Fine Mapping and Sequencing of a Variable Segment in the Inverted Repeat Region of Varicella-Zoster Virus DNA

THOMAS A. CASEY,¹ WILLIAM T. RUYECHAN,² MICHAEL N. FLORA,¹ WILLIAM REINHOLD,³ STEPHEN E. STRAUS,³ AND JOHN HAY^{1*}

Departments of Microbiology¹ and Biochemistry,² Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799, and Laboratory of Clinical Investigation, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20205³

Received 24 September 1984/Accepted 11 January 1985

A strain variation in the internal and terminal repeats which bind the short unique sequence of varicella-zoster virus (VZV) DNA was found to be due to an insertion or deletion of DNA sequences at a single site. DNA sequence analysis showed that the nucleotide sequence CCGCCGATGGGGAGGGGGCGCGGT ACC is tandemly duplicated a variable number of times in different VZV strains and is responsible for the observed variation in mobilities of restriction fragments from this region of VZV DNA. The variable region sequence shares some homology with tandemly repeated regions in the a and c sequences of herpes simplex virus type 1 and probably exists in a noncoding region of the VZV genome.

We have recently reported that varicella-zoster viruses (VZV) isolated from lesions of patients with varicella or zoster infections possess characteristic DNA restriction digestion profiles which permit differentiation of VZV strains (15). All viruses isolated from patients whose infection was acquired in different outbreaks could be distinguished from each other on this basis, whereas viruses from single, small outbreaks appeared identical (15; S. E. Straus, W. T. Ruyechan, and J. Hay, unpublished observations). In addition, this restriction endonuclease test has been used to show that viruses isolated from varicella and subsequent zoster infections in the same patient are also identical (17). Such strain differences in VZV DNA seem to be quite stable to passage in human cell culture (15, 18) and to at least limited passage in people (15). We have hypothesized that these strain variations arise from the insertion or deletion of DNA sequences at specific sites in VZV DNA.

Analogous strain differences in other herpesvirus DNAs have been reported (e.g., see references 1, 5, 9, and 14) and have been used as stable markers in molecular epidemiology studies (3, 10). The present study describes the detailed sequence analysis of one region of the genome in which strain differences have been observed.

Figure 1 shows the basic structure of VZV DNA. There are two unique sequences, U_L and U_S , and one set of inverted repeats, TR_S and IR_S (6, 7, 16). Recently, it has been suggested that both the U_L and U_S sequences may be capable of inversion, the former at low frequency (14a; A. Davison, personal communication). Regions of DNA strain difference are also identified in Fig. 1, on the maps, as specific fragments from restriction endonuclease digestions of whole VZV DNA (15). These differences occur in U_L and in the repeat (IR_S and TR_S) regions (15).

Based on the hypothesis that the DNA strain difference in the repeat region (Fig. 1) was the result of a single insertion or deletion and that it amounted to a maximum of about 250 base pairs (bp) (15), it was decided that heteroduplex analyses in the electron microscope would provide useful data on the location and, perhaps, on the size of the difference.

The *EcoRI* E fragment of VZV DNA, which contains the

repeat region strain difference (Fig. 1), was cloned into pBR325 by using DNA from a number of different VZV strains. Cloned DNA preparations were cut with *EcoRI*, and the E fragments from four strains were isolated, after electrophoresis, from 1% low-melting-point agarose gels by the method of Langridge et al. (8). Equal amounts of purified *EcoRI* E fragments derived from VZV Ellen and VZV Scott DNA were mixed with recrystallized formamide to yield final DNA and formamide concentrations of 1 μ g/ml and 80% (vol/vol), respectively, in a 100- μ l volume. The DNA was denatured by heating the mixture to 70°C for 5 min. A total of 50 μ l of 0.2 M Tris hydrochloride–0.02 M EDTA (pH 8.5) was added, and reannealing was allowed to proceed for 24 h at room temperature. After renaturation, 40 μ l of 0.1-mg/ml cytochrome *c* and 0.01 μ g each of simian virus 40 form II and fd DNAs were added. The heteroduplexes were mounted for electron microscopy either by standard formamide spreading or by touching a Parlodion-coated grid to a drop of the heteroduplex mixture which had been placed on a clean microscope slide. The grids were stained with uranyl acetate and observed in a Zeiss EM 10A transmission electron microscope. Heteroduplexes were photographed at a magnification of $\times 7,690$. The results are given in Fig. 2 and 3.

