Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 6817–6821, November 1984
Genetics

Spontaneous deletions and duplications of sequences in the genome of cowpox virus

(orthopoxvirus/white pock variants/rearrangement/recombination)

DAVID J. PICKUP, BARBARA S. INK, BARBARA L. PARSONS, WENSI HU, AND WOLFGANG K. JOKLIK

Department of Microbiology and Immunology, Duke University Medical Center, Durham, NC 27710

Contributed by Wolfgang K. Joklik, July 23, 1984

ABSTRACT **Examination of the genomes of 10 white-pock variants of cowpox virus strain Brighton red (CPV-BR) revealed that 9 of them had lost 32 to 38 kilobase pairs (kbp) from their right-hand ends and that the deleted sequences had been replaced by inverted copies of regions from 21 to 50 kbp long from the left-hand end of the genome. These variants thus possess inverted terminal repeats (ITRs) from 21 to 50 kbp long; all are longer than the ITRs of CPV-BR (10 kbp). The 10th variant is a simple deletion mutant that has lost the sequences between 32 and 12 kbp from the right-hand end of the genome. The limits of the inner ends of the observed deletions (between 32 and 38 kbp from the right-hand end of the CPV-BR genome) appear to be defined by the location of the nearest essential gene on the one hand and the location of the gene that encodes "pock redness" on the other. The genomes of the deletion/duplication white-pock variants appear to have been generated either by single crossover recombinational events between two CPV-BR genomes aligned in opposite directions or by the nonreciprocal transfer of genetic information. The sites where such recombination/transfer occurred were sequenced in four variants. In all of them, the sequences adjacent to such sites show no sequence homology or any other unusual strutural feature. The analogous sites at the internal ends of the two ITRs of CPV-BR also were sequenced and also show no unusual features. It is likely that the ITRs of CPV-BR and of its white-pock variants, and probably those of other orthopoxvirus genomes, arise as a result of nonhomologous recombination or by random nonreciprocal transfer of genetic information.**

The genomes of orthopoxviruses, which are about 200 kilobase pairs (kbp) long, comprise three distinct regions: a highly conserved central region that accounts for about half of the genome and two roughly similar-sized flanking regions (R1 and R2) that have diverged much more (1). At the termini of most but not all orthopoxvirus genomes, there are inverted terminal repeats (ITRs), which have been examined in some detail in the case of vaccinia virus and cowpox virus (CPV) (2–5). The ITRs of these viruses are about 10 kbp long and comprise two regions: an internal unique region about 7.5 kbp long that contains coding sequences and a terminal region that usually contains two blocks of four closely related sequences that are about 50 nucleotides long and are repeated up to about 30 times. In spite of not encoding proteins, these terminal regions of the ITRs of vaccinia virus and CPV are very similar—that is, they have been highly conserved (4).

Several orthopoxviruses, such as monkeypox virus (MPV), rabbitpox virus (RPV), and CPV, form red ulcerated pocks on the chorioallantoic membrane (CAM) of the developing chicken embryo; but up to about 1% of the pocks are usually white (6, 7). The genomes of many of these white-pock mutants or variants are rearranged: thousands of base pairs in either one flanking region or the other are deleted and replaced with inverted duplications of sequences from the other flanking region (8–11). Since these variants are capable of multiplying in a variety of cultured cells as well as in the cells of the CAM, much information encoded at either end of the orthopoxvirus genome is clearly not essential for virus multiplication *per se*. The purpose of this study was to determine how the genomes of a series of white-pock variants of CPV differ from that of the wild-type virus and to gain an understanding of the mechanism responsible for their generation.

MATERIALS AND METHODS

Virus and Viral DNA. CPV strain Brighton Red (CPV-BR) and white-pock variants isolated from it as described by Fenner (6) were grown either in the CAM of 11-day-old chicken embryos or in Vero cells. Virus was purified as described by Joklik (12). DNA was isolated from virus particles as described by Nevins and Joklik (13).

Bacterial and Phage Strains. *Escherichia coli* HB101 was used routinely as the host for the various plasmids. *E. coli* RS5033 (kindly provided by Paul Modrich) was used as the host for preparing plasmid DNAs lacking *dam* methylation at d(G-A-T-C) sequences (14). *E. coli* JM103 was used as the host for the phage M13 vectors mp8 and mp9 (15).

Plasmid DNA. Restriction endonuclease fragments of viral DNA were purified by agarose gel electrophoresis and inserted into pBR322 or pKB111 by standard procedures. Plasmid DNA was purified as described by Marko *et al.* (16). The terminal portion of the viral DNA was cloned as described by Pickup *et al.* (17).

