DNA Sequence and Genetic Content of the HindIII I Region in the Short Unique Component of the Herpes Simplex Virus Type 2 Genome: Identification of the Gene Encoding Glycoprotein G, and Evolutionary Comparisons

By DUNCAN J. McGECH,* HELEN W. M. MOSS, DAVID McNAB AND MARGARET C. FRAME

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

(Accepted 1 October 1986)

SUMMARY

The DNA sequence was determined of the HindIII I fragment of herpes simplex virus type 2 (HSV-2), which is located in the short unique region of the HSV-2 genome. HindIII I was found to comprise 9629 base pairs. Comparison with the previously determined corresponding sequence for herpes simplex virus type 1 (HSV-1), and limited mRNA mapping, showed that HindIII I contained six genes (termed US2 to US7) and part of another (US8). The HSV-1 and HSV-2 sequences were found to be generally colinear, with one major exception: the HSV-2 DNA contained an extra sequence of about 1460 base pairs, in the coding region of gene US4. By use of an antiserum raised against an oligopeptide representing amino acids near the C terminus of the predicted HSV-2 US4 polypeptide it demonstrated that this gene encodes the virion glycoprotein gG-2, while HSV-1 US4 encodes a much smaller virion glycoprotein with homology to the C-terminal portion of gG-2. Quantitative comparisons of the HSV-2 HindIII I and corresponding HSV-1 sequences showed that they had diverged by point mutation and by local addition and deletion, as well as by the major change in genes US4. It was found that within the HSV-2-specific part of gG-2 there was a locality showing sequence similarity to a glycoprotein of pseudorabies virus (gX), and weaker similarity to glycoproteins D of HSV-1 and HSV-2. These data were interpreted to suggest, first, that HSV-2 US4 represents an ancient gene of alphaherpesviruses, and, more tentatively, that the evolution of the genes for gG and gD may have proceeded through a duplication event.

INTRODUCTION

Since the 1960s it has been recognized that strains of herpes simplex virus (HSV) fall into two serotypes, designated type 1 and type 2 (HSV-1 and HSV-2) (Schneweis, 1962; Plummer, 1964; Roizman et al., 1981). These are primarily associated with orofacial and with genital lesions, respectively (Whitley, 1985).

It is known that both type-common and type-specific antigenic sites exist on many HSV-specified proteins (Honess et al., 1974). Biochemical comparisons of virion proteins and of virus-specified proteins in extracts of infected cells have shown that HSV-1 and HSV-2 encode many apparently equivalent proteins, although many minor changes are detectable, for instance in relative mobility on gel electrophoresis (Cassai et al., 1975; Halliburton et al., 1977; Marsden et al., 1978).

Analyses of genome DNA structure also indicated the close relations between the serotypes. In each case the genome DNA is a double-stranded molecule of Mr about 1·0 × 10⁸ (Kieff et al., 1971), and with the same general arrangement of large scale repeat and unique sequences (see Fig. 1). Relatedness between the DNA sequences of the two virus types was measured in
hybridization experiments, which indicated that approximately half of the sequences matched well whereas the remainder had little homology (Kieff et al., 1972). An electron microscopic analysis of hybrid DNA molecules gave a similar result (Kudler et al., 1983). Davison & Wilkie (1983) examined cross-hybridizing sequences in detail along the length of each genome, and concluded that the two genomic sequences were closely collinear. In line with these findings, it has been demonstrated that viable intertypic recombinant viruses can readily be isolated in the laboratory (Timbury & Subak-Sharpe, 1973; Halliburton et al., 1977; Marsden et al., 1978).

In recent years, DNA sequence analysis has allowed precise comparison between some corresponding regions in the two serotypes' genomes. So far, these analyses have been limited in scope to individual genes and parts of genes, and to other, relatively small, functional sequences, including promoters and origins of replication (for instance, Swain & Galloway, 1983; McLauchlan & Clements, 1983; Whitton et al., 1983; Whitton & Clements, 1984a, b; Swain et al., 1985). In general, comparisons have shown corresponding regions to have closely related sequences, with occasional divergences and addition/deletion changes.

At present, the largest published HSV DNA sequence is for the short region of HSV-1, of 26000 base pairs (bp) (McGeoch et al., 1985, 1986a). This comprises the short unique sequence (Us) together with two flanking short repeat sequences (see Fig. 1). HSV-1 Us is almost 13000 bp in length and contains the protein coding parts of 12 genes (McGeoch et al., 1985; Rixon & McGeoch, 1985). It seemed important to determine the sequence of a large part of the corresponding region of the HSV-2 genome in order to investigate relations between the sequences of the two serotypes over a relatively extended portion of their genomes. Of particular interest was the report, based on restriction analyses, that the Us region of HSV-2 was of the order of 1500 to 3000 bp larger than HSV-1 Us (Davison, 1981). Another datum was that the gene for a large glycoprotein, termed gG (or gG-2 to indicate its serotypic origin), was mapped to HSV-2 Us (Marsden et al., 1978, 1984; Roizman et al., 1984; Olofsson et al., 1986), and we could not identify a convincing counterpart to this in the HSV-1 sequence.

We chose to determine the sequence of the restriction fragment HindIII l of HSV-2 strain HG52, which contains most of Us, but does not include sequences at each end of the region (Cortini & Wilkie, 1978). In this paper we report the sequence of HindIII l, and its interpretation in terms of gene organization and relations with HSV-1 Us. HindIII l comprises 9629 bp and contains six genes plus part of a seventh. The gene encoding gG has been identified, and its relation with the HSV-1 sequence clarified. Finally, from these comparisons we have been able to propose evolutionary events which may have occurred in the generation of the present-day serotypes' genomes.

METHODS

DNA sequence analysis. A plasmid consisting of the HindIII l fragment of HSV-2 strain HG52 inserted into the HindIII site of pAT153, obtained from A. J. Davison, was propagated in Escherichia coli DH1. This cloned fragment was used for DNA sequence analysis by the M13/chain terminator method, as described (Sanger et al., 1980; McGeoch et al., 1985, 1986a; Quinn & McGeoch, 1985). In brief, isolated HindIII l DNA was sheared by sonication, and fragments of 200 to 600 bp were cloned into the Smal site of M13mp8. Recombinant bacteriophage from separate plaques were grown, and single-stranded DNA extracted. DNA sequences of inserted fragments were determined, and the whole sequence was assembled with the database program of Staden (1982).

