J. gen. Virol. (1986), **67**, 597–611. *Printed in Great Britain*

Key words: *HSV-1/VZV/genome evolution/inverted repeats*

Evolutionary Comparisons of the S Segments in the Genomes of Herpes Simplex Virus Type 1 and Varicella-Zoster Virus

By A. J. DAVISON† AND D. J. McGEOCH*

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

(*Accepted 19 December 1985*)

SUMMARY

The genomes of herpes simplex virus type 1 (HSV-1) and varicella-zoster virus (VZV) consist of two covalently joined segments, L and S. Each segment comprises an unique sequence flanked by inverted repeats. We have reported previously the DNA sequences of the S segments in these two genomes, and have identified protein-coding regions therein. In HSV-1, the unique sequence of S contains ten entire genes plus the major parts of two more, and each inverted repeat contains one entire gene; in VZV, the unique sequence of S contains two entire genes plus the major parts of two more, and each inverted repeat contains three entire genes. In this report, an examination of polypeptide sequence homology has shown that each VZV gene has an HSV-1 counterpart, but that six of the HSV-1 genes have no VZV homologues. Thus, although these regions of the two genomes differ in gene layout, they are related to a significant degree. The analysis indicates that the inverted repeats are evidently capable of large-scale expansion or contraction during evolution. The differences in gene layout can be understood as resulting from a small number of recombinational events during the descent of HSV-1 and VZV from a common ancestor.

INTRODUCTION

Members of the family *Herpesviridae* are classified into three subfamilies: the *Alpha-*, *Beta-* and *Gamma-herpesvirinae*. This subclassification is based on properties of the virus particle, on features of the replication cycle and on biological aspects, such as host range and site of latency (Matthews, 1982). These criteria have proved useful but, despite some attempts to include genome organization as an additional taxonomical factor, the evidence for relationships between the subfamilies has been circumstantial. Comparisons of currently emerging herpesvirus DNA sequences will allow the relationships to be investigated precisely and directly at the genetic level. For example, recent comparisons of sequences from herpes simplex virus type 1 (HSV-1) and Epstein–Barr virus, which are members of the *Alpha-* and *Gamma-herpesvirinae*, respectively, have demonstrated a degree of conservation in the predicted amino acid sequences of at least three virus-coded proteins: the two subunits of the ribonucleotide reductase (Gibson *et al.*, 1984), the DNA polymerase (Quinn & McGeoch, 1985; Baer *et al.*, 1984) and the major DNA-binding protein (Quinn & McGeoch, 1985). Nevertheless, the available evidence indicates that the *Alphaherpesvirinae* and *Gammaherpesvirinae* are rather distantly related. There are as yet no sequence data indicating the degree of relationship between the *Betaherpesvirinae* and the other subfamilies.

In contrast, studies of antigenic relatedness and DNA homology have shown members of the same subfamily to be much more closely related. For example, Davison & Wilkie (1983), using the technique of DNA–DNA hybridization, detected conservation of several genes in five members of the *Alphaherpesvirinae*. They proposed from the colinear arrangement of conserved

† Present address: Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205, U.S.A.

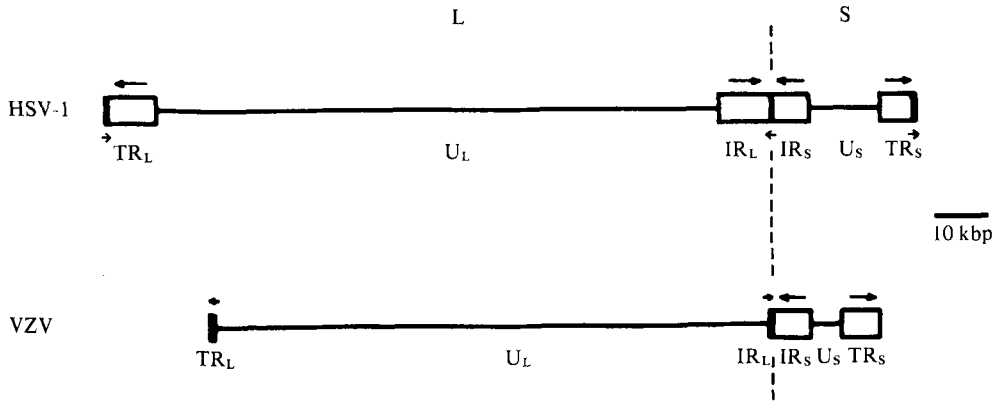


Fig. 1. HSV-1 and VZV genome structures. Both genomes contain two covalently joined segments (L, S), each comprising an unique sequence (U_L, U_S; indicated by solid lines) flanked by inverted repeats (TR_L, IR_L, TR_S, IR_S; indicated by open rectangles orientated by arrows). The HSV-1 DNA molecule contains a sequence of approximately 400 bp (indicated by small arrows) as a direct repeat at the termini and as an inverted repeat at the L–S joint, whereas the VZV genome is not terminally redundant (Davison & Scott, 1984). HSV-1 DNA populations contain equimolar proportions of four genome arrangements differing in the relative orientations of L and S. VZV DNA also contains four genome arrangements, but two predominate because, although S may be present in either orientation with equal probability, approximately 95% of molecules contain L in the same orientation (Davison & Scott, 1984).

genes that the gene layout might also be generally colinear in those regions for which no hybridization was detected. If proved true, this hypothesis would be of great value in advancing the experimental molecular genetics of less-studied members of the *Alphaherpesvirinae*. In this context, we are currently determining and comparing the DNA sequences of two of the more distantly related of the *Alphaherpesvirinae*, HSV-1 and varicella-zoster virus (VZV), which are both human pathogens.

A complete description of the relationships between HSV-1 and VZV involves several levels of analysis. One of the more important centres on the effects and possible causes of the striking difference between the base compositions of the two genomes: HSV-1 DNA contains 67% G+C (Becker *et al.*, 1968; Kieff *et al.*, 1971), whereas VZV DNA contains 46% G+C (Ludwig *et al.*, 1972). This aspect has been dealt with in some detail elsewhere (McGeoch, 1984; Honess, 1984; McGeoch *et al.*, 1986). In this paper, however, we focus primarily on a second important aspect: the relationship between HSV-1 and VZV at the level of the locations and organization of genes. This work consolidates and extends our preliminary results concerning this aspect of herpesvirus evolution (McGeoch, 1984).

We describe in this paper a comparative analysis of the genes in the S segments of HSV-1 and VZV. These regions were potentially the most interesting to compare because they differ significantly in size and have been shown by DNA hybridization experiments to be among the least related parts of the two genomes (Davison & Wilkie, 1983). We compared the amino acid sequences of predicted proteins in order to identify related genes. The underlying relationship between HSV-1 and VZV in these regions was determined, and many of the differences in gene layout were shown to be explicable as resulting from expansion and contraction of the inverted repeats during evolution.

METHODS

The analysis of predicted amino acid sequences was carried out using three programs running in a DEC PDP-11 computer operating under the RSX11M system:

The matrix program (Pustell & Kafatos, 1982). This program scores the number of residues in a window of length W from one sequence which are identical to those in a window of equal length from a second sequence. A weighted average is calculated by multiplying the value of each match by A^n (A is the weighting factor and n is the distance in nucleotides of the match from the centre of the window), and expressed as a percentage of the score that would be obtained from perfectly matched sequences. The windows are shifted by intervals of one residue and the results

