

Putative Glycoprotein Gene of Varicella-Zoster Virus with Variable Copy Numbers of a 42-Base-Pair Repeat Sequence Has Homology to Herpes Simplex Virus Glycoprotein C

PAUL R. KINCHINGTON,^{1*} JAMES REMENICK,² JEFFREY M. OSTROVE,³ STEVEN E. STRAUS,³ WILLIAM T. RUYECHAN,¹ AND JOHN HAY²

Departments of Biochemistry¹ and Microbiology,² Bethesda, Maryland 20814, and Uniformed Services University of the Health Sciences,² and Laboratory of Clinical Investigation, National Institute of Allergy and Infectious Diseases,³ Bethesda, Maryland 20205

Received 7 March 1986/Accepted 29 May 1986

A strain variation of varicella-zoster virus that maps to the U_L region of the genome was found to be due to different copy numbers of a high GC 42-base-pair repeat. DNA sequence analysis of this variable region showed the sequence to be 5-GCGGGATCGGGCTTTCGGG(A/T)AGCGGCCGAGGTGGGCGCGACG-3. Strains Scott and Webster both contain 7 and $\frac{32}{42}$ copies of the repeat, whereas strain Oka has exactly 4 copies less. Microheterogeneity exists within the repeated sequences, depending on the strain and the repeat number. Sequencing of the entire *EcoRI* P fragment (which contains the repeated sequences) and part of the adjacent *EcoRI* M and *EcoRI* Q fragments from strain Scott showed that the repeats are part of a large open reading frame that could code for a polypeptide core with a molecular weight of 66,000. Several potential TATA boxes exist upstream and two polyadenylation signals are found downstream of the open reading frame. The predicted protein bears several characteristics of a glycoprotein. The region is transcriptionally active in varicella-zoster virus-infected cells, specifying at least three RNA species of 1.7, 1.95, and 2.5 kilobases, which are transcribed from the same DNA strand. Part of the predicted protein has a high degree of homology to the herpes simplex virus type 1 glycoprotein gC.

Many of the members of the *Herpesviridae* family show variation in DNAs from different strains; some examples are herpes simplex viruses types 1 and 2 (HSV-1, HSV-2; 3, 17, 32, 33), equine herpesvirus type 1 (1), *Herpesvirus saimiri* (12), and Epstein-Barr virus (27). Strain variation seems to occur by at least two different mechanisms. First, there may be small changes in the DNA sequence at specific positions, which result in the loss or gain of particular restriction endonuclease recognition sites. Second, interstrain variation can arise by the insertion or deletion of sequences which do not profoundly alter the virus phenotype. It is clear that many variations are quite stable, both during replication of the virus in the natural host and in extensive propagation of the virus in tissue culture.

Similar variations in DNAs from different isolates of varicella-zoster virus (VZV) were also reported (4, 35, 37, 42). VZV is the causative agent of two distinct diseases: varicella, or chicken pox, and zoster, or shingles, the latter representing a reactivation of latent VZV. The genome of VZV is a linear double-stranded DNA molecule (125 kilobase pairs) composed of a long unique region (U_L) and a short unique region (U_S) bounded by inverted repeats (IR_S/TR_S). Both regions are capable of inverting with respect to each other, with U_L inverting at a much lower frequency than U_S (7, 8, 13, 14, 19, 34, 36).

Analyses of DNA prepared from virus isolated from patients with either disease and examined by restriction endonuclease digestion (4, 35, 37) suggest that each epidemiologically unrelated strain produces a distinctive set of restriction enzyme patterns, permitting clear differentiation of VZV strains. In contrast, DNA from strains in related outbreaks of VZV infection provide identical restriction

profiles. Three variable regions have been mapped on the genome; two of them have been localized to U_L , characterized by *EcoRI* fragments P and G (4, 35, 42). The third region of variation lies in IR_S/TR_S (4; Fig. 1). This region is characterized by different copy numbers of a high GC 27-base-pair (bp) repeat sequence in different strains and lies in an apparently noncoding region upstream from the 5' ends of two open reading frames coding for putative proteins of molecular weights of 140,000 and 30,000 (4, 9). It was suggested that this sequence may be part of a transcription-controlling element; it bears limited homology to similarly located sequences in HSV (4, 10, 24).

We analyzed a second variable region in VZV DNA that maps in the *EcoRI* P fragment. This region, like the IR_S/TR_S variation, was found to be composed of multiple copies of a high GC sequence, but, in contrast, seemed to lie inside the coding region for a viral glycoprotein that has some homology to HSV-1 glycoprotein gC.

MATERIALS AND METHODS

Cells and viruses. VZV strains Oka, Webster, and Ellen were obtained from the American Type Culture Collection. Strain Scott was a gift from G. Fischer, Department of Pediatrics, Uniformed Services University of the Health Sciences. All viruses were grown in human foreskin fibroblast cells as previously described (30, 34).

DNAs. VZV DNA fragments generated by *EcoRI* and *BamHI* were cloned into pBR325 or GT.wesB as previously described (29, 33) or into pUC9 (39). Clones containing *EcoRI*-P or *BamHI*-F were identified for further analysis after colony hybridization. DNA fragments were separated by electrophoresis in 1 or 1.5% agarose gels (SeaKem LE type; FMC Corp., Marine Div., Rockland, Maine) run at 40 mA for 18 to 24 h in 0.1 M Tris borate-0.01 M EDTA (pH

* Corresponding author.

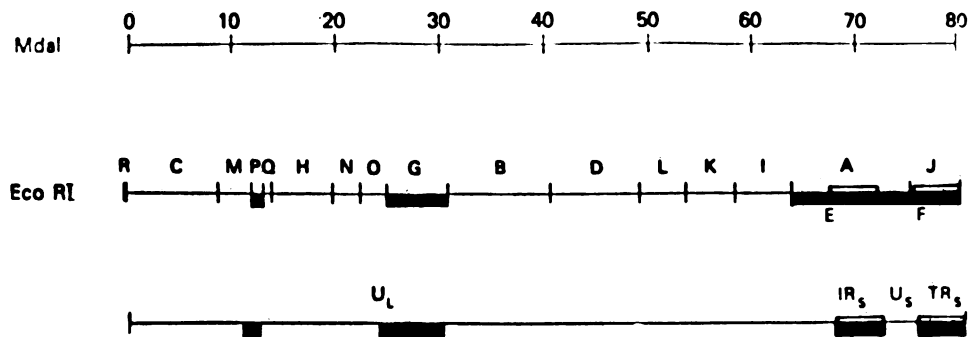


FIG. 1. Location of the strain-variable fragments on the physical map of VZV DNA for *EcoRI*. ■, The map positions of the VZV DNA restriction fragments from different strains that have altered mobilities. A summary of the variable regions derived from several restriction maps (4) is shown below. Above the map is a scale showing the molecular mass (in megadaltons [Mdal]) of the VZV genome.

8.3). Fragments were purified from agarose gels by electroelution with an analytical electroelution apparatus (model UEA; International Biotechnologies Inc., New Haven, Conn.) and concentrated by ethanol precipitation.