Computing. A DEC PDP11/44 computer running under the RSX11M operating system was used for sequence analysis, interpretation and comparison, as described (McGeoch et al., 1985, 1986a).

The following procedures were used for detection and evaluation of similarities between pairs of sequences (DNA and polypeptide). Initial comparisons were made with the matrix comparison program of Pustell & Kafatos (1982) in which the two sequences to be compared are regarded as the x and y axes of a graph, and the locations of similar sequences are marked on the graph at appropriate coordinates, so that extended similarity is seen as a diagonal line of slope $-1$. The original program scored only for identical residues, and was extended by our colleague P. Taylor, for amino acid sequences, to score for similar amino acids using the replacement probability table of Dayhoff et al. (1983). In addition, output on a graphics printer was implemented. Regions for which similarity was detected using the matrix program were then examined using the alignment optimizing program of Taylor (1984). Starting with assigned and separate penalties for introduction of mismatches and of gaps in one or other sequence, this program produces an optimal alignment of two sequences.
DNA sequence of HSV-2 HindIII l

Fig. 1. Location of HindIII l in the HSV-2 genome and organization of genes in HindIII l. The upper part of the figure shows a conventional representation of the genome of HSV-2. The long unique (UL) and the short unique (US) sequences are shown as solid lines, and the long repeat (TRL and IRs) and the short repeat (IRs and TRs) sequences as open boxes. The location of HindIII l in US is indicated, and the lower part of the figure expands the 9629 base pair HindIII l region, with numbering corresponding to Fig. 2. Proposed locations and orientations of transcripts are shown as arrows, with predicted protein-coding regions as open boxes. Numerals represent genes US2 to US8, as described in the text. At the bottom are indicated the published mapping limits for sequences encoding epitopes of glycoprotein G (Olofsson et al., 1986).

Output from the optimal alignment program was further used in several ways. First, ‘overall divergence’ between two sequences (DNA or polypeptide) was defined as the number of mismatches plus gapping characters, expressed as a percentage of the total length of the aligned sequences after introduction of gaps. Second, ‘mismatch divergence’ was defined as the number of mismatches between two aligned sequences, excluding gaps, expressed as a percentage of the aligned sequence length excluding regions of introduced gaps. Third, ‘length divergence’ was defined as the sum of introduced gapping characters in both sequences, expressed as a percentage of the total length including gaps. In initial descriptions of HSV-1 and HSV-2 amino acid sequences, percentage homologies are quoted: these are equivalent to (100 — overall divergence).

Northern blot analysis of HSV-2 mRNA. Confluent monolayers of BHK cells were infected with 10 p.f.u./cell of HSV-2 strain HG52. After 4 or 7 h at 37 °C, cells were collected and RNA was extracted as described (Kumar & Lindberg, 1972). Poly(A)-containing RNA was selected by chromatography on oligo(dT)-cellulose, and used for Northern blot analysis, essentially as described (Rixon & McGeoch, 1985). In the example shown in Fig. 4, the DNA probe was a 38-residue oligodeoxynucleotide which had been chemically synthesized and 5’-labelled by means of [y-32P]ATP and polynucleotide kinase.

Preparation of anti-oligopeptide serum and immunoprecipitation. The dodecapeptide NH2-Pro-Ser-Val-Arg-Tyr-Val-Cys-Leu-Pro-Pro-Glu-Arg-COOH (synthesized for us by Cambridge Research Biochemicals) was coupled to bovine serum albumin using water-soluble carbodiimide and used to raise an antiserum in rabbits as described (Frame et al., 1986).

Confluent monolayers of BHK cells were infected with 20 p.f.u./cell of HSV-1 strain 17 or HSV-2 strain HG52, and labelled with 100 µCi/ml [3H]mannose from 5 to 11 h post-infection. Immunoprecipitations were carried out as described (Frame et al., 1986) and precipitated proteins were analysed by electrophoresis in 5 to 12.5% gradient polyacrylamide gels containing sodium dodecyl sulphate (Marsden et al., 1978).

RESULTS

DNA sequence analysis of the HSV-2 HindIII l fragment

The location of the restriction fragment HindIII l within the US region of HSV-2 DNA is indicated in Fig. 1. The sequence of a plasmid-cloned copy of HSV-2 HindIII l was determined
The 24 gene US4 region. These comparisons enabled us to align restriction maps for the two serotypes. (McGeoch had previously been analysed by complete sequence determination and mRNA mapping terms of our sequence analysis.

Reported size difference in Us between the two serotypes. Kudler heteroduplexes of recombinants. Clearly, the 1460 residue extra sequence in HSV-2 Us accounts for the previously appropriate in comparisons of the two genomes, for example in analysis of intertypic 2940 to 4400, which had no counterpart in HSV-1. This discontinuity is located in the HSV-1 this: the HSV-2 sequence contained a section of about 1460 residues, approximately numbers HindlI1,

The base composition was 65.9 ~ G + C. This is the largest sequence so far obtained for HSV-2. Proposed amino acid sequences of encoded proteins are given in single-letter code. For the leftward oriented gene US2, the amino acid sequence is below the corresponding DNA sequence, while for the other, rightward genes the amino acid sequences are above the DNA sequence. Locations of genes are given at the left, on the first line containing amino acid sequence for that gene, irrespective of gene orientation. Proposed TATA boxes and polyadenylation-associated sequences AATAAA are underlined.

Fig. 2. The DNA sequence of HindIII l. The sequence is indicated as the rightward 5' to 3' strand only. Proposed amino acid sequences of encoded proteins are given in single-letter code. For the leftward oriented gene US2, the amino acid sequence is below the corresponding DNA sequence, while for the other, rightward genes the amino acid sequences are above the DNA sequence. Locations of genes are given at the left, on the first line containing amino acid sequence for that gene, irrespective of gene orientation. Proposed TATA boxes and polyadenylation-associated sequences AATAAA are underlined.

by the M13 shotgun method. The final sequence was supported by a database of 95000 sequence characters, and comprised 9629 residues (including the whole of both terminal HindIII sites). The base composition was 65.9% G + C. This is the largest sequence so far obtained for HSV-2. The HindIII l sequence is listed in Fig. 2.