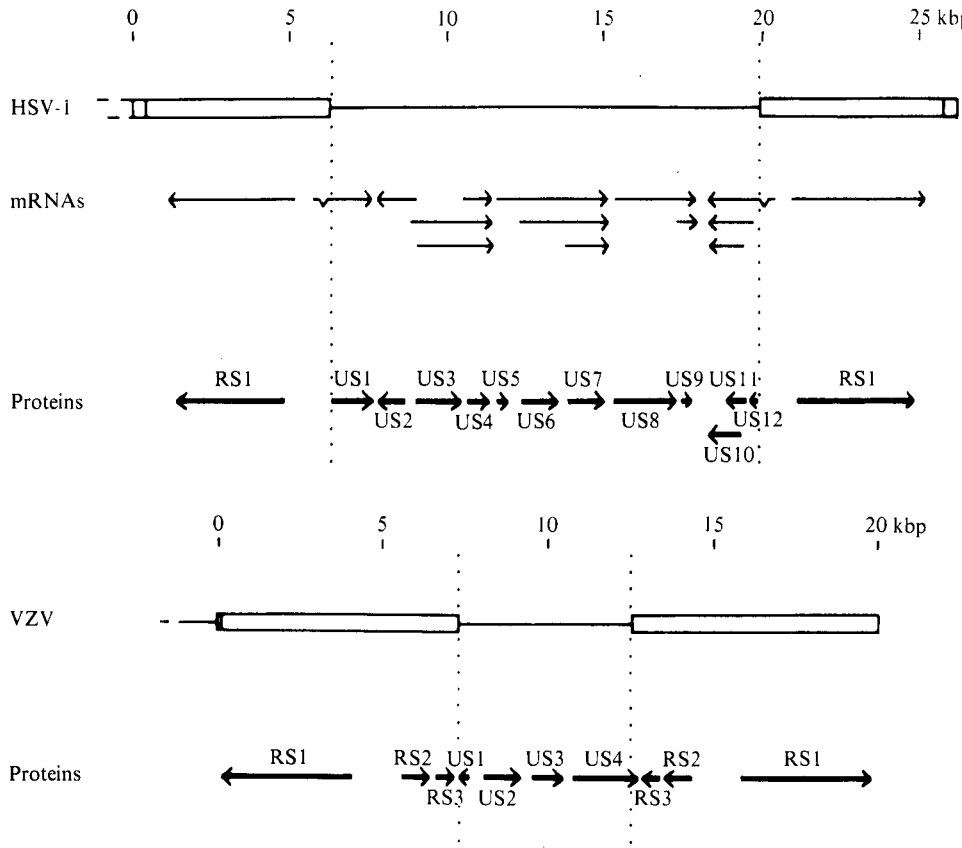


Fig. 2. Gene layout in the S segments of HSV-1 and VZV. The HSV-1 mRNA locations were reported by Rixon *et al.* (1982), Rixon & Clements (1982) and Rixon & McGeoch (1984, 1985). The locations of HSV-1 protein-coding regions (here called RS1 and US1 to US12) were described by Murchie & McGeoch (1982), and McGeoch *et al.* (1985, 1986). No mRNA map is available for VZV, but protein-coding regions (here called RS1 to RS3 and US1 to US4) were described by Davison (1983) and Davison & Scott (1985).

obtained as a two-dimensional (matrix) plot. A program devised by P. Taylor in the Institute of Virology was used to obtain matrix plots from a graphics printer. In this study, the values of W and A were 41 and 0·98, respectively. For most comparisons, the scores were plotted in four ranges: 0 to 24%, 25 to 49%, 50 to 74% and 75 to 100% homology.

The alignment program (Taylor, 1984). Regions for which matches were detected using the matrix program were aligned directly by inserting blank characters in the sequences at appropriate points.

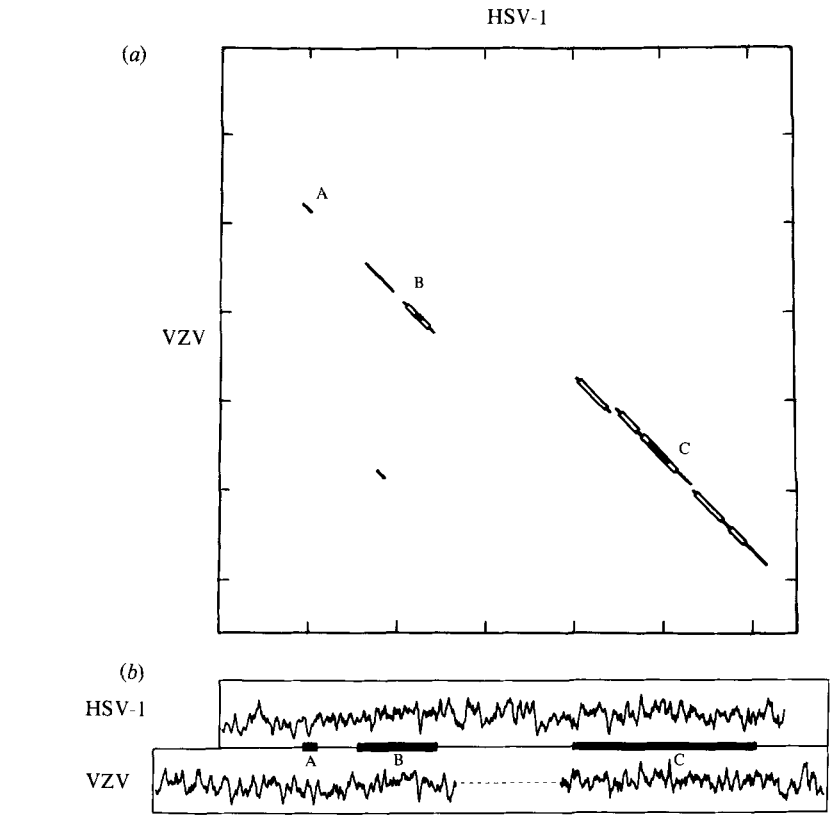
The hydrophobicity program (Kyte & Doolittle, 1982). Hydrophobicity values for each amino acid were summed over a window of nine residues shifted by intervals of one residue. Graphical results were obtained, with peaks representing hydrophobic regions.

The matrix and alignment programs were also used to analyse DNA sequence homology in selected regions.

RESULTS AND DISCUSSION

The genome structures of HSV-1 and VZV are depicted in Fig. 1, and the gene layout in the S segment of each virus is shown in Fig. 2. HSV-1 U_S contains ten entire genes plus the major parts of two others, and each copy of the short repeat (TR_S and IR_S) contains one gene. In contrast, VZV U_S contains two entire genes plus the major parts of two others, and each copy of the short repeat (TR_S and IR_S) contains three genes.

The matrix program was used to compare the amino acid sequences of predicted HSV-1 proteins with those of VZV proteins. Counterparts to all seven VZV genes were detected. Comparisons of each gene with all the other genes in the HSV-1 and VZV S segments showed no additional significant homologies. The alignment program was used to display regions of



(c)

```
183 DSGSSSSSSASSSSSSSDEDEDDDGNDA  HSV-1
      **** *** ** *****                 A
348 SREFVSSSSSSSSWGSSSEDEDDEPRRV  VZV

308 DGYVSGEPWPGAGPPPPGRVLYGGLGDSRPGLWGAPEAEEARRRFEASGAPAAVWAPELGDAAQQYALITRLLYTPDAEAMGWLQNPRVVPGDVALDQACFRISGAARNSSSFITGSVARAVPHLG
461 LMMPDGSPWPGSAPLPSNRVRFGPSGETREGHWEDEAVRAARARYEASTEPVPLYVPELGDPARQYRALINLIYCPDRDPIAWLQNPKLTGVNSALNQFYQKLLPPGR AGTAVTGSVASPVPHVG

                                    tsK
                                    ↑
434 YAMAAGRFGWGLAHAAAAVAMSRRYDRAQKGFLLTSLRRAYAPLLARENAALTGAAGSPGA  HSV-1
      *** *   * * ******************** * * ***** *     *      **                B
586 EAMATGEALWALPEAAAAVAMSRRYDRAQKHFILQSLRRAFASMAYPEATGSSPAARISRG  VZV

792 QARPRPVAVSRRPAEGPDPLGGWRRQPPGPSHTAAPAAAALEAYCSPRAVAELTDHPLFPVPWRPALMFDPRALASIAARCAGPAPAAQAACGGGDDDDNPHPHGAAGGRLFGPLRASGPLRRMAA
730 GPVVDPPAVITMPLDGPAPNGGFRRIPRGALHTPVPSDQARKAYCTPETIARLVDDPLFPTAWRPALSFDPGALAEIAAR          RPGGGD              RRPGPPSGVEALRRRCA

                ←—                                  tsD, tsT                          —→
                <                                                                     >
918 WMRQIPDPEDVRVVVLYSPLPGED    LAGGGASGGPPEWSAERGGLSCLLAALANRLCGPDTAAWAGNWTGAPDVSALGAQGVLLLSTRDLAFAGAVEFLG LLASAGDRRLIVVNTVRACDWP
833 WMRQIPDPEDVRLLIIYDPLPGEDINGPLESTLATDPGPSWSPSRGGLSVVLAALSNRLCLPSTHAWAGNWTGPPDVSALNARGVLLLSTRDLAFAGAVEYLGSRLASA RRRLLVLDAVALERWP

1039 ADGPAVSRQHAYLACELLPAVQCAVRWP   AARDLRRTVLASGRVFGPGVFARVEAAHARLYPDAPPLRLCRGGNVRYRVRTRFGPDTPVPMSPREYRRAVLPALDGRAAASGTTDAMAPGAPDF
958 RDGPALSQYHVYVRAPARPDAQAVVRWPDSAVTEGLARAVFASSRTFGPASFARIETAFANLYPGEQPLCLCRGGNVAYTVCTRAGPKTRVPLSPREYRQYVLPGFDGCKDLARQSRGLGLGAADF

1162 CEEEAHSHAACARWGLGAPLRPVYVALGREAVRAGPARWRGP  RRDFCARALLEPDDDAPPLVLRGDDD  HSV-1
                                                                                C
1084 VDEAAHSHRAANRWGLGAALRPVFLPEGRRPGAAGPEAGDVPTWARVFCRHALLEPDPAAEPLVLPPVAG  VZV
```