Electron micrographs of three heteroduplex DNA molecules (strains Ellen and Scott) are shown in Fig. 2. In each, two features are evident; first, there is a single-stranded (open or collapsed) loop present close to one terminus of the molecule, and second, the remainder of the molecule appears to be duplex DNA. Twenty-one such heteroduplexes were examined, and the location of the single-stranded loop was estimated to occur, on average, $80 \pm 4\%$ of the distance from one terminus of the *EcoRI* E fragment (Fig. 3). It was difficult to measure accurately the size of the single-stranded loop in these heteroduplexes, but it appeared to be of approximately the same magnitude (100 to 200 bp) as the strain difference (Ellen versus Scott) estimated previously from gel electrophoresis (15). These data suggest that the *EcoRI*-E strain difference is a single insertion or deletion close to one terminus of the fragment.

To orient this strain difference with respect to the termini of *EcoRI*-E and to obtain a better assessment of the size of

* Corresponding author.

FIG. 1. Locations of the strain-variable fragments on the *EcoRI*, *BamHI*, *SmaI*, *BglII*, *HindIII*, and *PstI* restriction maps of VZV DNA. The map positions of the VZV DNA restriction endonuclease fragments which have altered mobilities in different VZV strains are marked in black. The fragments which are hatched may also vary but are either too large for this small change to be seen or too small to be resolved in the gels. At the bottom, the three regions of VZV which vary are marked on model of the genomic organization of VZV DNA. There are two variable regions in the long unique segment (U_L), and the third variable region is present two times, once in each of the inverted repeats (IR_S and TR_S) which bound the short unique segment (U_S). The maps are redrawn from Dumas et al. (6) for *PstI* and *BglII*; from Ecker and Hyman (7) for *HindIII*; from Straus et al. (15, 16) for *EcoRI* and *BamHI*; and from Ruyechan et al. (14a) for *SmaI*. Mdal, Megadalton.

FIG. 3. Summary of heteroduplex molecules formed between VZV Scott and Ellen *EcoRI* E fragments. The position of the insertion or deletion loop is plotted as the fractional distance from one end of the molecule versus the number of molecules with a loop at that position. A total of 21 heteroduplexes are shown in this figure. The mean fractional position of the difference is also shown. At the top of this figure is a simplified restriction endonuclease map of the *EcoRI* E fragment which shows the positions of the *PstI* site, the *BglI* sites, and the *SalI* sites in this fragment.

the difference, *EcoRI* E fragments cloned in pBR325 were subjected to restriction enzyme digestion with a variety of enzymes. An example is given in Fig. 4, in which four different *EcoRI*-E clones, from VZV Scott, Ellen, Webster, and Oka, digested with *BglI* are shown.

It is clear from the digests that the strain differences are all localized to one fragment, *BglI*-K, which maps in the right half of *IR_S*, toward the *IR_S*-*U_S* junction (Fig. 3). An analogous digestion with *SalI* (not shown) further localized the strain differences to the *SalI* Q fragment (Fig. 3). From these digests, estimates of the sizes of the *EcoRI* E fragments from different strains can be made. Ellen E is approximately the same size as Webster E, whereas Scott E is 100 and Oka E is 200 bp larger.

As a means of determining the precise nature of these VZV DNA strain differences, segments of the *EcoRI* E fragments from VZV Ellen and Oka were cut with *SalI*, and the *SalI* Q fragments were ³²P end labeled with polynucleotide kinase before being subjected to Maxam-Gilbert (11) sequencing procedures.