Interpretation of the sequence was primarily by comparison with the HSV-1 Us region, which had previously been analysed by complete sequence determination and mRNA mapping (McGeoch et al., 1985; Rixon & McGeoch, 1985). Computerized comparison of DNA sequences in HSV-2 HindIII l and HSV-1 Us showed that there was high homology over much of HindIII l, indicating a generally colinear genetic organization. There was one major exception to this: the HSV-2 sequence contained a section of about 1460 residues, approximately numbers 2940 to 4400, which had no counterpart in HSV-1. This discontinuity is located in the HSV-1 gene US4 region. These comparisons enabled us to align restriction maps for the two serotypes. The alignment is shown in Fig. 3 to illustrate gross relations between the sequences, and is appropriate in comparisons of the two genomes, for example in analysis of intertypic recombinants. Clearly, the 1460 residue extra sequence in HSV-2 Us accounts for the previously reported size difference in Us between the two serotypes. Kudler et al. (1983) analysed heteroduplexes of HindIII l with HSV-1 DNA, but we have been unable to interpret their data in terms of our sequence analysis.
Our present view of gene organization in HSV-1 Us is based on evaluation of DNA sequence and mRNA mapping information. Any such interpretation of a large DNA sequence inevitably leaves some local qualifications and uncertainties. In interpreting the HSV-2 data, we used homology with HSV-1 to provide a framework. However, the comparison also provided a new means for critical examination of the current view of a large part of HSV-1 Us. This reappraisal did not uncover any basic conflict between the two sets of data. Thus, the HSV-2 sequence possessed open reading frames and control sequence elements where they were expected, and the natures of predicted amino acid sequences were consistent with the HSV-1 results. On the other hand, various sequences which in HSV-1 had been assigned a non-critical role were less conserved or not conserved.

Our interpretation of the HSV-2 HindIII l sequence in terms of the predicted locations of mRNAs and protein coding regions is summarized in Fig. 1 and explicitly listed in Fig. 2. We consider that HSV-1 genes US2, US3, US5, US6 and US7 all have close counterparts in HSV-2 HindIII l, that HSV-1 US4 has a less closely related counterpart, and that the right extremity of HindIII l contains the upstream half of a gene similar to HSV-1 US8. The HSV-2 genes were therefore named after the corresponding HSV-1 genes. We now describe regions of the sequence.

HSV-2 genes US2 and US3

The left end of HindIII l is situated in the region between genes US1 and US2, so that all of the leftward transcribed US2 is within HindIII l. The sequences in the proposed promoter region of HSV-2 US2 are highly similar to the HSV-1 version, so we propose that HSV-2 transcription is initiated at residue 1371, downstream of the candidate TATA box sequence at 1400 to 1394. We propose that translation initiates with the first ATG at 1003 to 1001, and terminates with TAA at 130. As with HSV-1, this gives a long 5' non-coding region. The 291 amino acid protein so encoded has an M_r of 32778, and is 76% homologous to the HSV-1 protein. Both proteins have an N-terminal hydrophobic region, which might be a signal sequence for translation on membrane-bound ribosomes (McGeoch, 1985). Their function is unknown.
Gene US3 of HSV-1 is transcribed rightwards from two separate promoters. The resulting mRNAs, termed US3a and US3b, lead into the same open reading frame, and the 5'-terminal regions of both overlap substantially, on the opposite strand, with the 5'-terminal sequence of US2 mRNA. This model appears to hold also for HSV-2 insofar as the promoter regions are well conserved. A qualification is that HSV-2 US3a mRNA, thought to initiate at 1039, would contain a possible translational initiation codon at 1059, which has no counterpart in HSV-1. This starts a 31 codon open reading frame, terminating at 1152. For the present we discount the possibility that this might be a genuine translated sequence. The next translation initiator in US3a, and the first in US3b, is at 1290. This opens a 481 codon open reading frame encoding a protein of $M_r$ 52674, which is the proposed US3 gene product, and which shows 76% identity with the HSV-1 counterpart. These polypeptides are homologous to the protein kinase-related family of eukaryotes and retroviruses (McGeoch & Davison, 1986), but have not yet been identified in infected cells.

The gene US4 region in HSV-2

In HSV-1, gene US4 lies to the right of US3 and is transcribed rightwards. US4 mRNA forms a 3'-coterminal set with the US3 mRNAs. HSV-1 US4 mRNA contains 804 residues and is predicted to encode a protein of $M_r$ 25237, referred to as 25K. Recently we have shown, using antiserum to an oligopeptide from this predicted amino acid sequence, that US4 encodes a previously unrecognized virion glycoprotein (Frame et al., 1986), as was considered likely from the amino acid sequence (McGeoch, 1985). The same glycoprotein has also been characterized by Richman et al. (1986), using a monoclonal antibody. In HSV-2, DNA sequences clearly equivalent to the US4 promoter, the 5'-terminus of the mRNA, and the 3'-terminal portion of the mRNA are present. However, relative to HSV-1, an additional 1460 nucleotide section is present between the mRNA 5'- and 3'-terminal sequences. It seemed most likely that the HSV-2 gene organization was similar to that of HSV-1, except that the HSV-2 US4 gene was much larger. Experiments were carried out to clarify the location and structure of transcripts for this region.

Early (4 h post-infection) and late (7 h) preparations of poly(A)-containing RNA from HSV-2-infected cells were separated by electrophoresis in an agarose gel containing formaldehyde, and transferred to a nitrocellulose membrane. This was probed with a $[5'-32P]$oligodeoxynucleotide complementary to a sequence near the 3' terminus of the putative gene US4 (residues 4828 to 4865 of Fig. 2). As shown in Fig. 4, two mRNA species were clearly detected, of estimated sizes 3700 and 2300 nucleotides [including poly(A) tracts]. The smaller, which was relatively much more abundant in the late sample, is presumed to be the counterpart of HSV-1 US4 mRNA. Analyses of protection of end-labelled DNA probes from S1 nuclease by formation of hybrids with mRNA preparations showed that an mRNA 5' terminus was present near residue 2800 and there were no interruptions before the $XhoI$ site at residue 4785 (data not shown). It is therefore proposed that US4 mRNA initiates close to residue 2810 (by comparison with HSV-1) and terminates downstream of AATAAA at 5014. The larger RNA species was present in abundance at early and late times, and is of an appropriate size for US3 mRNA initiating at residue 1039 or 1256 and 3'-coterminal with US4 mRNA.