Fig. 3. Comparative data for HSV-1 RS1 and VZV RS1. In the matrix plot (*a*) the axes are marked every 200 amino acid residues; a diagonal line indicates 37 to 49% homology, an open rectangle indicates 50 to 74% homology, and a filled rectangle indicates 75 to 100% homology. In this and the hydrophobicity plot (*b*) the three conserved regions (A, B, C) are indicated (these regions have been aligned by separating the two parts of the VZV plot by a dashed line), and the homology plot (*c*) shows the aligned amino acid sequences of these regions. Also indicated on the homology plot are the locations of the lesions in HSV-1 temperature-sensitive mutants *ts*K, *ts*D and *ts*T.

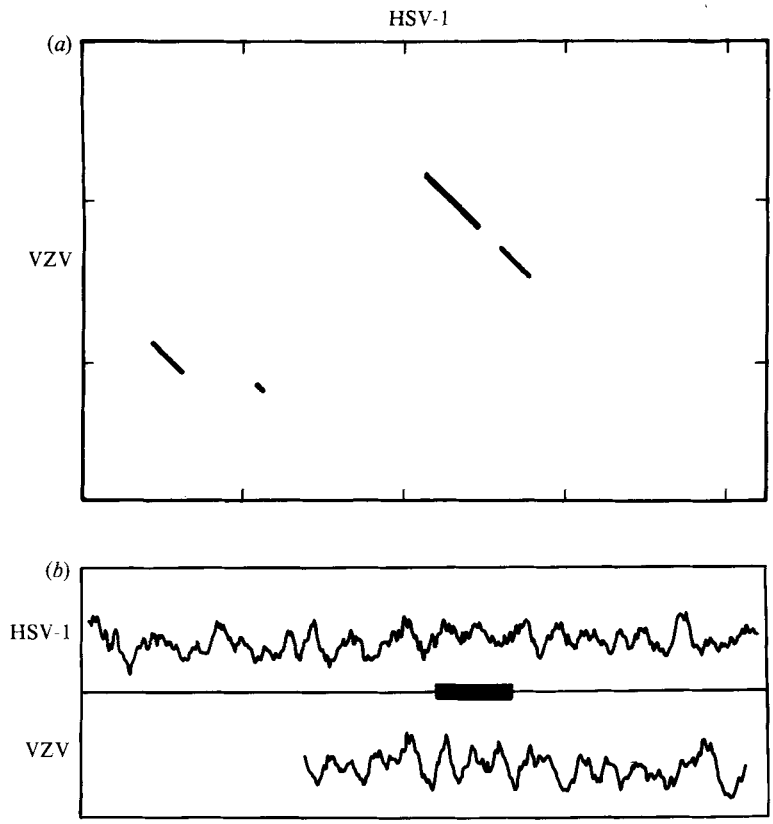


Fig. 4. Comparative data for HSV-1 US1 and VZV RS2. In the matrix plot (*a*) the axes are marked every 200 residues, and a diagonal line indicates 25 to 49% homology. The conserved region is denoted as a filled rectangle in the hydrophobicity plots (*b*), and shown as a direct alignment of amino acids in the homology plot (*c*).

similarity in detail, and hydrophobicity plots were compared. The results of the three types of analysis for the seven pairs of partially homologous proteins are shown in Fig. 3 to 9, and the relationships between genes in the S segments of HSV-1 and VZV are summarized in Fig. 10.

The pairs of genes exhibit a wide range of amino acid homology, ranging from the obvious similarity between HSV-1 RS1 and VZV RS1 (Fig. 3) to the very limited relationship between HSV-1 US10 and VZV RS3 (Fig. 5). Homologous amino acid residues tend to be localized rather than distributed evenly throughout each pair of proteins. The borderline between significant and fortuitous similarity is not easily assessed, and it is only the presence of less common amino acids [in this case, tryptophan (W) and cysteine (C) residues] in the sequence shared by HSV-1 US10 and VZV RS3 which suggests that the relationship is significant.

The amino-terminal portions of homologous proteins are among the least conserved, not only in sequence but also in the alignment of amino termini relative to conserved regions. The latter feature may be appreciated by extending the line indicating homology in each matrix plot towards the amino termini. Thus, HSV-1 US1, US3, US10 and VZV RS1, US1 and US4 have relatively more residues in this portion than their counterparts. Only HSV-1 US7 is similar to its counterpart at the amino terminus in this respect. In contrast, the carboxy termini of each pair are relatively better aligned, as may be visualized by extending the line indicating homology in

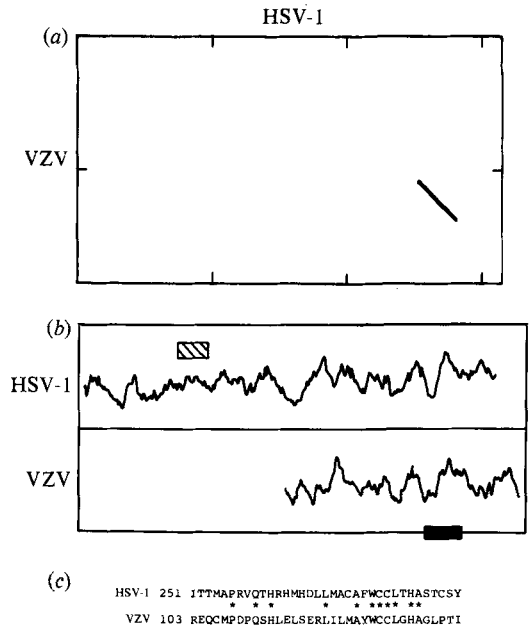


Fig. 5. Comparative data for HSV-1 US10 and VZV RS3. In the matrix plot (*a*) the axes are marked every 100 residues, and a diagonal line indicates 25 to 49% homology. The main conserved region is denoted as a filled rectangle in the hydrophobicity plots (*b*), and shown as a direct alignment of amino acids in the homology plot (*c*). The hatched box marks the location in HSV-1 US10 of a partially repeated amino acid sequence comprising seven copies of a tripeptide.

