amino acids with G + C-rich codons, and partly by a heavy bias towards G and C in the redundant third positions. The phenomenon has been discussed for IE gene 3 (McGeoch et al., 1986), and the arguments adduced then appear also to apply in the present case. Thus, it is clear in general that the base composition is not driven to extremity by requirements of amino acid sequence functionality. The nature of evolutionary forces underlying the effect remains obscure, although they are evidently potent. For both IE gene 1 and gene 3, it is a major repeat element of the HSV genome that is involved.

In both genes there are, within the coding sequence, regions of particularly high G + C content. In the present study, the central portion of the IE110 coding sequence (residues 3489 to 4564) has a base composition of 80.8%, while the upstream and downstream portions (allowing for introns) are 71.2% and 70.4% respectively. It is interesting that the upstream boundary of this high G + C region is at or near the start of exon 3: this could suggest a possible distinction in evolutionary history or in function of encoded polypeptide between exons 1 plus 2, on the one hand, and exon 3. In the case of IE gene 3, McGeoch et al. (1986) suggested that particularly extreme base composition might correlate with sequences encoding relatively non-crucial regions of protein. The plasmid pl10del14 (which has lost 483 bp from the region of high G + C) retained some transactivational activity, which would support this suggestion;
however, plasmid pl10de1l contains a deletion outside of the high G + C region and also retains some transactivational activity.

There is at present no firm model for the mode of action of IE110. The existence of a Cys-rich locality in IE110, similar to that found in many proteins which interact directly with nucleic acids (Berg, 1986), does suggest a rather direct mechanism of transcriptional activation, involving binding of IE110 to DNA. The finding that a residual activity remains after deletion of this region we regard as probably indicative of the complexity of functional interaction between IE110 and IE175, which is also present in our assay.

In conclusion, the mRNA mapping, DNA sequencing and sequence interpretation give a precise description of IE gene 1. Our studies on modification of the gene correlate well with the sequence analysis, and represent a start in molecular genetic analysis of IE110 function.

We wish to thank J. H. Subak-Sharpe for advice and encouragement during this work. Our thanks are also due to N. D. Stow, A. J. Davison, C. M. Preston and J. Russell for plasmids, and to A. Dolan, M. Wilkie, A. Orr and S. McCallum for technical assistance. L. J. P. was in receipt of a grant from The Shell Trust for Higher Education.

REFERENCES


(Received 23 May 1986)