Within the assigned US4 mRNA region, the first potential initiation codon is at residue 2842, and this opens a reading frame of 699 codons, closed by TAG at residue 4939. We propose that this encodes the US4 protein, which would have an $M_r$ of 72239 when unprocessed, and is referred to for the moment as 72K. The 72K amino acid sequence is in part homologous to the sequence of the HSV-1 US4 gene product, 25K, as summarized in Fig. 5(a). For purposes of comparison, each polypeptide can be regarded as comprising three parts. First, there is a limited similarity at the N termini, covering approximately the first 20 amino acids. This is proposed in each protein to contain a signal sequence (see below) and so could represent similarity deriving from a common evolutionary origin (homology) or from common function (analogy). Next, there are regions which show no similarity between the two sequences. In HSV-1 25K this comprises about amino acids 25 to 66. In 72K this second region includes the amino acids encoded by the extra DNA sequence of HSV-2, and is thus much longer, running to about residue 546.
Fig. 4. US3 and US4 mRNAs of HSV-2. Polyadenylated mRNAs were purified from HSV-2-infected cells 4 and 7 h after infection, fractionated by electrophoresis in agarose gel, and transferred to a nitrocellulose membrane. A $^{32}$P-labelled oligodeoxynucleotide complementary to a region in gene US4 was annealed to the mRNAs. The figure shows an autoradiograph of the mRNA species detected. Locations of marker ribosomal RNAs are indicated at the right. The upper and lower mRNA species have estimated chain lengths of 3700 and 2300 nucleotides, and are considered to represent transcripts of genes US3 and US4, respectively.

the C-terminal portions of the sequences are clearly homologous, as shown in Fig. 5(b). In effect, the HSV-2 protein is broadly equivalent to the HSV-1 protein with a large insert near the N terminus, but because the borders of the extra sequence are located in regions without detectable similarity between the serotypes it is not possible to delineate insertion or deletion sites with complete precision.

Both proteins have uncharged N-terminal regions which, by criteria of length and hydrophobicity, could be signal sequences for translation on membrane-bound ribosomes (McGeoch, 1985). Both also have sequences within their homologous C-terminal portions which could be transmembrane anchor regions, as previously noted for HSV-1 25K (McGeoch et al., 1985; see Fig. 5b). Like the HSV-1 25K protein, then, HSV-2 72K has features characteristic of a membrane-inserted species. Both sequences possess potential $N$-glycosylation sites (N–S or
Fig. 5. Comparisons of the amino acid sequences of proteins encoded by the US4 genes of HSV-1 and HSV-2. (a) Relations between the HSV-1 25K and HSV-2 72K proteins, as presented by the matrix comparison program. The sequences were compared using a window of 25 residues, using the replacement probability table of Dayhoff et al. (1983) to score for occurrence of similar amino acids. Similarity scores of 36 and higher are shown as diagonal lines. (b) Alignment of the sequences, first at the N termini and, secondly, in a larger region extending to the C termini. In this latter part the alignments were computed with the alignment optimizing program. Pairs of identical residues are marked by asterisks. The locations of proposed transmembrane sequences are shown by over- or underlining, and locations of sequences used in making antipeptide antisera are marked with colons.
DNA sequence of HSV-2 HindIII 1

HSV-2 genes US5, US6, US7 and US8

In our analysis of HSV-1 Us, we described a low abundance, rightward mRNA with its 5' terminus to the right of gene US4 (Rixon & McGeoch, 1985). Identification of the coding region of this region, termed US5, was complicated by the occurrence of potential translation initiation codons close to the assigned locus of the mRNA's 5' terminus. We decided that the region most likely encoded a previously unsuspected, small protein of 92 amino acids, with characteristics of a membrane-inserted glycoprotein (McGeoch et al., 1985; McGeoch, 1985). We now find that the HSV-2 sequence supports this interpretation, and think that HSV-2 encodes a similar, small protein, with 92 amino acids and M, 9509.

In HSV-1, genes US5, US6 and US7 form a rightward transcribed 3'-coterminal family, and we think that this is the case also for HSV-2. Gene US6 encodes the virion glycoprotein gD. Watson (1983) has reported the sequence of the gD gene region in HSV-2 strain G. This corresponds to residues 5706 to 7348 of our HSV-2 strain HG52 sequences, and the two sequences agree closely. We think that the gD coding region of HSV-2 gene US6 runs from ATG at 5980 to TAG at 7159, specifying a protein of 393 amino acids and M, 43188. Amino acid homology with HSV-1 gD is 82%. Within the coding region, Watson's strain G sequence is identical except for two nucleotide substitutions, one of which specifies a distinct amino acid, at residue 353. This position is within the predicted transmembrane sequence and in strain G comprises an alanine instead of a valine.

HSV-1 gene US7 was proposed to encode a presently unrecognized glycoprotein. We think that the coding region of HSV-2 gene US7 runs from ATG at 7363 to TAG at 8479, encoding a protein of 372 amino acids and M, 39553. This shows 66% homology to the HSV-1 protein and is 18 amino acids shorter. Most of the length difference and sequence variation are in residues 203 to 250 of the HSV-2 species. This is a region thought to be in the external domain of the protein, but close to the C-proximal transmembrane sequence. In HSV-1 there is a partially reiterated amino acid sequence in this locality not found in HSV-2.

The last gene represented in HindIII 1 is for the virion glycoprotein species gE. This is also transcribed rightwards. Translation is thought to start with ATG at 8807, to give 274 codons before the right HindIII site. Thus, only the N-terminal portion of HSV-2 gE, about half of the whole protein, is encoded within HindIII 1. This portion shows 66% homology to the corresponding HSV-1 polypeptide sequence.

Analysis of US4 gene products by a type-common oligopeptide antiserum

As shown in Fig. 1, the mapping bracket for sequences encoding epitopes of HSV-2 gG includes genes US4, US5, US6 and US7. All of these are thought to specify membrane-inserted glycoproteins, but only US6, encoding glycoprotein D, is assigned to a well-characterized species. Because of its large size, HSV-2 gene US4 was the primary candidate for encoding gG. We studied this by preparing an antiserum against an oligopeptide representing amino acids 687 to 698 of the US4 product (that is, a near C-terminal sequence but excluding the C-terminal residue itself). This antiserum (termed 14713) was used to precipitate proteins from extracts of HSV-2-infected, [3H]mannose-labelled cells, and precipitates were analysed by polyacrylamide gel electrophoresis. The results are shown in Fig. 6. A single labelled species was precipitated specifically by the antiserum, and this had a mobility identical to an authentic gG band as precipitated by monoclonal antibody LP5 (Marsden et al., 1984). Specificity was judged by the inhibition of gG precipitation on addition of the synthetic peptide to the antipeptide serum reaction mix. From this experiment we conclude, first, that our reading frame assignment for gene US4 at the C terminus is validated, and, second, that HSV-2 gene US4 does encode gG.