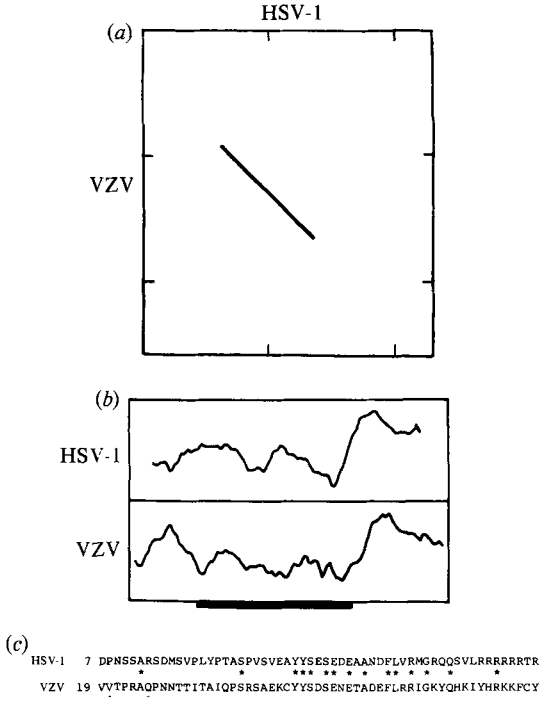
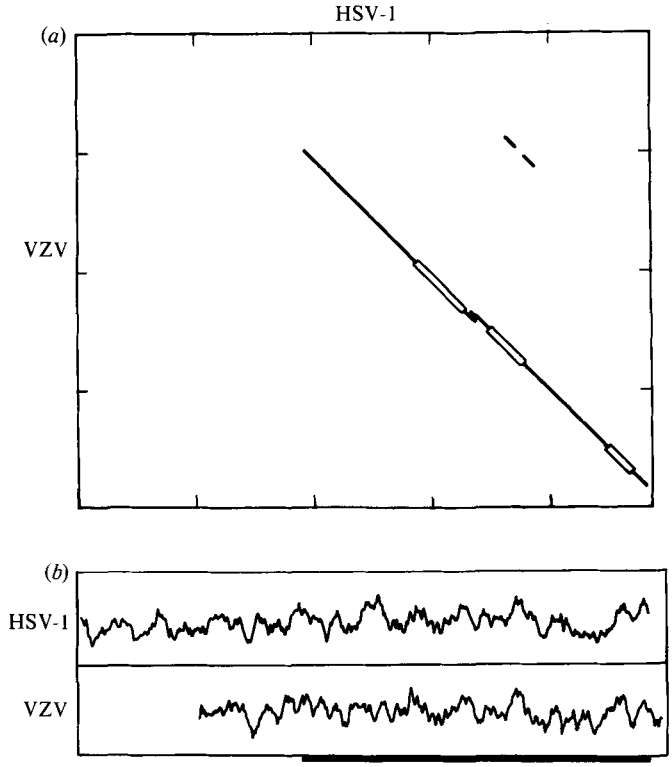


Fig. 6. Comparative data for HSV-1 US9 and VZV US1. In the matrix plot (*a*) the axes are marked every 40 residues, and a diagonal line indicates 25 to 49% homology. The conserved region is denoted as a filled rectangle in the hydrophobicity plots (*b*), and shown as a direct alignment of amino acids in the homology plot (*c*).



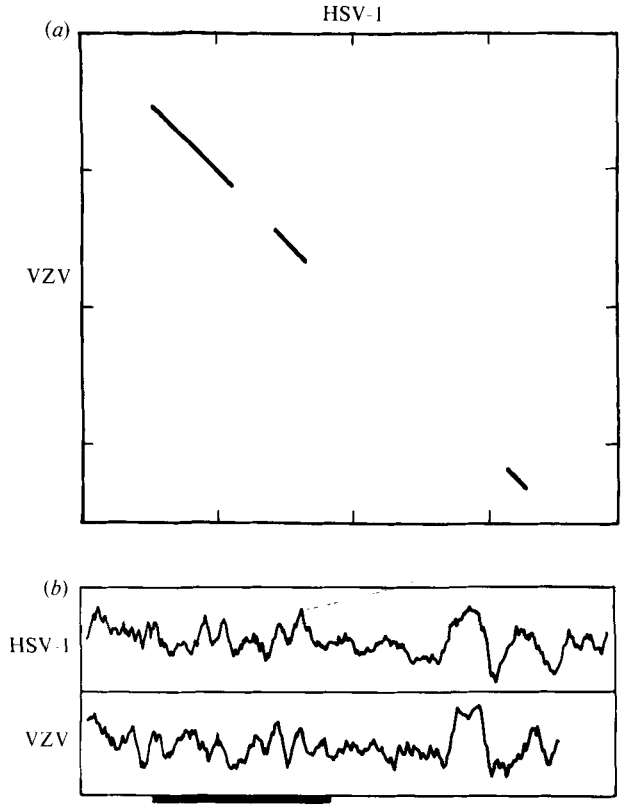
(c)

```
185 LVTGMGFTIHGALTPGSEGCVFDSSHPDYPQRVIVKAGWYTSTSHEARLLRRLDHPAILPLLDLHVVSGVTCLVLPKYQADLYTYLSRRLNPLGRPQIAAVSRQLLSAVDYIHRQGIIHRDIKTEN
87 GINKAGFVILKTFTPGAEGFAFACMDSKTCEHVVIKAGQRQGTATEATVLRALTHPSVVQLKGTFTYNKMTCLILPRYRTDLYCYLAAKRN LPICDILAIQRSVLRALQYLHNNSIIHRDIKSEN

311 IFINTPEDICLGDFGAACFVQGSRSSPFPYGIAGTIDTNAPEVLAGDPYTTTVDIWSAGLVIFETAVHNASLFSAPRGPKRGPCDS QITRIIRQAQVHVDEFSPHPESRLTSRYRSRAAGNNRP
212 IFINHPGDVCVGDFGAACF PVDINANRYYGWAGTIATNSPELLARDPYGPAVDIWSAGIVLFEMATGQNSLF ERDGLDGNCDSERQIKLIIRRSGTHPNEFPINPTSNLRRQYIGLAKRSSRK

435 PYTRPAWTRYYKMDIDVEYLVCKALTFDGALRPSAAELLCLPLFQQK HSV-1
335 PGSRPLWTNLYELPIDLEYLICKMLSFDARHRPSAEVLLNHSVFQTL VZV
```

Fig. 7. Comparative data for HSV-1 US3 and VZV US2. In the matrix plot (*a*) the axes are marked every 100 residues; a diagonal line indicates 25 to 49% homology, and an open rectangle indicates 50 to 74% homology. The conserved region is denoted as a filled rectangle in the hydrophobicity plots (*b*), and shown as a direct alignment of amino acids in the homology plot (*c*).

each matrix plot towards the carboxy termini. Only slight misalignment is apparent between HSV-1 US8 and US10 and their counterparts. This difference between amino and carboxy termini may reflect greater constraints either in DNA structure at the 3′ termini of the genes or in protein function at the carboxy termini. Although amino termini are predicted less reliably from DNA sequence data than are carboxy termini, because proteins need not be initiated at the first ATG in an open reading frame, there is good reason from the analyses of the DNA sequences (Murchie & McGeoch, 1982; Davison, 1983; Davison & Scott, 1985; McGeoch *et al.*, 1985, 1986) to suppose that the amino termini have been identified correctly. The transcript mapping data summarized in Fig. 2 allowed the amino termini of HSV-1 proteins to be identified with greater confidence than those of VZV proteins.

Although the relationship between HSV-1 and VZV genes is most apparent at the level of primary protein structure, it is observed to a lesser extent in the DNA sequences encoding conserved amino acid sequences. However, little similarity was detected either between DNA sequences encoding non-conserved amino acid sequences or between DNA sequences not coding for protein. Thus, no extensive homology was detected between presumptive gene promoter regions. Fig. 11 shows the similarity between the DNA sequences of HSV-1 RS1 and



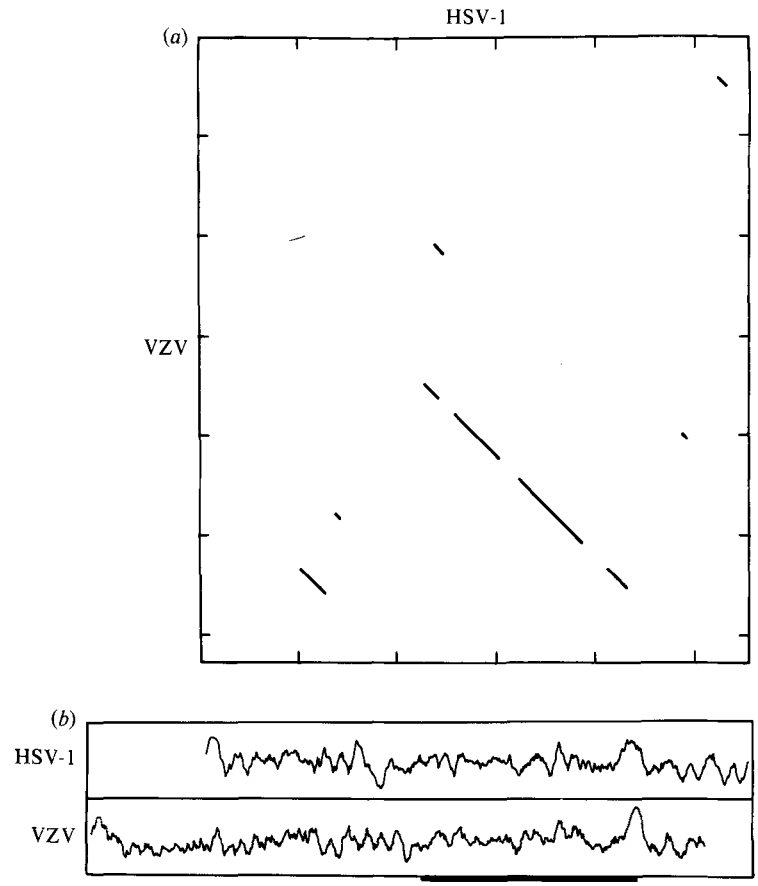
(c)

```
48 EEDLLILGELRFVGDQVPHTTYYDGGVELWHYPMGHKCPRVVHVVTVTACPRRPAVAFALCRATDSTH   SPAYPTLELNLAQQPLLRVQRATRDYAGVYVLRVWVGDAPNASLFVLG MAIAAE
46 DNYTEIKGQLVFIGEQLPTGTNYSGTLELLYADTVAFCFRSVQVIRYDGCPRIRTSAFISCRYKHSWHYGNSTDRISTE PDAGVMLKITKPGIND AGVYVLLVRLDHSRSTDGFILGVNVYTAG

170 GTLAYNGSAYGSCDPKL  HSV-1
170 SHHNIHGVIYTPSPLQN  VZV
```