Construction of deletion fragments and DNA sequencing. DNA fragments were generated by (i) restriction endonuclease digestion with enzymes purchased from Bethesda Research Laboratories, Inc., Gaithersburg, Md., or from Boehringer Mannheim Biochemicals, Indianapolis, Ind., and cloned into M13mp18 or M13mp19 or (ii) progressive deletion of sequences with BAL 31 (mixed; Boehringer Mannheim) as described by Poncz et al. (26). Briefly, *EcoRI*-P cloned into pUC9 or pBR325 in predetermined orientations was purified from CsCl-ethidium bromide gradients and linearized with *Bam*HI or *Pvu*II. The DNA was then digested with BAL 31 for predetermined times. Digestion was halted by the addition of EGTA (ethylene glycol-bis(β -aminoethyl ether)-*N,N,N',N'*-tetraacetic acid) to 0.02 M, and the DNA was then digested with *EcoRI*. The resulting fragments were separated by agarose gel electrophoresis and purified by electroelution. The resulting *EcoRI* P subfragments possessed one *EcoRI* terminus and one blunt terminus. These were subsequently cloned into M13. Single-stranded DNA purified from recombinant extracellular phage was sequenced by primer extension in conjunction with dideoxynucleotide chain termination as previously described (31).

Heteroduplex mapping by electron microscopy. Cloned *EcoRI* P fragments derived from VZV strains Oka and Scott were purified by extraction from low-melt agarose gels as previously described (22). Heteroduplex mapping by electron microscopy was performed as described for VZV *EcoRI* fragment E (4). Briefly, equal amounts of the DNA fragment from both strains were mixed with recrystallized formamide to yield final concentrations of 1 μ g of DNA per ml and 80% (vol/vol) formamide in a 100- μ l volume. The DNA was denatured at 70°C for 5 min and allowed to reanneal for 24 h after the addition of 50 μ l of 0.2 M Tris hydrochloride–0.02 M EDTA (pH 8.5). The resulting heteroduplexes were analyzed with a Zeiss EM10A transmission electron microscope at a magnification of $\times 8,000$.

Northern blot analysis of VZV RNAs. The transcripts encoded in the *EcoRI*-P region of VZV DNA were identified by hybridization of 32 P-labeled DNA as described by Ostrove et al. (25). Whole-cell RNA was purified from VZV Ellen-infected Flow 5000 cells (Flow Laboratories, Inc., McLean, Va.) using the guanidinium isothiocyanate–CsCl procedure. RNA (5 μ g) was electrophoretically separated on

1% formaldehyde-agarose gels and transferred to nitrocellulose (BA85; Schleicher & Schuell, Inc., Keene, N.H.).

Single-stranded DNA probes were prepared by primer extension of M13 clones containing VZV DNA in predetermined orientations. Single-stranded DNA purified from such recombinants was annealed with 17-base universal primer (Boehringer Mannheim). Primer extension was carried out at 15°C in a mixture containing 2.5 μ g of annealed DNA, 200 μ M each dGTP, dATP, and dTTP, 20 μ Ci of [32 P]dCTP (>3,000 Ci/mmol; New England Nuclear Corp., Boston, Mass.), 50 mM Tris hydrochloride (pH 7.4), 6 mM MgCl₂, 10 mM dithiothreitol, and 2 U of DNA polymerase 1 (Klenow fragment) for 1 h. The reaction was then halted by inactivation at 65°C for 10 min, digested with *EcoRI* and *Hind*III, and electrophoresed on 1% agarose gels. The resulting fragment containing single-strand-labeled VZV DNA was purified by electroelution, denatured, and used to probe Northern blots of VZV RNA. Double-strand-labeled DNA was prepared by nick repair with [32 P]dCTP.

RESULTS

Identification and cloning of the variable fragments. The *EcoRI*-generated fragments from VZV DNA which varied among different strains are shown in Fig. 1, along with consensus regions for the map positions of the variable fragments. In all the strains analyzed, the variable fragments fell into three regions. One of these, which lies in *EcoRI*-P approximately 18 kilobase pairs from the left-hand end of U_L, is the subject of this paper.

To initiate the study of this variation in *EcoRI*-P, the fragment was cloned from strains Oka, Ellen, Scott, and Webster. DNA from each clone was then analyzed by agarose gel electrophoresis after digestion with *EcoRI*. One clone from Ellen, two clones from Scott, two clones from Oka, and one clone from Webster are shown in Fig. 2. Strains Scott and Webster have *EcoRI* P fragments of approximately 1,750 bp, whereas the strain Oka *EcoRI* P fragment is 150 to 200 bp smaller. *EcoRI*-P from strain Ellen is approximately 50 bp smaller than those from Scott and Webster. The cloned fragments are identical in size to the fragments generated from whole VZV DNA. Furthermore, independent clones from the same strain and different preparations of the same recombinant plasmid appeared identical (data not shown), indicating that the strain differences are stable to cloning. This is in contrast to the variable region present in *EcoRI*-G which we found to be highly unstable in size upon cloning into pBR325 and pUC9 (unpublished observations).

Heteroduplex mapping by electron microscopy. To determine both the number of possible variable regions and their locations within *EcoRI*-P, electron microscope analyses of heteroduplexes between purified P fragments from Scott and Oka were performed. An electron micrograph of a typical heteroduplex molecule is shown in Fig. 3. Analysis of 46 heteroduplexes showed that each molecule consists of mostly double-stranded DNA, with a collapsed single-stranded loop located at a mean fractional distance of 64% ($\pm 8\%$) from one end of the heteroduplexes. This result suggests that the difference between the two DNAs lies at one site only and implies that a single deletion or insertion exists toward the center of the fragment.

Sequencing of *EcoRI*-P. To determine the source of these differences, the *EcoRI* P fragments from different clones were partially sequenced. Deletions were generated by digestion with BAL 31 from a single *HincII* restriction site. Cleavage of *EcoRI*-P at this site generated 1,350- and 450-bp fragments from purified Scott *EcoRI*-P and 1,150- and 450-bp fragments from purified Oka *EcoRI*-P. After digestion with *EcoRI*, the termini of the truncated P fragments consisted of a cleaved *EcoRI* end and a blunt end. These were then cloned into the *EcoRI* and *SmaI* sites of M13mp10 replicative form, and the recombinant virus DNA was sequenced (31).