The dodecapeptide sequence used to raise the antiserum is identical, except at one position, to the corresponding HSV-1 US4 protein (see Fig. 5b), so that it was (by intent) possibly type-common. We previously examined the product of HSV-1 gene US4 using antiserum to an internal peptide with an HSV-1 specific sequence (Frame et al., 1986; see Fig. 5b), which precipitated three mannose-labelled species from HSV-1-infected cell extracts and from virion preparations. Fig. 7 shows that the antiserum to the C-terminal region of the HSV-2 US4 protein specifically precipitates the same three species from HSV-1-infected cell extracts. Again,
Fig. 6. Immunoprecipitation of HSV-2 glycoprotein G by antipeptide antiserum and monoclonal antibody. Immunoprecipitations were carried out on extracts of [3H]mannose-labelled, HSV-2-infected BHK cells, and precipitated species were resolved by gel electrophoresis and detected by autoradiography. Lanes 1 and 8 show extracts from mock-infected cells. Lanes 2 and 7 show extracts from HSV-2-infected cells. Lane 3 shows gG-2, as precipitated by monoclonal antibody LP5 (Marsden et al., 1984). Lane 4 is a control using normal ascitic fluid. Lane 5 shows species precipitated with a serum (14713) raised to the C-terminal oligopeptide of 72K. Lane 6 is a control using antiserum 14713 with the addition of 50 μg of the C-terminal peptide. The location of gG-2 in lane 3 and of the equivalent band in lane 5 are marked ■.

Precipitation of these species was inhibited by addition of peptide. Thus, the new antiserum is type-common, and this result confirms the relation of the HSV-1 and HSV-2 US4 products. It also indicates that the three polypeptide species found for HSV-1 US4 are very probably all forms of the same polypeptide. We have thus identified the HSV-1 US4 protein as the HSV-1 counterpart of gG-2, and this is now designated gG-1.

Serafini-Cessi et al. (1985) analysed the carbohydrate content of gG-2 and found a very high level of O-linked oligosaccharide chains, there being approximately 25 times as many O-linked as N-linked chains. We find four potential N-glycosylation sites in our amino acid sequence, so
that, if all of these were fully substituted, there could be about 100 O-glycosylation sites. This would be a maximum value and is clearly approximate. O-linkage is through serine and threonine residues, and the predicted amino acid sequence contains 106 serines plus threonines. In particular, the region between residues 347 and 537 is very rich in these amino acids, containing 61 out of 191 total. We suggest that this region could be the main site of O-glycosylation, with most or all of the serine and threonine side chains being substituted.

**Analysis of the divergent evolution of HSV-1 and HSV-2 DNAs**

The sequences of HSV-2 HindIII I and its HSV-1 counterpart provide the largest continuous segments of the two genomes available for comparison. In this section some quantitative aspects of relations between the two sequences are presented, and interpreted in terms of evolution of the present sequences.

Each DNA sequence was subdivided into its constituent coding and non-coding regions, and corresponding parts were compared by the alignment optimizing program. Fig. 8 shows as an example the result for the non-coding DNA between US5 and US6 coding regions. Tables 1 and 2 summarize results for coding and non-coding regions, respectively.

We used two distinct aspects of the alignments. The first is the extent to which two sequences contain non-identical residues at equivalent positions, and the second is the proportion of 'gapping' characters which were introduced to obtain the alignment. It should be realized that in principle these are not completely independent, inasmuch as the aligning algorithm is provided with separately variable penalties to impose for introduction of gaps and of mismatches (Taylor,
HSV-1 5403 TAAATTCTACCCACACACCGAAAAAAAAGAATCTTACCCTGACACAGGTATGGCTTCAG AATAAA
HSV-2 5527 TAAATTCTACCCACACACCGAAAAAAAAGAATCTTACCCTGACACAGGTATGGCTTCAG AATAAA

Fig. 8. Comparisons of HSV-1 and HSV-2 DNA sequences between the coding regions of genes US5 and US6. The two DNA sequences shown run from the US5 termination codons to end before the US6 initiation codons, and were aligned with the alignment optimizing program.

Table 1. Comparisons of DNA sequences in coding regions

<table>
<thead>
<tr>
<th>Region</th>
<th>HSV-1 (bp)</th>
<th>HSV-2 (bp)</th>
<th>Aligned* (bp)</th>
<th>Overall divergence:</th>
<th>Length divergence:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no. of residues</td>
<td>no. of residues</td>
<td>mismatch divergence (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US2</td>
<td>873</td>
<td>873</td>
<td>880</td>
<td>206 (23.4%)</td>
<td>14 (1.6%)</td>
</tr>
<tr>
<td>US3</td>
<td>1443</td>
<td>1443</td>
<td>1443</td>
<td>283 (19.6%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>US5</td>
<td>276</td>
<td>276</td>
<td>281</td>
<td>97 (34.5%)</td>
<td>10 (3.6%)</td>
</tr>
<tr>
<td>US6</td>
<td>1182</td>
<td>1179</td>
<td>1182</td>
<td>196 (16.6%)</td>
<td>3 (0.3%)</td>
</tr>
<tr>
<td>US7</td>
<td>1170</td>
<td>1116</td>
<td>1174</td>
<td>321 (27.3%)</td>
<td>62 (5.3%)</td>
</tr>
<tr>
<td>US8 (part)</td>
<td>838</td>
<td>823</td>
<td>844</td>
<td>212 (25.1%)</td>
<td>27 (3.2%)</td>
</tr>
</tbody>
</table>

* Aligned lengths of sequences after introduction of gaps.

1984). Obtaining an appropriate alignment is thus a matter of judgement, and there is in general no such entity as an absolute optimal alignment. However, it is clear from inspection of the HSV-1/HSV-2 comparisons that the results obtained are in practice appropriate and meaningful. There is a further aspect of the alignments which could also be used to evaluate relations. This is the number of separate gaps introduced in a given length of sequence to obtain the alignment, which could be used to give a measure of the number of separate addition/deletion events during divergent evolution. However, we have chosen not to pursue this topic explicitly in the present paper.