Fig. 8. Comparative data for HSV-1 US7 and VZV US3. In the matrix plot (*a*) the axes are marked every 100 residues, and a diagonal line indicates 25 to 49% homology. The main conserved region is denoted as a filled rectangle in the hydrophobicity plots (*b*), and shown as a direct alignment of amino acids in the homology plot (*c*).

VZV RS1 in one of the major regions of amino acid conservation. The observation that overall DNA sequence conservation is correspondingly greater in this pair of genes than in the others is in accord with the previous results of Davison & Wilkie (1983), who detected DNA–DNA hybridization between the S segments of HSV-1 and VZV only in the regions occupied by the RS1 genes.

The results of analyses of amino acid sequences using the matrix and alignment programs are limited in that they show only regions containing identical residues. Inspection of the alignment plots shows that many of the non-identical residues in conserved regions are similar in structure, charge or hydrophobicity, and thus represent conservative differences. Moreover, the hydrophobicity plots demonstrate similarities not only in regions containing identical residues but also in regions of little or no precise homology. For example, the carboxy termini of HSV-1 US9 and VZV US1 are similarly hydrophobic, and regions immediately preceding the conserved sequences in HSV-1 US1 and VZV RS2, and HSV-1 US10 and VZV RS3, have similar hydrophobicity profiles. Such similarities suggest the existence of a greater degree of structural conservation than would be concluded from the extent of identical residues alone. It is clear that the significance attached to the analysis is limited by the algorithms used, and these in



(c)

```
215 EVSHVRGVTVRMETPEAILFSPGETFSTNVSIHAIAHDDQTYSMDVVWLRFDVPTSCAEMRIYESCLYHPQLPECLSPADAPCAASTWTS      RLAVRSYAGC SRTNPPPRCSAEAHMEPVPG
332 VTPQPRGAEFHMWNYHSHVFSVGDTFSLAMHLQYKIH EAPFDLLLEWLYVPIDPTCQPMRLYSTCLYHPNAPQCLSHMNSGC    TFTSPHLAQRVASTVYQNCEHADNYTAYCLGISHMEPSFG

335 LAWQAASVNLEFRDASPQHSGLYLCVVYVNDHIHAWGHITISTAAQYRNAVVEQPLPQRGADLAEPTHPHVGAPPHAPPTHGALRLGAVMG  AALLLSALGLSVW  HSV-1
454 LILHDGGTTLKFVDTPESLSGLYVFVVYFNGHVEAVAYTVVSTVDHFVNAIEERGFP PTAGQPPATTKPKEITPVNPGTSPLLRYAAWTGGLAAVVLLCLVIFLI  VZV
```

Fig. 9. Comparative data for HSV-1 US8 and VZV US4. In the matrix plot (*a*) the axes are marked every 100 residues, and a diagonal line indicates 25 to 49% homology. The main conserved region is denoted as a filled rectangle in the hydrophobicity plots (*b*), and shown as a direct alignment of amino acids in the homology plot (*c*).

turn are constrained by our ignorance of the extent to which protein function is influenced by primary amino acid sequence. One suspects that more subtle analytical procedures developed in the future will strengthen the evidence for functional, as well as structural, conservation.

Structural conservation of proteins suggests at least a degree of common function, and the more highly conserved regions probably correspond to important functional sites. Unfortunately, the functions of most of the proteins coded by the HSV-1 and VZV S segments are unknown. Two of the genes in the VZV S segment encode glycoproteins: VZV US4 encodes a major envelope glycoprotein (Ellis *et al.*, 1985) which is homologous to the product of HSV-1 US8, also an envelope glycoprotein, gE (Hope *et al.*, 1982; Lee *et al.*, 1982; Para *et al.*, 1982; Hope & Marsden, 1983; McGeoch *et al.*, 1985); VZV US3 is a glycoprotein gene (Davison *et al.*, 1985) and its counterpart, HSV-1 US7, has been predicted to encode a glycoprotein (McGeoch *et al.*, 1985; McGeoch, 1985). The hydrophobicity plot for each of these proteins exhibits a hydrophobic region near the amino terminus, corresponding to a signal sequence for translation on membrane-bound ribosomes, and towards the carboxy terminus there is a region containing a

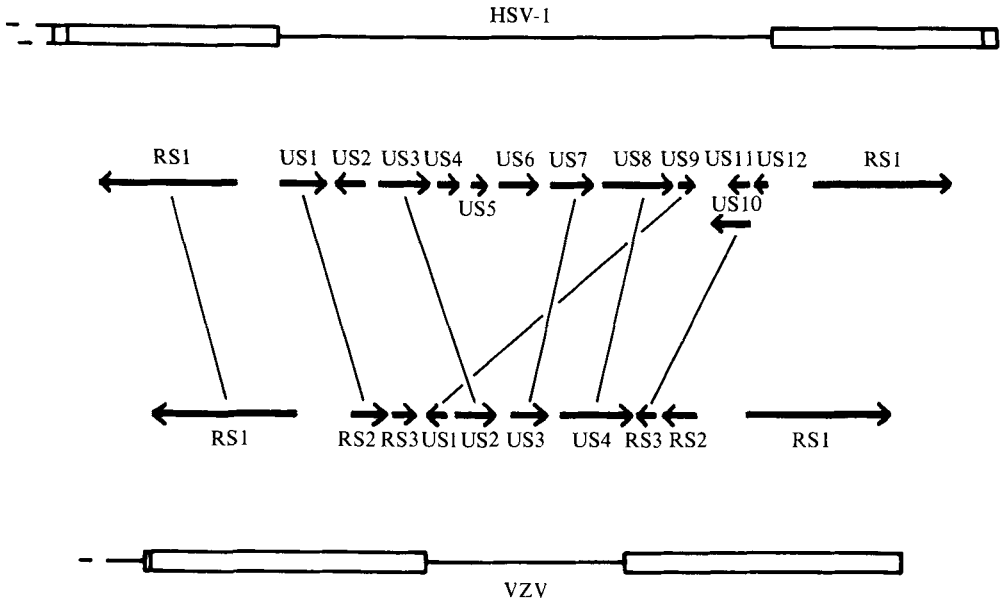


Fig. 10. Relationships between genes in the S segments of HSV-1 and VZV. For clarity, relationships are indicated for only one copy of genes in the inverted repeats.

highly hydrophobic domain followed by basic residues, which is thought to be a transmembrane anchor sequence. The locations of potential glycosylation sites of sequence asparagine, any amino acid, serine (NXS) or asparagine, any amino acid, threonine (NXT) are not conserved in these proteins.

HSV-1 RS1 encodes an immediate-early protein, $V_{mw}175$, which is involved in transcriptional activation of early and late genes (Preston, 1979; Watson & Clements, 1980; Dixon & Schaffer, 1980). The VZV counterpart may function similarly, since plasmids containing VZV RS1 are able to activate an early HSV-1 gene (Everett, 1984). This pair of genes is the most conserved in the S segments of HSV-1 and VZV, and two main regions of homology are present (regions B and C in Fig. 3), separated by a non-homologous region which is approximately 200 residues longer in HSV-1. The two proteins also share a serine-rich sequence (region A in Fig. 3) closer to the amino terminus. The lesions in several temperature-sensitive HSV-1 mutants have been mapped in HSV-1 RS1. One of these, *ts*K, abolishes the ability of $V_{mw}175$ to activate later genes at the non-permissive temperature (Marsden *et al.*, 1976). The lesion maps in conserved region B, and results in the replacement of an alanine (A) by a valine (V) residue, as indicated in Fig. 3 (Davison *et al.*, 1984). The corresponding residue in VZV RS1 is also alanine. A second group of mutants containing *ts*D and *ts*T, which are less affected in their ability to activate later genes at the non-permissive temperature than *ts*K, has been mapped to conserved region C (Fig. 3) (Preston, 1981). These observations indicate that regions B and C correspond to two important domains, which may fulfil different, and possibly physically separable, functions.