Preliminary results (Fig. 4) indicate that the strain difference is due to different copy numbers of a repeating se-

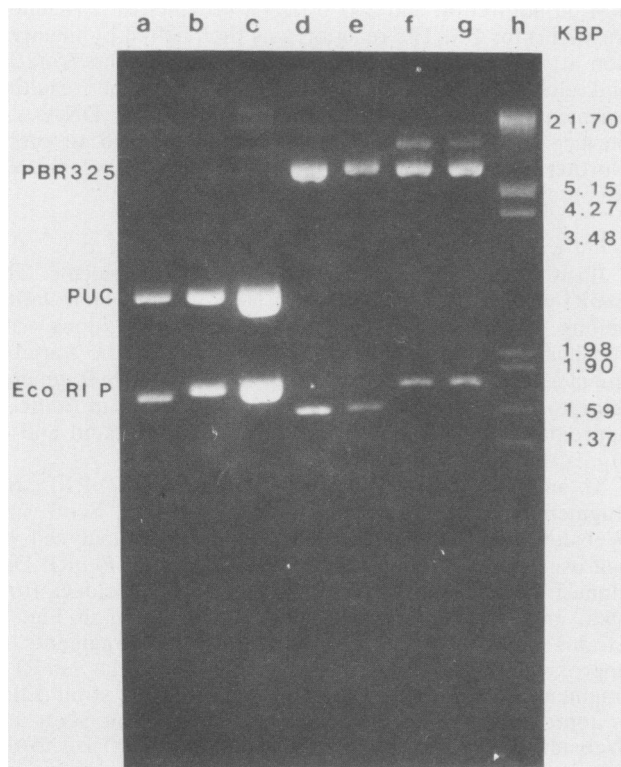


FIG. 2. Cloned *EcoRI* P fragments from DNA of VZV strains Ellen (a), Scott (b and c), Oka (d and e), and Webster (f and g). Two separate clones of Scott and Oka are shown. *EcoRI* P fragments of VZV Ellen and Scott were cloned into pUC9 (PUC), whereas fragments from Oka and Webster were cloned into pBR325. Lane h contains lambda DNA cut with *EcoRI* and *HindIII* to indicate fragment size. The sizes of the resulting fragments are shown at the right in kilobase pairs (kbp).

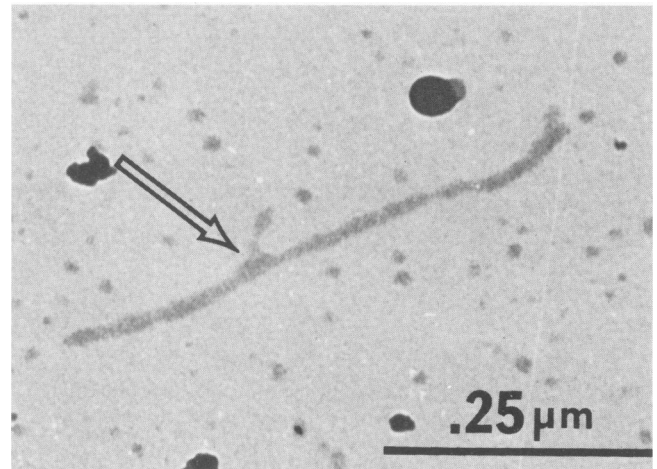


FIG. 3. Electron micrograph showing a heteroduplex molecule formed from the annealing of equal amounts of cloned purified *EcoRI*-P obtained from VZV Oka and Scott (described in Materials and Methods). An arrow indicates the location of the collapsed single-stranded region described in the text.

quence. The top of a gel in which clones of the Oka and Scott 1,150- and 1,350-bp fragments were sequenced from the *HincII* end is shown in Fig. 4. Although the nucleotide bands are too close to allow determination of the precise sequence, a regular ladder pattern can be seen, particularly in the C lanes, which represent repeating sequence units. There appear to be eight repeating units in strain Scott and four in Oka. Sequencing of additional deletions showing the repeating sequence to be 5'-GCGGGATCGGGCTTTTCGG(A/T)AGCGGCCGAGGTGGGCGCGACG-3' (reading from the single internal *HincII* site). It is 42 bp long and contains a single *HaeIII* site. Preliminary work from this laboratory previously showed that digestion of variable *EcoRI* P fragments with this enzyme removes the variation but yields a multimolar fragment of approximately 40 bp (T. A. Casey, Ph.D. thesis, Uniformed Services University of the Health Sciences, Bethesda, Md., 1984). Taken together, these data suggest that the difference between the two strains of VZV in *EcoRI*-P can be entirely explained by different copy numbers of a 42-bp repeat.

Precise sequencing of the repeat regions also showed (Fig. 5) that strains Scott and Webster contain 7 and $\frac{32}{42}$ copies of the repeat, whereas Oka contains precisely 4 copies less. In addition, microheterogeneity exists between different copies of the repeats, depending on the position and the strain. For example, the second repeat of strain Scott has a C to A substitution at position 12 (the only example of this change found so far) and in all strains base 20 is either A or T, depending on the strain and the location.

At this point, the question arose as to the location of this variable region relative to VZV genes, since the previously investigated variable region (in *TR_S/IR_S*) seemed to lie outside the polypeptide coding sequences (4, 9). Therefore, the DNA sequence of the region around the 42-bp repeat was determined. This eventually involved sequencing the entire P fragment plus a *KpnI*-*EcoRI* subfragment of *EcoRI*-M and part of *EcoRI*-Q. The nucleotide sequence is shown in Fig. 6, along with the amino acid sequence predicted for a large open reading frame found within. Interestingly, the repeat sequences lie completely within the open reading frame close to the amino terminus of the putative protein and code

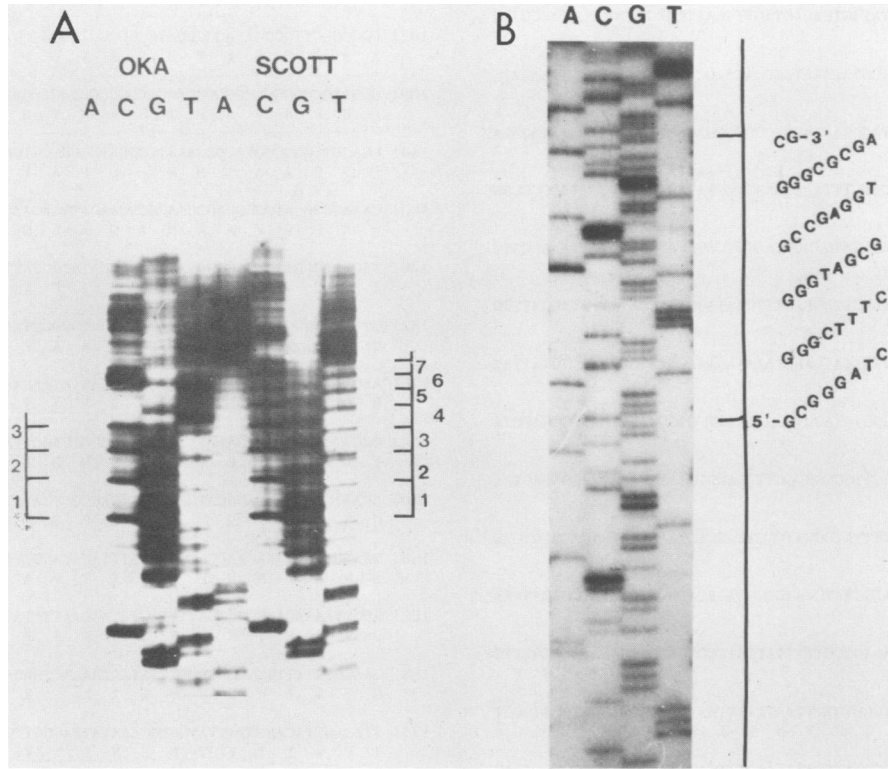


FIG. 4. Identification and sequencing of the repeats found in *EcoRI*-P from strains Oka and Scott. (A) The top of a sequencing gel in which the larger *HincII* fragments were cloned and sequenced from the unique *HincII* site. The number of repeats found in the clones of each strain is indicated. For the Oka clone, lane A did not fully extend to the top of the sequencing gel. (B) Part of an autoradiogram which displays the sequence of one of the repeats.