Results of these sequence studies indicated that the *SalI* Q fragment from strain Ellen contained approximately 6 copies of a 27-bp sequence, while strain Oka contained about 15 copies of the same sequence. The sequence is *U_L*—

FIG. 4. *BglI* and *BglI*-*EcoRI* digestion of the *EcoRI* fragments from four VZV strains. Recombinant plasmid DNA containing the *EcoRI* fragment from VZV strain Ellen (E), Scott (S), Webster (W), or Oka (O) was digested with *BglI* or *BglI* and *EcoRI*, and the bands were separated by electrophoresis in a 1.5% agarose gel at 50 V for 18 h. The gel was stained with ethidium bromide and photographed with UV light. *BglI*-*EcoRI*-digested pBR325 DNA (P) was used as a control. Lanes 1 and 2 contain adenovirus type 2 DNA digested with *EcoRI* and *HindIII* as molecular weight markers. Kb, Kilobase.

(CCGCCGATGGGGAGGGGGCGCGGTACC)_n—*U_S* and occurs in *SalI*-Q, approximately 20% from the right-hand terminus of the fragment (Fig. 3); *n* appears to be not necessarily a whole number.

The strain differences in VZV DNA which we have described above appear to have close parallels in the DNA of equine herpes virus type 1, at least in terms of their location on the genome (1). Equine herpes virus type 1, like VZV, possesses a genome which contains a single set of inverted repeat sequences, and the two viruses are somewhat related at the DNA sequence level.

Several other herpesviruses contain, in their inverted repeat sequences which bound a short unique region, DNA sequences which vary in length depending on the strain. Examples of these are pseudorabies virus (2) and herpes simplex virus type 1 (HSV-1) (13), in addition to equine herpes virus type 1 and VZV. The nature of a repeat sequence strain variation has been investigated in only one other instance. In HSV-1, multiple repeats of a short DNA sequence in the c region exist upstream from coding sequences for a small polypeptide (4, 13). It is hypothesized that variable copy numbers of this short sequence (22 bp) are responsible for HSV-1 DNA strain variation in that region.

In the present study, we found an analogous situation in VZV DNA and were able, by analysis of several strains, to confirm the hypothesis that strain variation in the repeat

FIG. 2. Electron micrographs of heteroduplex molecules formed between the *EcoRI* E fragments from VZV Scott and Ellen. Plasmids containing the *EcoRI* E fragment from strains Ellen and Scott were digested with *EcoRI*, and the bands were separated in 1% low-melting-temperature agarose gels. The *EcoRI* E bands were cut out and extracted from the gels (8). Equal amounts of each DNA were mixed, denatured, and allowed to reanneal before being spread and mounted for examination in the electron microscope. Shown are three representative heteroduplex molecules (A, B, C). There is a single-stranded insertion or deletion loop near one end of each molecule (arrows) which is the strain difference. The remainder of each of the molecules is double stranded. Bar, 1 μ m.

FIG. 5. Tandemly repeated DNA sequence which is responsible for variability in the mobilities of restriction fragments that contain VZV IR_S or TR_S is compared with the sequences found in the a and c regions of HSV-1 DNA. The relative positions of these sequences in the IR_S and TR_S of VZV and HSV-1 are also shown.

region is due to variable quantities of a short DNA sequence. Davison (personal communication) has recently observed the presence of five to six copies of this 27-bp sequence in one strain of VZV (Dumas). Examination of the VZV sequence responsible for strain variation and comparison with the parts of the HSV-1 a and c sequences in the short repeat regions (12, 13) reveals some interesting similarities (Fig. 5).

Several points arise. First, they are all strikingly similar in overall G+C content. The HSV-1 sequences are 86.4% (c) and 83.3% (a) G+C, while the VZV sequence is 81.4% G+C. In contrast, the overall base composition of these two viral genomes is 67 and 48%, respectively. At a more detailed level, both HSV repeat sequences contain the following stretch of purines: GAGGAGGGGG; a corresponding sequence, GGGGAGGGGG, identical to the HSV sequence except for the A/G switch at position 2, occurs in the center of the VZV repeat. Alignment of these two purine stretches results in an 18-base overlap between the HSV c and VZV sequences (Fig. 5), which have a base homology of 67%. The overall homology between the HSV and VZV genomes has been estimated to be only a few percent. It would seem likely that such an unusual and apparently highly conserved region of DNA has a role to play in the life cycle of these viruses. At present, however, what this role might be is unclear.