HSV-1 US1 also encodes an immediate-early protein which, although presumably essential in the normal life-cycle of the virus, is apparently dispensable for virus growth *in vitro* in some cell lines (Post & Roizman, 1981; Sears *et al.*, 1985). Its counterpart, VZV RS2, may be functionally equivalent, although a degree of doubt should be sensibly entertained in view of the low degree of amino acid conservation. Similar hesitation regarding functional equivalence may be applied to the barely homologous products of HSV-1 US10 and VZV RS3. An unusual feature of the protein coded by HSV-1 US10, which is not shared by VZV RS3, is that it contains a partially repeated amino acid sequence comprising seven copies of the tripeptide proline, glycine, X (PGX), where X is leucine (L), serine (S) or proline (P) (Rixon & McGeoch, 1984). The location of this region is indicated in Fig. 5. Also, HSV-1 US10 partially overlaps HSV-1 US11 (Fig. 2),

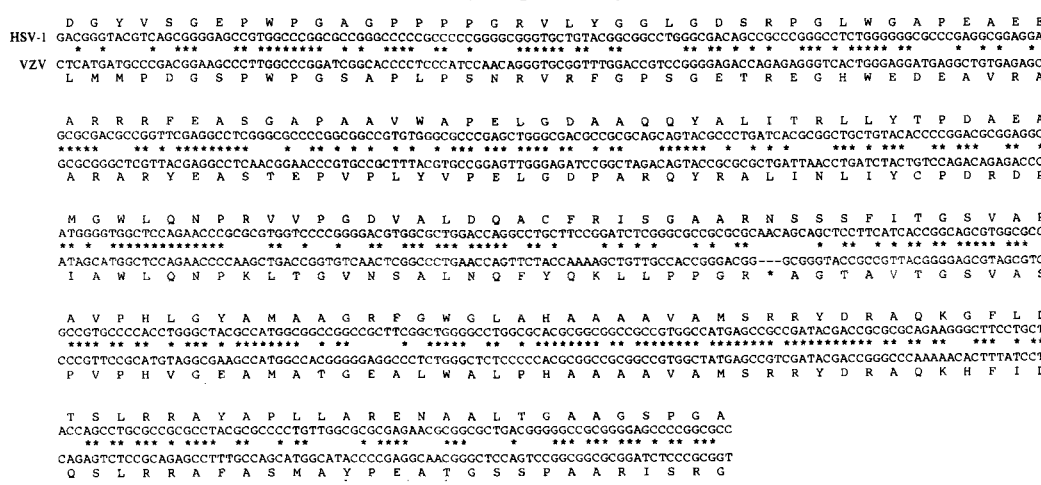


Fig. 11. Homology plot of the DNA sequences of HSV-1 RS1 and VZV RS1 in conserved region B (see Fig. 3). The encoded amino acid sequences are also shown. In order to obtain alignment, three adjacent blank characters, indicated by dashes, were inserted into the VZV DNA sequence.

which has no counterpart in VZV. HSV-1 US9 encodes a small virion protein which is extensively modified by phosphorylation (Frame *et al.*, 1986). Despite a high degree of conservation between HSV-1 US3 and VZV US2, there is no experimental indication of the functions of this pair of genes. However, both predicted polypeptides share homology with several eukaryotic protein kinases (McGeoch & Davison, 1986).

It is apparent from Fig. 10 that the HSV-1 S segment contains six genes for which no counterparts were detected in the VZV S segment. Our unpublished sequence data indicate that counterparts are not present elsewhere in the VZV genome. Of these HSV-1 genes, US6 encodes the most extensively studied HSV-1 glycoprotein, gD (Watson *et al.*, 1982; McGeoch *et al.*, 1985). US12 is a small immediate-early gene (Watson *et al.*, 1979; Marsden *et al.*, 1982; Murchie & McGeoch, 1982; Watson & vande Woude, 1982). US11 encodes a protein which binds specifically to sequences at the HSV-1 genome termini and L–S joint (Dalziel & Marsden, 1984). The structures of the proteins coded by US4, US5 and perhaps US2 suggest that they are glycoproteins (McGeoch *et al.*, 1985; McGeoch, 1985). The absence of counterparts in the VZV S segment might correlate in part with differences between HSV-1 and VZV in their biological behaviour. However, a full explanation is likely to encompass a combination of the differences in expression and structure of many of the virus-encoded proteins.

The observed relationship between genes in the S segments of the HSV-1 and VZV genomes may be explained as resulting from descent of these two viruses from an ancestral herpesvirus. An understanding of the processes by which the S segments might have diverged hinges on the relative locations of the junctions between U_S and the inverted repeats. In HSV-1, the junctions are located 8 and 40 bp, respectively, from the initiating ATG codons in US12 and US1 (Murchie & McGeoch, 1982). Whitton & Clements (1984) proposed that the inverted repeats are capable of expansion during evolution to include sequences from U_S, since the junctions in HSV-2 are located 1 and 33 bp from the corresponding ATG codons. They also concluded that expansion of the repeats may be limited by the locations of adjacent protein coding regions in U_S. The VZV data support this conclusion because the termination codon for US1 is located at the junction, and that for US4 is located within the repeat (Davison, 1983), such that any further expansion of the repeats would interfere drastically with the coding potential of these genes. However, whereas the HSV-1/HSV-2 comparison indicates that the inverted repeats may be able to expand by up to 7 bp, the HSV-1/VZV data imply that the repeats may expand or contract during evolution to include or expel entire genes. Thus, two genes in HSV-1 U_S (US1 and US10) have counterparts in the VZV inverted repeat (RS2 and RS3).

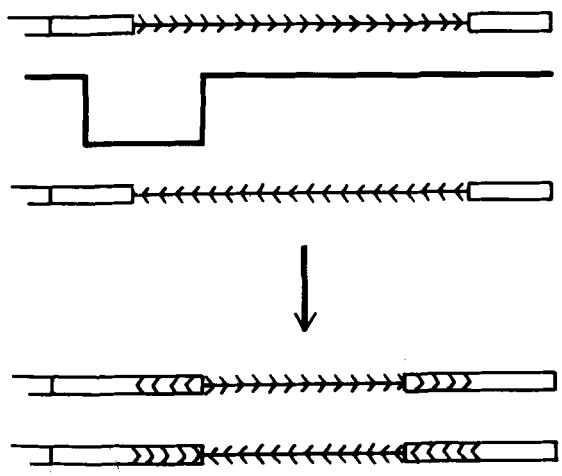


Fig. 12. A simple recombination scheme to account for movement of the junctions between unique and repeat sequences. The upper part of the figure shows the S segment of the two genomes aligned in opposite orientations, as indicated by the arrowheads. The heavy lines between them denote two crossovers, the first between completely homologous sequences in the inverted repeat and the second between dissimilar sequences in the unique regions. The lower part shows the two progeny molecules. In this example, approximately equal lengths of DNA are exchanged, resulting in expansion of the repeats. Certain types of unequal exchanges involving recombination at the second crossover between sequences in the inverted repeat of one molecule and the unique region of the other would result in contraction of the repeat.

The process by which movement of the endpoints of the repeats has occurred is likely to be mediated primarily by recombination. For many years both intratypic and intertypic recombination in HSV-1 and HSV-2 have been known to occur with ease (Wildy, 1955; Timbury & Subak-Sharpe, 1973). Homologous recombination between the inverted repeats is probably the chief mechanism by which they retain their sequence identity, and by which mutations in one repeat are transmitted to the other (Wilkie *et al.*, 1977). Recombination may also serve to disseminate changes in base composition throughout a herpesvirus population (McGeoch *et al.*, 1986). An explanation of how recombination between S segments in opposite orientations could lead to movement of the endpoints of the repeats is given in Fig. 12. One recombination event occurs between completely homologous sequences in the repeat, and the second event takes place between dissimilar sequences in U_S which perhaps possess limited homology. The process may also be envisaged as taking place by gene conversion. Although this type of partially non-homologous recombination is therefore different from normal homologous recombination, evidence for its occurrence in laboratory stocks of HSV-1 has been reported recently by Umene & Enquist (1985).