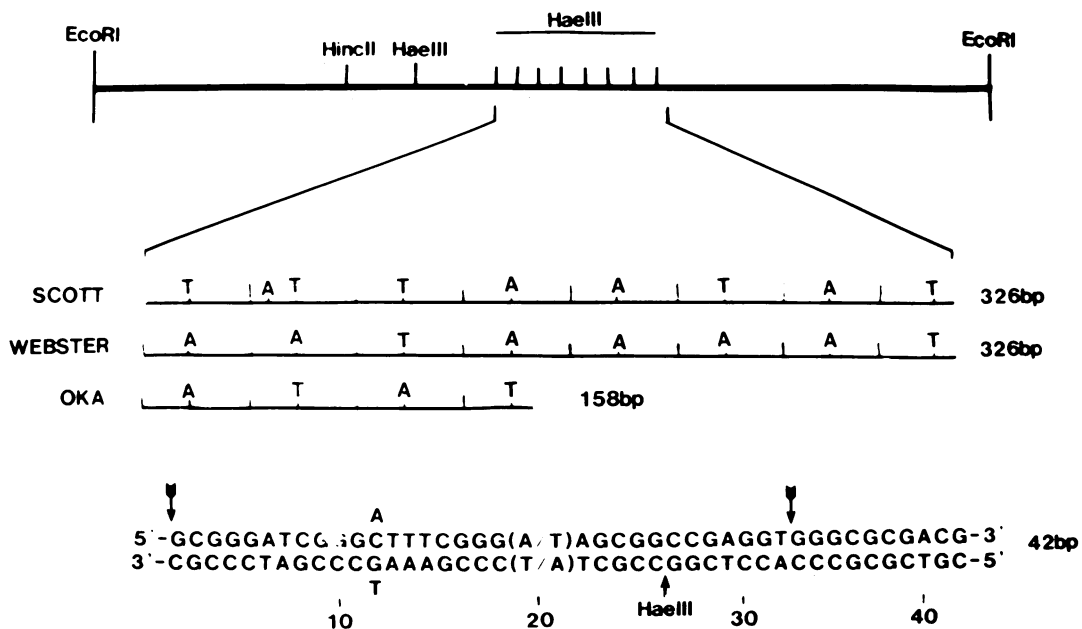


FIG. 5. Summary of the number and the microheterogeneity found in the repeats of *EcoRI*-P from strains Scott, Webster, and Oka. The location of the *HincII* and *HaeIII* sites discussed in the text is shown at the top. Below that is a graphic representation of the repeat region, showing the number of repeats, the length of the repeat region, and the identities of the variable bases found at positions 12 and 20, as discussed in the text. The double-stranded sequence of the repeat with the termini of the repeat region indicated by arrows above the sequence is shown at the bottom. The *HaeIII* site is also indicated below the sequence.

1 TTATTGGATTGTGTATATCTATTGTGCACACTAGTTTCAATTGTCOGATGCCGCCACCGATT
 61 TGCAAGATACGTGTATATATCGTAAAAATCGCCATAAGCAATTAACACTTATACAGATT
 121 TAGGTTTTGCCGTTGTATATACACAAAATGACCGGGGAGAGTATGTGACCATAGAGAAA
 Eco RI *****
 181 GTTCCCGGACCCCTAAACCGCGTGTAAAGGAATTCGTATAATGTCTGTTATACCCCGG

 241 TGTTATATATAGTTACCCCAATTAATGTGGCAATCTCACATATAAATAAATTAATCATT
 301 TTATCAAACTTACACAAGTAACGTTAGCAGTTTCAATAGGAGGTCATATTATAGCATTTG
 361 GGTTACAGGTTTTGCCGTTTTATATCAAGAAAAAAAACCTATGGGTAATTGTATTAT
 421 ATACAACGACCTCGGTGACGGGTATAGCTGCAACATTGCCGGCATTTCATGGGAGCTA

 481 TTATAAATCTAACATCAACAGTTCGGCGGGTTTTGAOGTGTATTAGATGATGAGACTAA
 541 GCGTTAAACCTATTGACTGTTTTATGGCATCTCATATCACTAAAAGTATATCACGTGTGTG

 601 TTTATATTATAATAAATCTATGCTATCTATGTGGTACATATGTATCGTAATCGAGATAAA
 661 TAAAGTTTTTAAAGTTGCAAAAGCCGTTTTTATTATCCCAATGTCGAAAAAAACGTTTC
 M S K K T F
 721 CATCATTAAATCCCGCGTGGGTGTTTTAATCTTTTATTAAAGGGAGCGTGGATGGGT
 P S F K F R G G C F N L L F K G S V D G
 781 CAATAAAAACCGAGTGAAGCGGATACAAAATAAATTTAATTTAACGATCGCGTGTATAC
 S I K T R M K R I Q I N L I L T I A C I
 841 AATTATCGACTGAATCTCAACCCACACCGTAAGTATAACTGAATTATATAGCTCGGCGG
 Q L S T E S Q P T P V S I T E L Y T S A
 901 CTACCCGAAAGCCCGATCCCGCGTCCGCGCCACCTCGGCGCGTTCCCGAAAGCCCGATC
 A T R K P D P A V A P T S A A S R K P D
 961 CCGCGTCCGCGCCACCTCGGCGCTACCCGAAAGCCGATCCCGCGTCCGCGCCACCT
 P A V A P T S A A T R K P D P A V A P T
 2101 GTATTATTGATGGACACATAGTATGTACGGCAAAATGTGTGCCACGTGGCGTTGTACATT
 G I I D G H I V C T A K C V P R G V V H
 2161 TCGTATGGTGGTTAACGATTCTCCCAATTAACCACGAAAACAGTGAGATTACTGGGGTGT
 F V W W V N V S P I N H E N S E I T G V
 2221 GTGATCAAAACAAACGGTTGTAAACATGCAAAATCTTGTCCAACATCGGAACTCGACG
 C D Q N K R F V N M Q S S C P T S E L D
 2281 GACATACACATATTCGTGTCATCTAGATGGTTACCCTAAAAAATCCCTCCGTTTTCCG
 G P I T Y S C H L D G Y P K K F P P P S
 2341 CCGTTTATACCTACGATGCACTACCTACGCCACTACATTTCCGTTGTAGCAGTTATAA
 A V Y T Y D A S T Y A T T F S V V A V I
 2401 TTGGTGTGATATCTATCCTTGGGACATTGGGCTTATCGCAGTTATCGCAACCCATGCA
 I G V I S I L G T L G L I A V I A T L C
 2461 TCCGTTGCTGTTCAATAACAGAAACCAACCAACCGCTGTGTATATCATTTTATTACA
 I R C C S -
 2521 TTCGAACACATCTACTGTCTTGACAACATTTAAAAATCCATTAAAGAGCCATTTCACATT
 2581 TTTAGGGGGGGTGTGGATTATATCCATCAAGCTGAAAAATCGTCCCAATTTAAAGTCGTTT
 2641 ATATCTGTTACATTTGCAATAATTTAAAGCAAGGAAAGGTTTTGGGATCGAGCTAGC
 2701 TGCACTTTTAAAGCATCTATATGATCAAGTAAATATGTGCATCCCCCATTTGTATGAATT
 2761 AAATCTCCGTTTTAAGTCTGTAACATGCGCTACTATGTAGTAAGAAGTCATATCCA
 Kpn I
 2821 GCAATGTTGAACGGTACC