Restriction Endonuclease Cleavage Site Mapping. Restriction endonuclease recognition sites were mapped by standard techniques. Mapping of *Cla* I cleavage sites was done either by using plasmid DNA prepared in the Dam$^-$ *E. coli* strain RS5033 or by mapping the sites in viral DNA. To map *Cla* I sites in viral DNA, the procedure of Smith and Birnstiel (18) was used with the following modification: instead of resolving the products of a partially digested, end-labeled fragment of DNA, a DNA probe labeled by nick-translation and specific for the end of the region to be mapped was hybridized to a blot of the resolved products of a partial restriction endonuclease digest.

Characterization of the DNAs of White-Pock Mutants of CPV-BR. Viral DNAs were digested with various restriction endonucleases. The resultant fragments were resolved by gel electrophoresis and transferred to nylon membranes (Biodyne A, Pall Ultrafine Filtration, Glen Cove, NY) by the procedure of Southern (19). Various cloned fragments of viral DNA were labeled by nick-translation and then were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CPV, cowpox virus; CPV-BR, CPV strain Brighton red; kbp, kilobase pair(s); MPV, monkeypox virus; RPV, rabbitpox virus; CAM, chorioallantoic membrane.

used as hybridization probes for the DNAs on these blots. In addition, some of the novel junction fragments present in the DNAs of the white-pock variants were cloned, mapped, and subjected to nucleotide sequence analysis.

Nucleotide Sequence Analysis. Fragments of DNA to be sequenced were cloned into phage M13 vectors mp8 and mp9 (15). Nucleotide sequence determination was done by the method of Sanger *et al.* (20), with the modifications described by Biggin *et al.* (21).

RESULTS

Restriction Endonuclease Maps of the Flanking Regions of CPV-BR DNA. The two flanking regions of CPV-BR DNA were designated region 1 and region 2 with respect to their locations relative to the positions of the two *Sma* I cleavage sites (Fig. 1).

Since previous work on white-pock variants of RPV, MPV, and CPV (8–10, 22) had indicated that their genomes differed from those of the parental wild-type virus strains in their terminal regions, a detailed restriction endonuclease cleavage map of regions R1 and R2 of CPV-BR DNA was prepared first. Overlapping and adjacent restriction endonuclease fragments spanning these two regions were cloned into plasmid vectors, and maps of the cloned inserts were constructed and used to create the composite restriction endonuclease cleavage maps shown in Fig. 2. By comparing the maps of regions 1 and 2, and by nucleotide sequence analysis of the regions on the internal side of the *Pst* I cleavage sites located closest to the termini of the DNA, the ITRs were found to extend 375 nucleotides inwards from these *Pst* I sites. Thus, the ITRs in the DNA of wild-type CPV-BR are about 9.7 kbp long. No remarkable features are present in the nucleotide sequences at the junctions between the ITRs and the main body of the genome (Fig. 3).

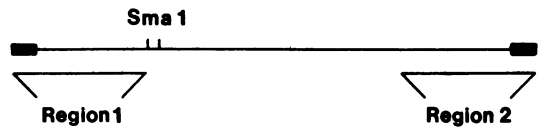

FIG. 1. The genome of CPV-BR. The locations of the *Sma* I cleavage sites are indicated, and the two flanking regions are designated region 1 and region 2 depending on their location relative to them. The black rectangles represent the ITRs. The size of each flanking region is about 40 kbp.

Isolation of White-Pock Variants and the Characterization of Their DNAs. Virus was isolated from white pocks picked from the CAM of infected chicken embryos. Some of the isolates were homogeneous as judged by restriction endonuclease analysis, but others were heterogeneous. Heterogeneity was eliminated in some, but not all, of the latter by a cycle of pock or plaque purification. Isolates that were homogeneous were then grown up, and their DNAs were extracted and analyzed by restriction endonuclease digestion. Ten variants (W1 to W10) that differed from each other in the structure of their DNAs were examined in detail. Fig. 4 shows agarose gel electrophoretograms of the products of *Pst* I digestion of the DNAs of these white-pock variants (except very small fragments). Three features stand out. First, three bands, corresponding to *Pst* I fragments G, L, and O, are not present in the DNA of any variant, and several variants contain novel *Pst* I fragments. This indicates that large