Fig. 13 shows a scheme which relates the structures of the S segments of HSV-1 and VZV by a discrete number of steps involving expansion or contraction of the repeats. The steps imply no temporal orientation since, in considering the differences between present-day genomes, the assignment of directions in time tends to be arbitrary and misleading. This device is intended to retrace forms from HSV-1 towards an ancestral herpesvirus and thence to VZV. Gene loss in one direction would be seen in the other direction as input from an outside source. Thus, starting with HSV-1, the following conversions to generate the VZV gene arrangement are shown in Fig. 13.

(A) Gene 1 is taken into the repeats, and genes 11 and 12 are lost from the opposite end of U_S. (B) Genes 9 and 10 are taken into the repeats, and gene 2 is lost from the opposite end of U_S. (C) Gene 9 is lost from the repeats, the remaining copy appearing at the opposite end of U_S from its position in HSV-1. (D) The downstream end of gene 8 is taken into the repeats, and the termination codon of gene 9 is moved adjacent to the junction. (E) Genes 4, 5 and 6 are deleted.

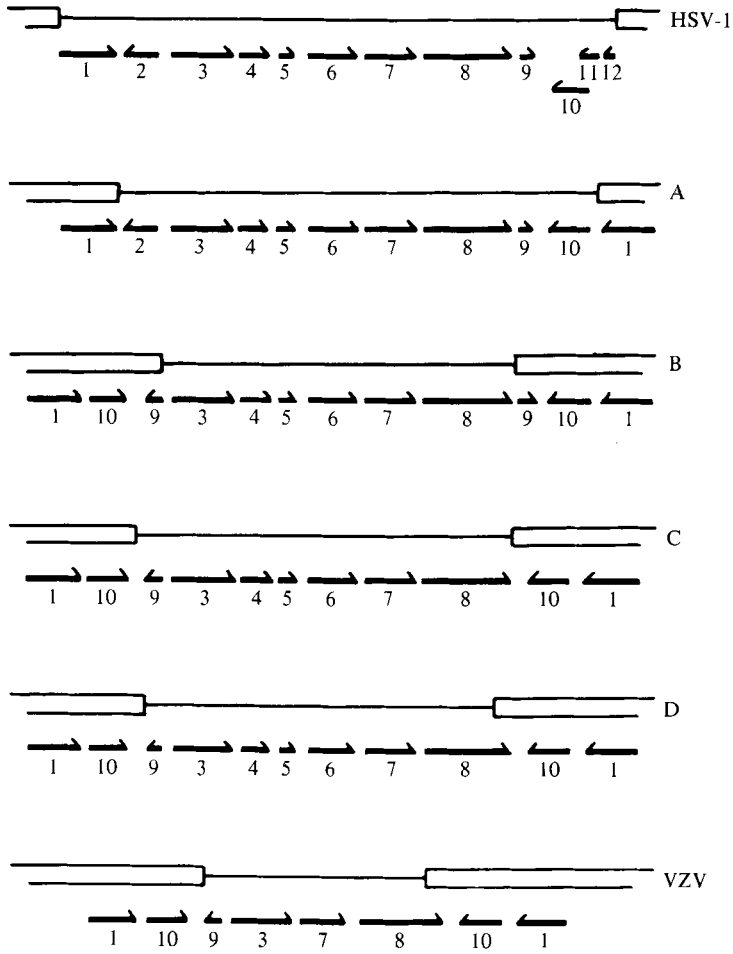


Fig. 13. Descent of the S segments of HSV-1 and VZV from that of an ancestral herpesvirus. The HSV-1 genes US1 to US12 are indicated, and this numbering scheme has been retained in the remainder of the figure. No commitment to direction in time has been made.

This step could also be accomplished by a more elaborate movement of the repeats, and need not necessarily occur after steps A to D.

Thus, the inverted repeat is evidently a dynamic structure capable of varying in extent during evolution. The forces which operate to bring this about are largely undefined and probably complex. Nonetheless, it is clear that the coding potential of regions close to the junctions forms one important element in defining the extent of the repeats. The flexible nature of the inverted repeat indicates that the presence in the genome of two copies of the genes encoded therein may not be of primary importance in virus growth, and renders it unlikely that the repeat is an element in which certain distinct functions are compartmentalized. Finally, it would be unwise to overlook the potential contribution to be made by other members of the *Alphaherpesvirinae* to the study of evolution of the subfamily. For example, the inverted repeat of pseudorabies virus (PRV) is 15 kbp in size, considerably larger than those of HSV-1 and VZV, and PRV U_S is only slightly smaller, at 9 kbp, than that of HSV-1 (Ben Porat *et al.*, 1979). It is thus possible that the PRV S segment contains more genes than the HSV-1 and VZV S segments, and a future detailed examination of this and other members of the *Alphaherpesvirinae* would be expected to add to our concept of herpesvirus evolution.

A fundamental relationship between the S segments of HSV-1 and VZV has been identified, but the gene arrangements are substantially different. Hence, one could not deduce the layout of genes in the S segment of one virus from that of the other. In this region of the genomes the predictive value of the hypothesis put forward by Davison & Wilkie (1983) is limited, as was suggested by these authors in noting the differences in size between regions of the genome. Nevertheless, it is apparent from a comparison of the complete VZV DNA sequence (A. J. Davison & J. E. Scott, unpublished results) with available HSV-1 data that the two genomes share a similar gene layout in the majority of the L segment, thus consolidating the value of the hypothesis in locating the majority of genes in the genomes of other members of the *Alphaherpesvirinae*.

The authors thank Philip Taylor for willing and extensive help with computer programming.

REFERENCES

BAER, R., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHWELL, S. C., SEGUIN, C., TUFFNELL, P. S. & BARRELL, B. G. (1984). DNA sequence and expression of the B95-8 Epstein–Barr virus genome. *Nature, London* **310**, 207–211.

BECKER, Y., DYM, H. & SAROV, I. (1968). Herpes simplex virus DNA. *Virology* **36**, 184–192.

BEN-PORAT, T., RIXON, F. J. & BLANKENSHIP, M. L. (1979). Analysis of the structure of the genome of pseudorabies virus. *Virology* **95**, 285–294.

DALZIEL, R. G. & MARSDEN, H. S. (1984). Identification of two herpes simplex virus type 1-induced proteins (21K and 22K) which interact specifically with the *a* sequence of herpes simplex virus DNA. *Journal of General Virology* **65**, 1467–1475.

DAVISON, A. J. (1983). DNA sequence of the U_S component of the varicella-zoster virus genome. *EMBO Journal* **2**, 2203–2209.

DAVISON, A. J. & SCOTT, J. E. (1984). Structure of the genome termini of varicella-zoster virus. *Journal of General Virology* **65**, 1969–1977.

DAVISON, A. J. & SCOTT, J. E. (1985). DNA sequence of the major inverted repeat in the varicella-zoster virus genome. *Journal of General Virology* **66**, 207–220.

DAVISON, A. J. & WILKIE, N. M. (1983). Location and orientation of homologous sequences in the genomes of five herpesviruses. *Journal of General Virology* **64**, 1927–1942.

DAVISON, A. J., WATERS, D. J. & EDSON, C. M. (1985). Identification of the products of a varicella-zoster virus glycoprotein gene. *Journal of General Virology* **66**, 2237–2242.

DAVISON, M.-J., PRESTON, V. G. & McGEOCH, D. J. (1984). Determination of the sequence alteration in the DNA of the herpes simplex virus type 1 temperature-sensitive mutant *ts* K. *Journal of General Virology* **65**, 859–863.

DIXON, R. A. F. & SCHAFFER, P. A. (1980). Fine structure mapping and functional analysis of temperature sensitive mutants in the gene encoding the herpes simplex virus type 1 immediate early protein VP175. *Journal of Virology* **36**, 189–203.

ELLIS, R. W., KELLER, P. M., LOWE, R. S. & ZIVIN, R. A. (1985). Use of a bacterial expression vector to map the varicella-zoster virus major glycoprotein gene, gC. *Journal of Virology* **53**, 81–88.

EVERETT, R. D. (1984). Transactivation of transcription of herpes virus products; requirements for two HSV-1 immediate early polypeptides for maximum activity. *EMBO Journal* **3**, 3135–3141.