-----r4-----r5
 1021 CGGCCGTTCCCGAAAGCCCGATCCCGCGTCCGCGCCACCTCGGCGGCTTCCCGAAAGC
 S A A S R K P D P A V A P T S A A S R K
 -----r6-----
 1081 CCGATCCCGCGTCCGCGCCACCTCGGCGCTACCCGAAAGCCCGATCCCGCGTCCGCGG
 P D P A V A P T S A A T R K P D P A V A
 -----r7-----
 1141 CCACCTCGGCGCTACCCGAAATCCCGATCCCGCGTCCGCGCCACCTCGGCGCTACCC
 P T S A A T R N P D P A V A P T S A A T
 -----r7.6-----
 1201 GAAAGCCCGATCCCGAGCCAAACGCCAATTCACAAACCTTTTCTATTGTAAATA
 P K P D P A A N A Q H S Q P P F L F E N
 1261 TACAATGCGTTCACGGGGAATACAATCCATACCTATTTCACACATTTATCATGCGTT
 I Q C V H G G I Q S I P Y F H T F I M P
 1321 GTTACATGCGTCTAACGACCGGACAACAGGCGGCTTTAAGCAGCAACAAAAACATAG
 C Y M R L T T G Q Q A A F K Q Q Q K T Y
 1381 AACAAATTTCTTAGATCCGGAAGGTTCAAAATAACAAGGTGGAAGTCGCTTATACGCC
 E Q Y S L D P E G S N I T R W K S L I R
 Hinc II
 1441 CCGATCTCATATGAAGTTGGTTACGCGTACCTTATAGATCCGACCGTCAACTGG
 P D L H I E V W F T R H L I D P H R Q L
 1501 GCAATGCGTTAATACGATGCCAGATTTACCGGTTATGTTATATAGCAACAGTCCCGATT
 G N A L I R M P D L P V M L Y S N S A D
 1561 TAACTTAATAAACAACCCCTGAGATATTTACACACCGCTAAGGAAAATTTATGTAATACCAG
 L N L I N N P E I F T H A K E N Y V I P
 1621 ATGTTAAACAACGCTGATTTTCTGTAAACATTTTATCTATGATGCTACCACGGAGG
 D V K T T S D F S V T I L S M D A T T E
 1681 GAACGTATATTTGGCGAGTCGTTAATACAAAACCTAAGAAGTCAATCGGAACACAGTA
 G T Y I W R V V N T K T K N V I S E H S
 1741 TTACAGTTACAACGTATTATCGTCCAAATATTACCGTTGTCGGGATCCAGTCTTAACCG
 I T V T T Y Y R P N I T V V G D P V L T
 1801 GACAGACATACGACGCTACTGTAAAGTATCAAGTATTATCCACCGCACTCGGTACGTTG
 G Q T Y A A Y C N V S K Y Y P P H S V R
 1861 TTCGGTGGACTCAAGGTTGGTAAACATCGGAAAAAATTTTATAACCGATGCAATACAAG
 V R W T S R F G N I G K N F I T D A I Q
 Eco RI
 1921 AATATGCCAATGGATTATTAGTTATGTTTCCGCGTACGAAATCCACAGCAAAAACAAA
 E Y A N G L F S Y V S A V R I P Q Q K Q
 1981 TGGATTACCCACCCAGCCATCCAATGTAATGTTTATGGAATCGGGATGCGGCTCTCTA
 M D Y P P P A I Q C N V L W I R D G V S
 2041 ATATGAAATATTCGTGTGTTACCCCTGACGCTATCCATTTCCCAACGTTGCTATAG
 N M K Y S A V V T P D V Y P F P N V S I

FIG. 6. Sequence of the VZV DNA region containing the entire open reading frame associated with *EcoRI*-P. The sequence is derived from the *EcoRI* P fragment and parts of the adjacent *EcoRI* Q and *EcoRI* M fragments from strain Scott. The identities of the predicted amino acid residues of the open reading frame are shown by single letters below the coding sequence. The two potential start sites for the protein are indicated by dots above the ATG codons. Potential glycosylation signals predicted from the amino acid sequence Asn-X-Ser/Thr are underlined. The locations of the repeat sequences are shown above the sequence. Restriction sites discussed in the text are also indicated above the sequence. Four of the potential TATA boxes and two likely polyadenylation signals are starred and underlined, respectively.

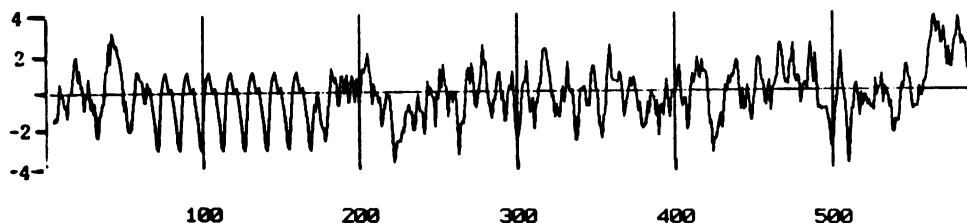


FIG. 7. Hydropathy plot of the protein predicted from the sequence coding for the open reading frame shown in Fig. 6. The numbers below the plot indicate the amino acid number of the reading frame. Positive values for the plot indicate hydrophobicity, and negative values indicate hydrophilicity.

for a repeating 14-amino-acid peptide Thr-Ser-Ala-Ala-Thr-Arg-Lys-Pro-Asp-Pro-Ala-Val-Ala-Pro (reading from amino to carboxy). The variation (A/T) in base 20 of the repeat described above causes a conserved amino acid change of threonine to serine, whereas the single base change in the second repeat of strain Scott results in the replacement of a lysine with an asparagine residue. The open reading frame has two potential ATG start sites, with the second lying 31 amino acids downstream from the first. Comparisons of these sequences with the nucleotides around the ATG start sites in eucaryotic genes, as shown by Kozak (20), suggest that the second ATG is more likely to be the functional start of the protein. In addition, there are two putative polyadenylation signals (ATTAAG) located downstream of the termination codon. The putative protein is 591 amino acids long and codes for a predicted protein of 66,000 molecular weight. There are five potential N-glycosylation sites within the open reading frame located at amino acids 236, 322, 351, 462, and 490, as determined by the predicted N-glycosylation signal sequence of Asn-X-Ser/Thr.

Further information on the predicted protein can be obtained from the hydropathy plot (Fig. 7). The protein has a short region of relatively nonhydrophobic residues at the

amino terminus followed by a highly hydrophobic region of some 13 or 14 amino acids, which could be a signal sequence for membrane-coupled protein synthesis. This hydrophobic region would constitute the amino terminus and be a likely signal sequence if initiation occurred at the second ATG codon. This arrangement of two possible ATG start sites was also predicted from the sequence of a glycoprotein found in the U_S region of VZV (6). In addition, there is a highly hydrophobic region consisting of some 30 amino acids located at the carboxy terminus of the protein. Taken together, these characteristics strongly suggest that the putative product of this open reading frame is a glycoprotein.