Table 1 presents the results for protein coding regions, excluding the very distinct US4 genes. Overall divergences are given as the number of non-identical aligned residues, expressed as a percentage of the total length of each sequence after introduction of gaps for alignment. These range from 16.6% to 34.5%, with a total value of 22.7%. There is a low level of length divergence between pairs of coding sequences, giving introduced gapping characters at 0.0% to 5.3% of overall aligned lengths. The highest value is for the US7 genes, and results in part from a reiterated sequence in HSV-1. The weighted overall average is 2.0%. Another measure of divergence is to present the number of non-identical, aligned residues as a percentage of sequence length excluding all sections where gaps have been introduced. This attempts to measure sequence divergence by point mutation as opposed to addition/deletion changes, and is here termed mismatch divergence. Values for this measure range from 16.4% to 32.1%, with a total value of 21.1%. The similarity of these figures to the previous overall divergences reflects, of course, the low level of gaps introduced.

In Table 2 the corresponding data are presented for the seven regions bounding the protein coding sequences. These non-coding sequences fall into four classes: (i) intergenic sequence for genes in a tail-to-tail arrangement (the partial intergenic sequence for US1/US2), (ii) for head-to-head arrangement (US2/US3 sequence), (iii) for tail-to-head, with a polyadenylation signal (US4/US5 and US7/US8), and (iv) tail-to-head, lacking a polyadenylation sequence (US3/US4, US5/US6 and US6/US7). The regions must therefore contain a variety of functional sequences,
Table 2. Comparisons of DNA sequences in non-coding regions

<table>
<thead>
<tr>
<th>Region</th>
<th>Length (bp)</th>
<th>Overall divergence:</th>
<th>Length divergence:</th>
<th>Mismatch divergence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HSV-1</td>
<td>HSV-2</td>
<td>Aligned*</td>
<td>no. of residues</td>
</tr>
<tr>
<td>US1/US2 (part)</td>
<td>97</td>
<td>130</td>
<td>135</td>
<td>62 (45.9%)</td>
</tr>
<tr>
<td>US2/US3</td>
<td>293</td>
<td>285</td>
<td>312</td>
<td>117 (37.5%)</td>
</tr>
<tr>
<td>US3/US4</td>
<td>79</td>
<td>110</td>
<td>110</td>
<td>48 (43.6%)</td>
</tr>
<tr>
<td>US4/US5</td>
<td>273</td>
<td>312</td>
<td>332</td>
<td>150 (45.2%)</td>
</tr>
<tr>
<td>US5/US6</td>
<td>412</td>
<td>453</td>
<td>482</td>
<td>224 (46.5%)</td>
</tr>
<tr>
<td>US6/US7</td>
<td>184</td>
<td>204</td>
<td>210</td>
<td>76 (36.2%)</td>
</tr>
<tr>
<td>US7/US8</td>
<td>288</td>
<td>328</td>
<td>333</td>
<td>144 (43.2%)</td>
</tr>
</tbody>
</table>

* Aligned lengths of sequences after introduction of gaps.

Table 3. Comparisons of predicted amino acid sequences

<table>
<thead>
<tr>
<th>Region</th>
<th>Length</th>
<th>Overall divergence:</th>
<th>Length divergence:</th>
<th>Amino acid mismatch divergence (%)</th>
<th>DNA mismatch divergence (%)†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HSV-1</td>
<td>HSV-2</td>
<td>Aligned*</td>
<td>no. of residues</td>
<td>no. of characters</td>
</tr>
<tr>
<td>US2</td>
<td>291</td>
<td>291</td>
<td>293</td>
<td>71 (24.2%)</td>
<td>4 (1.4%)</td>
</tr>
<tr>
<td>US3</td>
<td>481</td>
<td>481</td>
<td>481</td>
<td>122 (25.4%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>US5</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>52 (56.5%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>US6</td>
<td>394</td>
<td>393</td>
<td>394</td>
<td>71 (18.0%)</td>
<td>1 (0.3%)</td>
</tr>
<tr>
<td>US7</td>
<td>390</td>
<td>372</td>
<td>391</td>
<td>132 (33.8%)</td>
<td>20 (5.1%)</td>
</tr>
<tr>
<td>US8 (part)</td>
<td>279</td>
<td>274</td>
<td>279</td>
<td>96 (34.4%)</td>
<td>5 (1.8%)</td>
</tr>
</tbody>
</table>

* Aligned lengths of sequences after introduction of gaps.
† From Table 1.

although for our present purpose they are treated as one class. The non-coding regions exhibit overall divergences of 36.2% to 53.8% (total value 42.9%), and are thus substantially more diverged than the coding sequences. This is seen as mismatch divergence (20.9% to 33.2%; total 29.4%) and more particularly as length divergence (9.6% to 31.9%; total 19.0%).

Thus, in considering the divergent evolution of the HSV-1 and HSV-2 DNA sequences (and putting aside for the moment the case of the US4 genes) it is seen that both coding and non-coding sequences have undergone both point mutation and addition/deletion changes. Overall, non-coding sequences are more mutable than coding sequences: the accumulation of point mutations in a given length of sequence is about 1.4 times greater, and the length divergence about 9.5 times greater than in coding sequence.

One aspect of the size variability in non-coding regions concerns the occurrence of homopolymer tracts. In these regions of the HindIII l sequence, homopolymer runs of five or more residues represent 13.1% of the DNA, as opposed to 7.2% in coding DNA (including the gene US4). Homopolymer tracts appear to be particularly susceptible to length variation, as judged by comparisons in three situations. First, differences can be seen between lengths of corresponding tracts in DNAs of the two serotypes. Several examples of this are visible in Fig. 8. Second, variations are seen between two strains of one serotype: there is a 19 nucleotide G tract at residues 5708 to 5726 (Fig. 2 and 8) of the HSV-2 HG52 sequence, whereas in the corresponding sequence of strain G there is a G tract of 13 nucleotides (Watson, 1983). Lastly, similar variations have been observed between independent clones of the same DNA sequences from a single strain of HSV-1 (McGeoch et al., 1985; L. J. Perry & D. J. McGeoch, unpublished data).