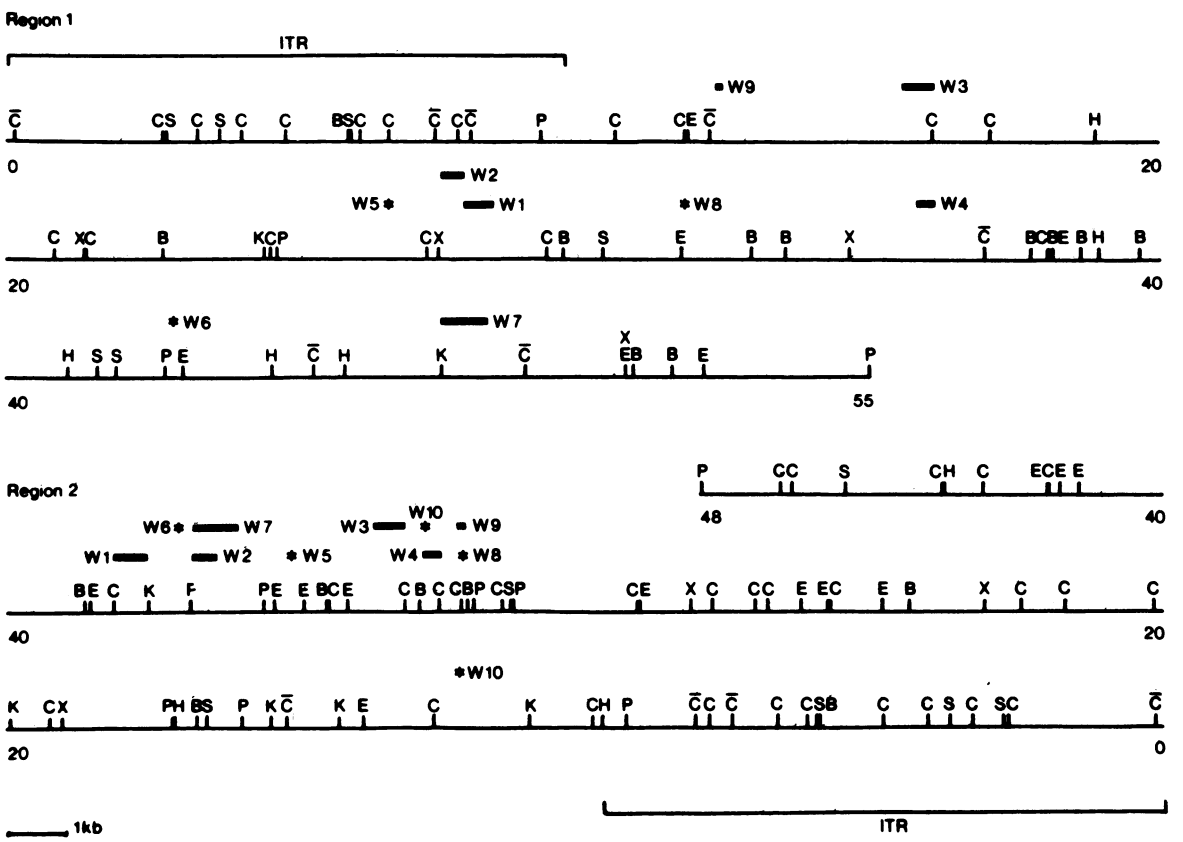

FIG. 2. Restriction endonuclease cleavage maps of region 1 and region 2 of CPV-BR DNA. For both maps, the coordinate 0 corresponds to the terminus of the viral DNA. The numbers correspond to the distances in kbp from the terminus. Recognition sites for restriction endonucleases are indicated as follows: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn* I; P, *Pst* I; S, *Sal* I; X, *Xho* I; C, *Cla* I; $\bar{C}$, *Cla* I sites that may be methylated in plasmid DNA extracted from *Dam*$^+$ strains of *E. coli*. Thickened bars and asterisks indicate the map locations of the regions where the novel junction sequences present in the DNAs of individual white-pock variants are located. Asterisks denote junctions that were identified by nucleotide sequence analysis.

```
region 1-ITR    TCGTAGCTAAAACTCAAGTAAGAGGGTTTTATTATCTCCGTCATACGTAAATGCCTTCTTAAGCTATTTG
region 2-ITR    CCTTAAAGCTTCTGATGGTAACTGTGTTACATGTGCTCCGTCATACGTAAATGCCTTCTTAAGCTATTTG
```

FIG. 3. Nucleotide sequences of the junctions between the ITRs and unique region DNA in regions 1 and 2 of the genome of CPV-BR. Both sequences are shown in a 5'-to-3' orientation such that the 5' end would correspond to the inner end of that sequence in the genome.

regions of DNA present in CPV-BR DNA are not present in the DNAs of white-pock variants. Second, *Pst* I cleaves within the ITRs at about 9.3 kbp from the termini of CPV-BR DNA. Therefore, the terminal fragment is present in viral DNA in two copies. It is also present in two copies in the DNAs of all white-pock variants. Third, most of the variants exhibit other bands that are also present in double molar amounts. Therefore, certain sequences are duplicated in the genomes of white-pock variants.