To identify a possible role for this putative glycoprotein in VZV infection, we compared the protein sequence with that of other herpesvirus proteins. Data presented by Davison and Wilkie (11) and A. J. Davison (personal communication) suggest that the genomic organization of VZV is very similar to that of HSV, except that the U_L region of VZV is inverted with respect to the prototype arrangement of HSV as defined by Roizman (29). Based on this, it was possible to demonstrate that the map location for this putative VZV glycoprotein corresponds approximately to the map location of glycoprotein gC in HSV-1. Thus, a direct comparison of the

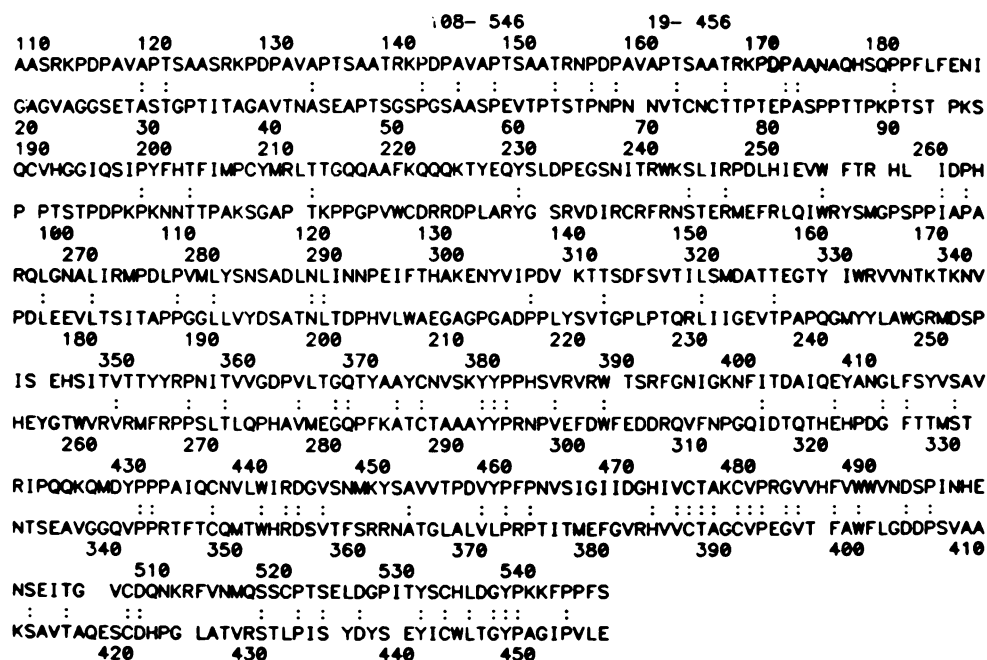


FIG. 8. Homologous amino acids predicted from parts of the potential VZV protein (top lines) and the amino acid sequence of HSV-1 gC (bottom lines; 15). The parts compared are indicated beneath each title. Gaps have been inserted to better show the homology. Residues that are identical in the two aligned sequences are indicated by two dots between the residues.

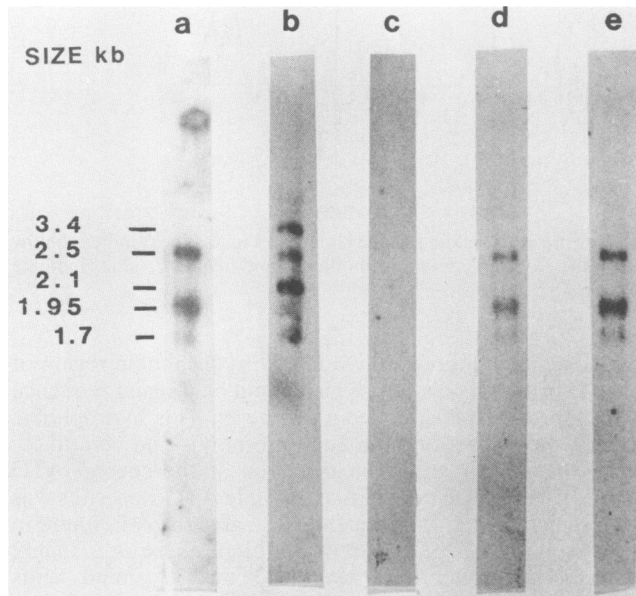


FIG. 9. Northern blot analysis of the *EcoRI*-P region. Northern blots of VZV RNA were probed with the double-stranded *EcoRI* P fragment (a), the double-stranded *HaeIII*-generated 42-bp repeat fragment (b), the sense-strand-labeled *EcoRI*-P (c), the missense-strand-labeled *EcoRI*-P (e), or the missense-strand-labeled 1,150-base *EcoRI*-*HincII* fragment of *EcoRI*-P (d). The approximate sizes of the RNA species are shown.

amino acid sequences of HSV-1 gC (15) and the putative VZV glycoprotein was made using the SEQHP program method (16). The locally homologous amino acids of parts of the two proteins are shown in Fig. 8 (gaps are inserted to better show the homology). Comparison of 439 amino acids of the potential VZV open reading frame (from amino acids 108 to 546) with 438 amino acids (19 to 446) of the HSV-1 gC showed that 92 amino acids (21%) are perfectly matched between the two proteins. Small regions show a much higher degree of homology, particularly towards the carboxy termini of the proteins. For example, there is a 15-amino-acid stretch in VZV (from 471 to 485) that has over 70% homology to HSV-1 gC. When the probability that so close a resemblance would occur by chance alone was calculated and used to classify the homologies by statistical significance (16), computer analysis indicated that the homology shown in Fig. 8 is greater than 26 standard deviations from the average matching of 100 random sequences, suggesting that the homology is highly significant.

Analysis of the transcription from the *EcoRI*-P region. We showed with the above data that the VZV sequences incorporating the *EcoRI*-P variable region are, in theory, capable of coding for a glycoprotein. We therefore examined transcription from the *EcoRI*-P region in vivo to determine whether this was an active region of VZV DNA. RNA was prepared from cells infected with VZV strain Ellen and transferred to nitrocellulose as previously described (25). We used single-stranded probes generated from this region of VZV DNA which allowed the determination of the direction of transcription of these RNA species. The results are shown in Fig. 9. Double-stranded DNA of *EcoRI*-P was found to hybridize to at least three transcripts of 1.7, 1.95, and 2.5 kilobases (kb) (Fig. 9, lane a). Probes of *EcoRI*-P DNA labeled only on the same (sense) strand as the open reading frame did not hybridize to any RNA species detect-

able on Northern blots (lane c). However, missense strand-labeled *EcoRI*-P DNA hybridized to at least three RNA species of 1.95 and 2.5 kb, with a minor species of 1.7 kb (lane e). The RNA species also hybridized missense strand-labeled DNA from the large *HincII*-*EcoRI* fragment of *EcoRI*-P (lane d). These results suggest that the large open reading frame described above is actually transcribed in vivo into at least two or possibly three RNA species. We also used the 42-bp repeat sequence to probe Northern blots and found that each of the three RNA species hybridized this probe. However, two other RNA species of 2.1 and 3.4 kb were also detected with the repeat probe, probably indicating nonspecific homology to other RNA species that also contain high G+C regions. We are now attempting to determine whether the three transcripts of 1.7, 1.95 and 2.5 kb are different processed forms of the same transcribed region and whether such transcripts are translated.