We next address the subject of differences in the amino acid sequences encoded by the HSV-1 and HSV-2 DNA sequences, and of how these are related to the DNA sequence divergences. Table 3 presents data obtained from alignment of corresponding amino acid sequences, after the manner of Table 1. Gene US4 is again excluded. We emphasize that the DNA and amino acid sequence alignments were evaluated independently and need not represent corresponding alignments in any given case. The extents of mismatch divergence range from 17.8% to 56.5%.
Table 4. Comparisons of DNA sequences in the 1st, 2nd and 3rd codon positions of genes US3 and US6

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. of codons</th>
<th>US3</th>
<th>US6</th>
<th>US3 + US6</th>
</tr>
</thead>
<tbody>
<tr>
<td>US3</td>
<td>481</td>
<td></td>
<td></td>
<td>874</td>
</tr>
<tr>
<td>US6</td>
<td>393</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Position</th>
<th>No. of changes</th>
<th>%</th>
<th>No. of changes</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79</td>
<td>16.4</td>
<td>49</td>
<td>12.5</td>
<td>14.6</td>
</tr>
<tr>
<td>2</td>
<td>54</td>
<td>11.2</td>
<td>31</td>
<td>7.9</td>
<td>9.7</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>31.2</td>
<td>114</td>
<td>29.0</td>
<td>30.2</td>
</tr>
</tbody>
</table>

with a total of 27.1%. The high value is for gene US5, and correlates at least in part with the fact that the US5 amino acid sequences are particularly short (92 residues).

Comparison of the amino acid and DNA sequence mismatch divergences shows that in each case the percentage difference in amino acid sequences is greater than the difference for DNA sequences. On the surface, this result is unexpected: one might have expected that DNA changes would be largely silent, through confinement to the redundant third position of codons. This property of coding sequences was analysed further by computing divergences independently for first, second and third positions in codons. To avoid possible errors through inappropriate alignment of the DNA sequences where introduction of gaps was required, the analysis was limited to genes US3 (no gaps introduced) and US6 (gap of one codon only). Results are presented in Table 4. In both genes the greatest divergences have occurred in codon position 3, while position 2 is most conserved. Thus, as expected, there is least resistance to fixation of mutations in the redundant third position; mutations accumulate in this position at a rate close to that for substitution mutations in the total non-coding DNA sequences. The pattern of DNA divergences in codon positions 1 and 2 is consistent with a model in which each protein is, overall, not highly resistant to the introduction of amino acid changes, but in which replacement at a given site by a similar amino acid is most likely to be tolerated.

Homologies of HSV-2 US4 with other herpesvirus genes

We searched for amino acid sequence homologies between gG-2 and other herpesvirus glycoproteins, using the matrix comparison program, and detected two cases of limited similarity. The first of these was with a section of pseudorabies virus (PRV) glycoprotein, termed gX, whose gene lies in the US region of the PRV genome (Rea et al., 1985). The similarity is restricted to a region of 81 amino acids, residues 92 to 172 of gG-2, as shown in Fig. 9. Within this region the sequences exhibit notable identities, especially in the positions of three cysteine residues and in the occurrence of an eight-residue identical sequence. The similar sequences lie in approximately equivalent positions with respect to the N termini of the two polypeptide chains. The PRV polypeptide contains 498 amino acids in its unprocessed form, and like gG...
DNA sequence of HSV-2 HindIII

possesses an N-terminal signal sequence and a C-proximal transmembrane sequence (Rea et al., 1985). This similarity is in the part of gG-2 with no counterpart in gG-1.

The other similarity detected was between gG-2 and gD of either HSV-1 or HSV-2. This correspondence occupies approximately the same location in the gG-2 sequence as the PRV example, namely residues 79 to about 147 (shown in Fig. 9 for gD-2 only). It is, however, weaker than the alignment of gG-2 and PRV gX. Again, the similar sequences are in near equivalent positions with respect to the N termini. When gD-2 and PRV gX were compared, similarity was also detected (Fig. 9) in this region and not elsewhere. The sets of identical residues seen with each of the three comparisons intersect, but each set possesses some unique identities.

We interpret the strongest correspondence, between gG-2 and PRV gX, as indicating that these genes have a common evolutionary origin in whole or part. More tentatively, we suggest that the correspondence between gG-2 and gD-2 (plus gD-1) may indicate an evolutionary relation between at least parts of these HSV glycoprotein genes.

DISCUSSION

In determining the sequence for HSV-2 HindIII, we set out to examine relations between the genomes of the two HSV serotypes in a relatively substantial section of their genomes, in terms of gene organization and evolutionary change. The sequence showed clearly that the gene arrangements in most of the HSV-1 and HSV-2 Us regions were closely comparable, the only large scale exception to this being the presence of a single, large block of additional sequence in HSV-2 gene US4. Our results are thus qualitatively in accord with the hybridization comparison of the whole genomes by Kieff et al. (1972), which suggested that the DNAs of the two serotypes were partly closely related, and partly quite distinct. Our sequence analysis did not examine regions at the extremities of HSV-2 Us. These have been partly sequenced by Whitton & Clements (1984a). It is clear that at the left end of HSV-2 Us there exists a counterpart of HSV-1 gene US1, and that this accounts for all of this region of Us. At the right end of HSV-2 Us, Whitton & Clements detected counterparts of HSV-1 genes USt0 and US12. In addition, from the sizes of the HSV-2 sequences, we think it likely that HSV-2 versions of US9 and US11 also exist.

Subsidiary goals in our analysis were to locate the additional DNA sequences in HSV-2 Us and to identify the gene encoding gG-2. We found that the HSV-2-specific sequence lay within the coding region of gene US4, so that the HSV-2 US4 product was, essentially, equivalent to the corresponding HSV-1 protein with insertion of a large amino acid sequence near the N terminus. It was thus predicted that the HSV-2 US4 gene encoded a protein of approximate Mr 72000 and that this possessed sequence features appropriate for a virion glycoprotein. US4 was therefore the primary candidate for the gG-2 gene, previously mapped to this locality, and this assignment was confirmed using an oligopeptide antiserum. HSV-1 therefore encodes a much smaller protein as its equivalent of gG-2 (Frame et al., 1986; Richman et al., 1986). The scale of this size difference between two corresponding glycoproteins is without parallel in other HSV glycoproteins (although it is known that gC-1 is somewhat larger than gC-2; Swain et al., 1985). Our analyses have not yielded any information on possible functions of gG-1 and gG-2, but they do establish a basis for such studies.