FRAME, M. C., McGEOCH, D. J., RIXON, F. J., ORR, A. C. & MARSDEN, H. S. (1986). The 10K virion phosphoprotein encoded by U_S gene 9 from herpes simplex virus type 1. *Virology* (in press).

GIBSON, T., STOCKWELL, P., GINSBURG, M. & BARRELL, B. (1984). Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes. *Nucleic Acids Research* **12**, 5087–5099.

HONESS, R. W. (1984). Herpes simplex and 'the herpes complex': diverse observations and a unifying hypothesis. *Journal of General Virology* **65**, 2077–2107.

HOPE, R. G. & MARSDEN, H. S. (1983). Processing of glycoproteins induced by herpes simplex virus type 1: sulphation and nature of oligosaccharide linkages. *Journal of General Virology* **64**, 1943–1953.

HOPE, R.G., PALFREYMAN, J. W., SUH, M. & MARSDEN, H. S. (1982). Sulphated glycoproteins induced by herpes simplex virus. *Journal of General Virology* **58**, 399–415.

KIEFF, E. D., BACHENHEIMER, S. L. & ROIZMAN, B. (1971). Size, composition and structure of the deoxyribonucleic acid of herpes simplex virus subtypes 1 and 2. *Journal of Virology* **8**, 125–132.

KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.

LEE, G. T. Y., PARA, M. F. & SPEAR, P. G. (1982). Location of the structural genes for glycoproteins gD and gE and for other polypeptides in the S component of herpes simplex virus type 1 DNA. *Journal of Virology* **43**, 41–49.

LUDWIG, H., HAINES, H. G., BISWAL, N. & BENYESH-MELNICK, M. (1972). The characterization of varicella-zoster virus DNA. *Journal of General Virology* **14**, 111–114.

McGEOCH, D. J. (1984). The nature of animal virus genetic material. In *The Microbe 1984. Part I: Viruses*, pp. 75–107. Edited by B. W. J. Mahy & J. R. Pattison. Cambridge: Cambridge University Press.

McGEOCH, D. J. (1985). On the predictive recognition of signal peptide sequences. *Virus Research* **3**, 271–286.

McGEOCH, D. J. & DAVISON, A. J. (1986). Alphaherpesviruses possess a gene homologous to the protein kinase gene family of eukaryotes and retroviruses. *Nucleic Acids Research* (in press).

McGEOCH, D. J., DOLAN, A., DONALD, S. & RIXON, F. J. (1985). Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. *Journal of Molecular Biology* **181**, 1–13.

McGEOCH, D. J., DONALD, S., DOLAN, A. & BRAUER, D. H. K. (1986). Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Research* (in press).

MARSDEN, H. S., CROMBIE, I. K. & SUBAK-SHARPE, J. H. (1976). Control of protein synthesis in herpesvirus-infected cells: analysis of the polypeptides induced by wild type and sixteen temperature-sensitive mutants of HSV strain 17. *Journal of General Virology* **31**, 347–372.

MARSDEN, H. S., LANG, J., DAVISON, A. J., HOPE, R. G. & MacDONALD, D. M. (1982). Genomic location and lack of phosphorylation of the HSV immediate-early polypeptide IE 12. *Journal of General Virology* **62**, 17–27.

MATTHEWS, R. E. F. (1982). Classification and nomenclature of viruses. *Intervirology* **17**, 1–200.

MURCHIE, M.-J. & McGEOCH, D. J. (1982). DNA sequence analysis of an immediate-early gene region of the herpes simplex virus type 1 genome (map coordinates 0·950 to 0·978). *Journal of General Virology* **62**, 1–15.

PARA, M. F., GOLDSTEIN, L. & SPEAR, P. G. (1982). Similarities and differences in the Fc binding glycoprotein (gE) of herpes simplex virus types 1 and 2 and tentative mapping of the viral gene for this glycoprotein. *Journal of Virology* **41**, 137–144.

POST, L. E. & ROIZMAN, B. (1981). A generalized technique for deletion of specific genes in large genomes: α gene 22 of herpes simplex virus 1 is not essential for growth. *Cell* **25**, 227–232.

PRESTON, C. M. (1979). Control of herpes simplex virus type 1 mRNA synthesis in cells infected with wild type virus or the temperature sensitive mutant *ts*K. *Journal of Virology* **29**, 275–285.

PRESTON, V. G. (1981). Fine-structure mapping of herpes simplex virus type 1 temperature-sensitive mutations within the short repeat region of the genome. *Journal of Virology* **39**, 150–161.

PUSTELL, J. & KAFATOS, F. C. (1982). A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Research* **10**, 4765–4782.

QUINN, J. P. & McGEOCH, D. J. (1985). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. *Nucleic Acids Research* **13**, 8143–8163.

RIXON, F. J. & CLEMENTS, J. B. (1982). Detailed structural analysis of the two spliced HSV-1 immediate-early mRNAs. *Nucleic Acids Research* **10**, 2241–2256.

RIXON, F. J. & McGEOCH, D. J. (1984). A 3′ co-terminal family of mRNAs from the herpes simplex virus type 1 short region: two overlapping reading frames encode unrelated polypeptides one of which has a highly reiterated amino acid sequence. *Nucleic Acids Research* **12**, 2473–2487.

RIXON, F. J. & McGEOCH, D. J. (1985). Detailed analysis of the mRNAs mapping in the short unique region of herpes simplex virus type 1. *Nucleic Acids Research* **13**, 953–973.

RIXON, F. J., CAMPBELL, M. E. & CLEMENTS, J. B. (1982). The immediate-early mRNA that encodes the regulatory polypeptide Vmw175 of herpes simplex virus type 1 is unspliced. *EMBO Journal* **1**, 1273–1277.

SEARS, A. E., HALLIBURTON, I. W., MEIGNIER, B., SILVER, S. & ROIZMAN, B. (1985). Herpes simplex virus 1 mutant deleted in the α22 gene: growth and gene expression in permissive and restrictive cells and establishment of latency in mice. *Journal of Virology* **55**, 338–346.

TAYLOR, P. (1984). A fast homology program for aligning biological sequences. *Nucleic Acids Research* **12**, 447–456.

TIMBURY, M. C. & SUBAK-SHARPE, J. H. (1973). Genetic interactions between temperature-sensitive mutants of types 1 and 2 herpes simplex virus. *Journal of General Virology* **18**, 347–357.

UMENE, K. & ENQUIST, L. W. (1985). Isolation of novel herpes simplex virus type 1 derivatives with tandem duplications of DNA encoding immediate-early mRNA-5 and an origin of replication. *Journal of Virology* **53**, 607–615.

WATSON, R. J. & CLEMENTS, J. B. (1980). A herpes simplex virus type 1 function continuously required for early and late virus RNA synthesis. *Nature, London* **285**, 329–330.

WATSON, R. J. & VANDE WOUDE, G. F. (1982). DNA sequence of an immediate-early gene (IEmRNA-5) of herpes simplex virus type 1. *Nucleic Acids Research* **10**, 979–991.

WATSON, R. J., PRESTON, C. M. & CLEMENTS, J. B. (1979). Separation and characterization of herpes simplex virus type 1 immediate-early mRNAs. *Journal of Virology* **31**, 42–52.

WATSON, R. J., WEIS, J. H., SALSTROM, J. S. & ENQUIST, L. W. (1982). Herpes simplex virus type-1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. *Science* **218**, 381–384.

WHITTON, J. L. & CLEMENTS, J. B. (1984). The junctions between the repetitive and the short unique sequences of the herpes simplex virus genome are determined by the polypeptide-coding regions of the two spliced immediate-early mRNAs. *Journal of General Virology* **65**, 451–466.

WILDY, P. (1955). Recombination with herpes simplex virus. *Journal of General Microbiology* **13**, 346–360.

WILKIE, N. M., STOW, N. D., MARSDEN, H. S., PRESTON, V., CORTINI, R., TIMBURY, M. C. & SUBAK-SHARPE, J. H. (1977). Physical mapping of herpes simplex virus coded functions and polypeptides by marker rescue and analysis of HSV-1/HSV-2 intertypic recombinants. In *Oncogenesis and Herpesviruses III*, vol. I, pp. 11–31. Edited by G. de The, W. Henle & F. Rapp. Lyon: IARC Scientific Publications, no. 24.

(Received 7 November 1985)