DISCUSSION

The results presented here identify the structural basis for a variation that is observed in the DNAs of different VZV strains. The presence of the variable region, located in the U_L region of the genome approximately 18 kilobase pairs from the left-hand end, results in a characteristic mobility difference in agarose gels of a DNA restriction fragment (*EcoRI*-P) that is due to the presence of different copy numbers of a repeating nucleotide sequence. The difference in mobility between the largest and the smallest (strains Scott and Oka, respectively) was originally estimated to be approximately 150 to 200 bp. Detailed heteroduplex analyses and nucleotide sequencing confirmed these early observations, and it is now clear that Oka and Scott differ in this region by 168 nucleotides (four complete repeat units).

The repeating sequences are composed of head-to-tail 42-bp units of 5'-GCGGGATCGGGCTTTCGGG(A/T)AGC-GGCCGAGGTGGGCGCAGC-3'. At the 3' terminus of each group of repeats, an incomplete sequence (32 nucleotides) is invariably found. The unit sequence is unusual for VZV, in that its 77% G+C content is much higher than the average for the whole genome (47% G+C). In several respects, this sequence is similar to repeat sequences found elsewhere in the genome of VZV and in other herpesviruses. We previously described 27-bp head-to-tail repeat units located in the IR_S/TR_S region of VZV whose copy numbers vary from strain to strain (4). The 27-bp sequence has only limited nucleotide homology with the 42-bp sequence, but it is similar in that it is also high in G+C (81%; 4, 9). Repeated high G+C sequences were also found in the IR_S/TR_S regions of other herpesviruses, including the *a* and *c* sequences of HSV and pseudorabies virus (2, 23, 24). The function of such repeated elements remains obscure, although their location outside yet close to transcribed sequences may suggest a role in the control of gene expression.

A major difference between the 42-bp repeat and those described above is that the 42-bp repeat seems to form part of the coding sequence for a protein(s). It should, however, be noted that the location of the 42-bp repeat does not exclude the possibility that it also may play a role in transcriptional control, analogous to certain transcriptional enhancer sequences encountered in other eucaryotic genomes (21). Intragenic repeat elements have been reported in a DNA-binding protein gene of HSV-1, but their role or significance is not known (28); they also show strain variation.

Expression of the putative glycoprotein gene in strains Oka and Scott should lead to the production of two proteins

which differ in size by 56 amino acids or about 10% of their sequence. Without knowledge of the polypeptide structures, it is difficult to predict how this might affect the activity of the proteins, and the parameters affected have yet to be determined.

Several potential TATA-box like sequences can be identified upstream from the open reading frame with characteristics of other herpesvirus promoters. This is presumably because of the relatively high A+T content of the virus genome; it is difficult to predict which may be used by the virus. Downstream of the termination codon lie two predicted polyadenylation signal sequences.

It is clear that the region is transcriptionally active in VZV-infected cells, coding for transcripts of 2.5, 1.95, and 1.7 kb, all read from the missense strand. However, we are not yet certain whether these transcripts are translatable or whether some of the species may be spliced or processed. It would be valuable to determine whether transcription in vivo from this region is early or late, allowing detailed comparisons of the controlling (upstream) regions of HSV with VZV.

The glycoprotein predicted to result from translation of these transcripts has not, as yet, been identified in VZV-infected cells. To continue these studies, it is likely that it will be necessary to obtain an immunological reagent capable of recognizing this protein, and this is being pursued. From the sequence, the polypeptide core of the protein has a molecular weight of about 66,000, but addition of sugars is likely to increase that figure substantially. Comparison of its amino acid sequence with that of HSV-1 gC showed homology between the two proteins, particularly in a region located toward the carboxy termini of the proteins. In contrast, there is less than 20% overall homology outside this region; this may suggest that an important part of these two molecules lies in these conserved regions.

It is not yet known what function gC has in HSV-1 infection. Furthermore, it was shown that HSV-1 gC is not essential for virus infection, since several virus isolates were found that lack this glycoprotein (5, 18). The HSV-2 counterpart of gC was also reported to be nonessential (41). We are now attempting to determine whether this putative VZV gene is also nonessential.

Finally, it may be of significance that strain Oka, with the smaller gene described in this paper, is the attenuated vaccine strain of VZV (38), whereas the other VZV strains analyzed have been passed relatively few times outside human cell culture. Strain Oka was reported to have many irregularities in its DNA (40), although none of the reported differences between the parent and vaccine strains corresponds to the variable region reported here. It will therefore be important to determine whether the number of copies of the 42-bp repeat in the parent Oka strain is the same as in the vaccine strain. It may be that the characteristics of this VZV glycoprotein are an important determining factor in the pathogenesis of varicella or zoster infection.

ACKNOWLEDGMENTS

We thank M. N. Flora and M. Wellman for technical assistance, J. Maizel and D. Chatderjee for help in the comparisons of amino acid sequences, and J. Weir and R. Peterson and for helpful discussions. We also thank A. J. Davison for communication of unpublished results.

This work was supported by Public Health Service grant AI18449 from the National Institute of Allergy and Infectious Diseases (to W.T.R. and J.H.).