We have examined evolutionary relations between the corresponding HSV-1 and HSV-2 Us sequences at several levels. First, it is seen, as expected, that the sequences differ by many single nucleotide changes, and this certainly represents accumulation of mutations in both strains since their divergence from a common progenitor. As might be expected, the mismatch divergence is higher in non-coding than in coding sequences. Within coding DNA, the largely redundant third positions of codons are most diverged, while the second positions, whose nature strongly correlates with the type of encoded amino acid, are most conserved. However, it is clear that at the whole gene level many amino acid substitutions are tolerable. A similar result can be obtained from comparison of other genes of the serotypes, for instance with the thymidine kinase genes, located in the long unique region (Swain & Galloway, 1983).

Turning to addition/deletion changes between the two sequences, we can discern several such types of change. The most large scale is the US4 gene event; this is discussed separately below.
On a smaller scale, we distinguish two classes of addition/deletion. These are, first, changes which have occurred in sequences having a repetitive nature and, second, changes in sequences lacking any visible repetitive element. The first of these classes is seen in non-coding sequences, where changes in lengths of homopolymeric runs are a prominent feature, and in one instance in coding sequence. This is in gene US7, where a reiterated sequence in HSV-1 is absent in HSV-2 (McGeoch et al., 1985). In principle, such changes in homopolymer tracts or in copy number of reiterations could result either from aberrant replicative copying or from unequal recombination.

The single most interesting feature in the divergence of the sequences is the large, HSV-2-specific sequence found in HSV-2 gene US4 relative to its HSV-1 counterpart. We have tried to address the question of whether this resulted from an insertion in a progenitor of HSV-2 or from a deletion in an HSV-1 precursor. To this end, characteristics of the HSV-2-specific sequence were examined. However, we were unable to reach any conclusion from this approach (not shown in Results). Thus, although the HSV-2-specific segment has a G+C content above average, it is not above average in the composition of its codon third positions. The high relative G+C content is seen in the first and second positions, and is therefore a reflection of the amino acids encoded. Since these must be presumed to be subject to functional constraints, the DNA cannot be regarded as 'foreign' because it has an atypical base composition.

The other data bearing on the evolution of the US4 genes are the proposed relations of gG-2 with PRV gX and HSV gD. The rather convincing similarity between gG-2 and gX is most simply accounted for by proposing that HSV-2 US4, including the HSV-2-specific sequence, represents an ancient gene of alphaherpesviruses, present before divergence of HSV and PRV. This then suggests that present-day HSV-1 US4 is a deleted version generated at or after divergence of HSV-1 and HSV-2. An alternative to this scheme would involve HSV-2 acquiring the extra sequences from some PRV-like virus in an 'inter-species' transfer, a possibility which we regard as less satisfactory.

Turning to the weaker similarity observed between gG-2 and gD, we favour the view that this too indicates a common evolutionary origin, at least in part, for these genes; that is, we propose that at some point gene duplication has played a part in alphaherpesvirus evolution. In this connection, we note that genes US4, US5, US6, US7 and US8 are adjoining genes proposed all to encode glycoproteins. This arrangement has no known parallel elsewhere in HSV, and duplication plus divergence would certainly constitute an attractive mechanism for evolution of such an array. We have not detected any other, comparable examples of sequence similarity to support this notion (however, we note here that the US7 protein sequences do bear a low level resemblance to the gG-2 sequence near their N termini, most noticeably in the occurrence of cysteine residues; analysis of this will be pursued separately). Gene duplication would proceed by mechanisms of aberrant recombination, which have already been invoked to account for the differences in gene organization in the short regions of the genomes of HSV and varicella-zoster virus (VZV) (Davison & McGeoch, 1986).

The observed sequence similarities between gG-2, PRV gX and gD are restricted to regions near the N termini, of the order of 80 amino acids long. This size is consistent with these regions representing structural or functional domains. We can thus discern several regions within the whole gG-2 sequence, as follows: (i) the N terminus, including the signal sequence, (ii) the region with counterparts in PRV gX and gD, (iii) the region between residues 347 and 537, where 32% of the residues are serine plus threonine, and which is proposed to be the major locus of O-glycosylation, and (iv) the C-terminal region, including the proposed transmembrane sequence, which is clearly similar to gG-1. We propose that, to some degree, these regions may all represent distinct modules in gG-2 structure and function.

Since HSV-1 apparently lacks sequences corresponding to most of gG-2, it may well be that the major, HSV-2-specific part of gG-2 is not essential for virus growth and viability. This does not preclude it from having some serotype-characteristic role, and we regard it as quite possible that such a large change in a surface glycoprotein could account in part for biological differences between the HSV serotypes. In this connection, it is interesting that VZV lacks glycoprotein genes corresponding to HSV US4, US5 and US6 (McGeoch, 1984; Davison & McGeoch, 1986).
It is known that the $U_S$ region of the genome of PRV contains genes for at least three glycoproteins. These are: $g_X$, of 498 amino acids (Rea et al., 1985); gp50, with an estimated $M_r$ of 50000 (Wathen & Wathen, 1985); and $g_A$, with an estimated $M_r$ of 122000 (Mettenleiter et al., 1985; Lukacs et al., 1985). From mapping analyses, the genes can be deduced to lie in the order $g_X$–gp50–$g_A$, although they are not necessarily contiguous. In addition, it is known that the $g_X$ gene is transcribed towards gp50 (Rea et al., 1985). We have shown that $g_X$ appears to be related to $g_G$-2. On the basis of size estimates of the PRV proteins and by comparison with the arrangement of glycoprotein genes in HSV $U_S$, we suggest that $g_A$ may be equivalent to HSV $g_E$, and that gp50 may be related either to HSV $g_D$ or to the US7 gene product.

Finally, we note that the HSV-2-specific part of gene US4 represents, as far as we know, an extensive DNA sequence without a counterpart in the HSV-1 genome, and it is thus of potential use as a serotype-specific analytical reagent.

We owe thanks to A. J. Davison for provision of the HindIII I plasmid and for discussion, to F. J. Rixon for extensive discussion and advice, and to J. H. Subak-Sharpe, F. J. Rixon and H. S. Marsden for criticism of the manuscript.

REFERENCES


DNA sequence of HSV-2 HindIII I 37


(Received 5 June 1986)