In HSV-1, these sequences appear to occur between or outside coding regions of the DNA and may be controlling factors for transcription (13). Preliminary evidence from this laboratory (J. Ostrove, personal communication) and others (A. Davison, personal communication) suggests that these variable sequences in VZV have a similar role in that they lie between the 5' termini of two mRNAs.

Finally, preliminary data suggest that the other two variable regions in VZV DNA (both in U_L; Fig. 1) are also the result of variable copy numbers of repeated elements, but their sequences appear not to be identical to those of the variable region analyzed here. Whether these other variable regions occur within or between coding sequences is being actively investigated.

We wish to thank Andrew Davison for making his data available to us before publication.

This work was supported in part by Public Health Service grant AI18449 from the National Institutes for Allergy and Infectious Diseases to W. T. R. and J. H.

LITERATURE CITED

- Allen, G. P., M. R. Yeargan, and J. T. Bryans. 1983. Alterations in the equine herpesvirus 1 genome after in vitro and in vivo passage. *Infect. Immun.* **40**:436-439.
- 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.
- Buchman, 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.
- Davison, A. J., and N. M. Wilkie. 1981. Nucleotide sequences of the joint between the L and S segments of herpes simplex type I and II. *J. Gen. Virol.* **55**:315-331.
- Desrosiers, R. C., and L. A. Falk. 1982. *Herpesvirus saimiri* strain variability. *J. Virol.* **43**:352-356.
- Dumas, A. M., J. L. M. C. Geelen, M. W. Weststrate, P. Wertheim, and J. van der Noordaa. 1981. *Xba*I, *Pst*I, and *Bgl*II restriction enzyme maps of the two orientations of the varicella-zoster virus genome. *J. Virol.* **39**:390-400.
- Ecker, J. R., and R. W. Hyman. 1982. Varicella zoster virus DNA exists as two isomers. *Proc. Natl. Acad. Sci. U.S.A.* **79**:156-160.
- Langridge, J., P. Langridge, and P. L. Bergquist. 1980. Extraction of nucleic acids from agarose gels. *Anal. Biochem.* **103**:264-271.
- Locker, H., and N. Frenkel. 1979. *Bam*I, *Kpn*I, and *Sal*I restriction enzyme maps of the DNAs of herpes simplex virus strains Justin and F: occurrence of heterogeneities in defined regions of the viral DNA. *J. Virol.* **32**:429-441.
- Lonsdale, D. M., S. M. Brown, J. H. Subak-Sharpe, K. P. Warren, and H. Koprowski. 1979. The polypeptide and DNA restriction enzyme profiles of spontaneous isolates of herpes simplex type I from explants of human trigeminal, superior, cervical and vagus ganglia. *J. Gen. Virol.* **43**:151-171.
- Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**:560-564.
- 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.
- Murchie, M. J., and D. J. McGeoch. 1982. DNA sequence analysis of an immediate-early gene region of the herpes simplex virus type I genome (map coordinates 0.950 to 0.978). *J. Gen. Virol.* **62**:1-15.
- 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.
- Ruyechan, W. T., T. A. Casey, W. Reinhold, A. C. Weir, M. Wellman, S. E. Straus, and J. Hay. 1985. Distribution of G+C rich regions in varicella-zoster virus DNA. *J. Gen. Virol.* **66**:43-54.
- Straus, S. E., J. Hay, H. Smith, and J. Owens. 1983. Genome differences among varicella-zoster virus isolates. *J. Gen. Virol.* **64**:1031-1041.
- Straus, S. E., J. Owens, W. T. Ruyechan, H. E. Takiff, T. A. Casey, G. F. VandeWoude, and J. Hay. 1982. Molecular cloning and physical mapping of varicella-zoster virus DNA. *Proc. Natl. Acad. Sci. U.S.A.* **79**:993-997.
- 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 DNA's from varicella and subsequent zoster infections in the same patient. *N. Engl. J. Med.* **311**:1362-1364.
- 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.