LITERATURE CITED

1. Allen, G. P., M. R. Yeargen, and J. T. Bryans. 1983. Alterations in the equine herpesvirus type 1 genome after in vitro and in vivo virus passage. *Infect. Immun.* **40**:436-439.
2. Ben-Porat, T., A. Deatly, R. A. Veach, and M. L. Blankenship. 1984. Equalization of the inverted repeat sequences of the pseudorabies virus genome by intermolecular recombination. *Virology* **132**:304-314.
3. Buchmann, T. G., B. Roizman, G. Adams, and B. H. Stover. 1978. Restriction endonuclease fingerprinting of herpes simplex virus DNA: a novel epidemiological tool applied to a nosocomial outbreak. *J. Infect. Dis.* **138**:488-498.
4. Casey, T. A., W. T. Ruyechan, M. N. Flora, W. Reinhold, S. E. Straus, and J. Hay. 1985. Fine mapping and sequencing of a variable segment of the inverted repeat region of varicella-zoster virus DNA. *J. Virol.* **54**:639-642.
5. Cassai, E., R. Manservigi, A. Corallini, and M. Terni. 1976. Plaque dissociation of herpes simplex viruses. Biochemical and biological character of the viral variants. *Intervirology* **6**:212-223.
6. Davison, A. J. 1983. DNA sequence of the U_S component of the varicella zoster genome. *EMBO J.* **2**:2203-2209.
7. Davison, A. J. 1984. Structure of the genome termini of varicella zoster virus. *J. Gen. Virol.* **65**:1969-1977.
8. Davison, A. J., and J. E. Scott. 1983. Molecular cloning of the varicella zoster virus genome and derivation of six restriction endonuclease maps. *J. Gen. Virol.* **64**:207-220.
9. Davison, A. J., and J. E. Scott. 1985. DNA sequence of the major inverted repeat in the varicella zoster virus genome. *J. Gen. Virol.* **66**:207-220.
10. Davison, A. J., and N. M. Wilkie. 1981. Nucleotide sequences of the joint between the L and S of herpes simplex viruses types 1 and 2. *J. Gen. Virol.* **55**:315-331.
11. Davison, A. J., and N. M. Wilkie. 1983. Location and orientation of homologous sequences in the genomes of five herpesviruses. *J. Gen. Virol.* **64**:1927-1942.
12. Desrosiers, R. C., and L. A. Falk. 1982. *Herpesvirus saimiri* strain variability. *J. Virol.* **43**:352-356.
13. Dumas, A. M., J. L. M. C. Geelen, M. W. Westrate, P. Wertheim, and J. Van Der Noordaa. 1981. *Xba*I, *Pst*I, and *Bgl*III restriction enzyme maps of the two orientations of the varicella-zoster virus genome. *J. Virol.* **39**:390-400.
14. Ecker, J. R., and R. W. Hyman. 1982. Varicella zoster virus exists as two isomers. *Proc. Natl. Acad. Sci. USA* **79**:993-997.
15. Frink, R. J., R. Eisenberg, G. Cohen, and E. K. Wagner. 1983. Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. *J. Virol.* **45**:634-647.
16. Goad, W. B., and M. I. Kanechisa. 1982. Pattern recognition in nucleic acid sequences. 1. A general method for finding local homologies and symmetries. *Nucl. Acids Res.* **10**:247-263.
17. Halliburton, I. W., L. S. Morse, B. Roizman, and K. E. Quinn. 1980. Mapping of the thymidine kinase genes of types 1 and 2 herpes simplex viruses using intertypic recombinants. *J. Gen. Virol.* **49**:235-253.
18. Heine, J. W., R. W. Honess, E. Cassai, and B. Roizman. 1974. Proteins specified by herpes simplex virus. XII. The virion polypeptides of type one strains. *J. Virol.* **14**:640-651.
19. Kinchington, P. R., W. C. Reinhold, T. A. Casey, S. E. Straus, J. Hay, and W. T. Ruyechan. 1985. Inversion and circularization of the varicella-zoster virus genome. *J. Virol.* **56**:194-200.
20. Kozak, M. 1983. Comparison of the initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47**:1-45.
21. Laimins, L. A., M. Kessel, N. Rosenthal, and G. Khoury. 1983. Viral and cellular enhancer elements, p. 28-37. *In* Y. Gluzman and T. Shenk (ed.), *Current communications in molecular biology*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. Langridge, J., P. Langridge, and P. L. Bergquist. 1980. Extraction of nucleic acids from agarose gels. *Anal. Biochem.* **103**:264-271.
23. Mocarski, E. S., and B. Roizman. 1982. Structure and role of the herpes simplex virus DNA termini in inversion, circularization

- and generation of virion DNA. *Cell* 31:89-97.
24. Murchie, M. J., and D. J. McGeoch. 1982. DNA sequence analysis of an immediate early gene region of the herpes simplex type 1 genome (map coordinates 0.950 to 0.978). *J. Gen. Virol.* 62:1-15.
 25. Ostrove, J. M., W. Reinhold, C.-M. Fan, S. Zorn, J. Hay, and S. E. Straus. 1985. Transcriptional mapping of the varicella-zoster virus genome. *J. Virol.* 56:600-606.
 26. Poncz, M., O. Solowiejczyk, M. Ballantine, E. Schwartz, and S. Surrey. 1982. Non-random DNA sequence analysis in bacteriophage M13 by the dideoxy chain termination method. *Proc. Natl. Acad. Sci. USA* 79:4298-4302.
 27. Raab-Traub, N., R. Pritchett, and E. Kieff. 1978. DNA of Epstein-Barr virus. III. Identification of restriction enzyme fragments that contain DNA sequences which differ among strains of Epstein-Barr virus. *J. Virol.* 27:388-398.
 28. Rixon, F. J., and D. J. McGeoch. 1984. A 3' coterminal 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 Res.* 12:2473-2487.
 29. Roizman, B. 1979. The structure and isomerization of the herpes simplex virus genomes. *Cell* 16:481-494.
 30. Ruyechan, W. T., T. A. Casey, W. Reinhold, A. C. Weir, M. Wellman, S. E. Straus, and J. Hay. 1985. Distribution of G+C regions in varicella zoster virus DNA. *J. Gen. Virol.* 66:43-54.
 31. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161-178.
 32. Skare, J., W. P. Summers, and W. C. Summers. 1975. Structure and function of herpesvirus genomes. I. Comparison of five HSV-1 and two HSV-2 strains by cleavage of their DNA with *EcoR* I [sic] restriction endonuclease. *J. Virol.* 15:726-732.
 33. Spear, P. G., and B. Roizman. 1980. Herpes simplex viruses. *In* J. Tooze (ed.), *Molecular biology of tumour viruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 34. Straus, S. E., H. S. Aulakh, W. T. Ruyechan, J. Hay, T. A. Casey, G. F. Vande Woude, J. Owens, and H. A. Smith. 1981. Structure of varicella-zoster virus DNA. *J. Virol.* 40:516-525.
 35. Straus, S. E., J. Hay, H. A. Smith, and J. Owens. 1983. Genome differences among varicella zoster virus isolates. *J. Gen. Virol.* 64:1031-1041.
 36. Straus, S. E., J. Owens, W. T. Ruyechan, H. E. Takiff, T. A. Casey, G. F. Van deWoude, and J. Hay. 1982. Molecular cloning and physical mapping of varicella zoster virus DNA. *Proc. Natl. Acad. Sci. USA* 79:993-997.
 37. Straus, S. E., W. Reinhold, H. A. Smith, W. T. Ruyechan, D. K. Henderson, R. M. Blaese, and J. Hay. 1984. Endonuclease analysis of viral DNAs from varicella and subsequent zoster infections in the same patient. *N. Engl. J. Med.* 311:1362-1364.
 38. Takahashi, M., Y. Okuno, T. Otsuka, J. Osame, A. Takamizawa, T. Sasada, and T. Kubo. 1975. Development of a live attenuated varicella vaccine. *Biken J.* 18:25-33.
 39. Veira, J., and J. Messing. 1982. The pUC plasmid, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19:259-268.
 40. Vlazny, D. A., and R. W. Hyman. 1985. Errant processing and structural alterations of genomes present in a varicella-zoster virus vaccine. *J. Virol.* 56:92-101.
 41. Zezulak, K. M., and P. G. Spear. 1984. Mapping of the structural gene for the herpes simplex virus type 2 counterpart of herpes simplex virus type 1 glycoprotein C and identification of a type 2 mutant which does not express this glycoprotein. *J. Virol.* 49:741-747.
 42. Zweerink, H. J., D. H. Morton, L. W. Stanton, and B. J. Neff. 1981. Restriction endonuclease analysis of the DNA from varicella zoster virus; stability of the DNA after passage *in vitro*. *J. Gen. Virol.* 55:207-211.