Comparison of the restriction endonuclease maps of regions 1 and 2 of CPV-BR DNA with those of the 10 white-pock variants revealed that, although in all of them region 1 was intact, all of them had undergone DNA rearrangements in region 2 (Fig. 5). In variants W1 to W9, most of region 2 has been deleted. The regions deleted are similar in size; they range from 38 kbp in variant W1 to 32 kbp in variant W8 (see also Fig. 2). These deleted sequences are replaced by sequences of region 1 inserted in the opposite sense, so that the presence at the end of the genomes of the characteristic CPV-BR ITR is maintained. In contrast to the rather uniform size of the region 2 deletions, the lengths of the duplicated region 1 sequences vary widely; excluding the ITR, they range in size from about 2.5 kbp for variant W9 to about 40 kbp for variant W7 (see also Fig. 2). The net effect of these deletions/duplications is that the sizes of the genomes of the white-pock variants vary greatly, the genome of variant W9 being 20 kbp smaller than that of CPV-BR, while that of variant W7 is 12 kbp larger. Further, all nine variants contain ITRs that are larger than that of CPV-BR; the smallest is the ITR of variant W9, which is about 12 kbp long, and the largest is that of variant W7, which is almost 50 kbp long.

Variant W10 differs from the other nine variants in that its genome contains the same ITRs as CPV-BR; the DNA of this variant simply possesses an internal deletion about 20 kbp long in region 2. Significantly, the inner end point of this deletion is close to that of the deletions in the other white-pock variants. Thus, the genomes of all 10 white-pock variants contain large overlapping deletions.

Collectively, these data indicate that at least 28 kbp of the DNA present in region 2 of CPV-BR are not essential for virus multiplication in cells of the CAM of the developing chicken embryo or in Vero cells.

Nucleotide Sequences at the Novel Junctions in the DNAs of White-Pock Variants. A simple mechanism for generating the genomes of the white-pock variants might involve a single intermolecular recombination event (8). Two CPV-BR genomes would be postulated to align in opposite orientations, and recombination between opposed regions 1 and 2 would then produce the molecules described above.

In order to determine whether the regions where recombination occurred contain unusual features, and to ascertain whether the proposed recombinations are homologous or

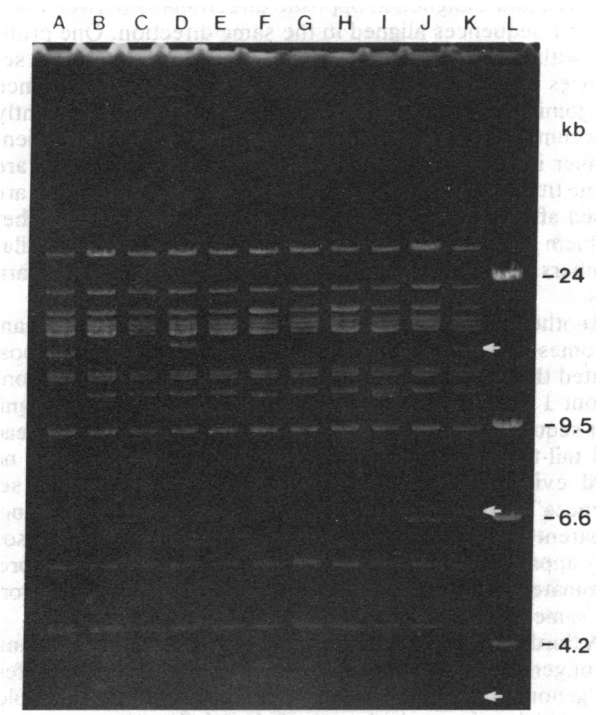

FIG. 4. Agarose gel electrophoretograms of the products of *Pst* I digestion of various CPV DNAs. Lanes: A–J, DNAs of variants W1 to W10, cleaved with *Hin*dIII; K, CPV-BR DNA; L, bacteriophage λ DNA restricted with *Hin*dIII. The white arrows indicate *Pst* I fragments G, L, and O of CPV-BR DNA that are absent from the DNAs of white-pock variants. The *Pst* I fragment that is present in the ITR and, therefore, is present in viral DNA in two copies is that near the 9.5-kbp marker fragment.

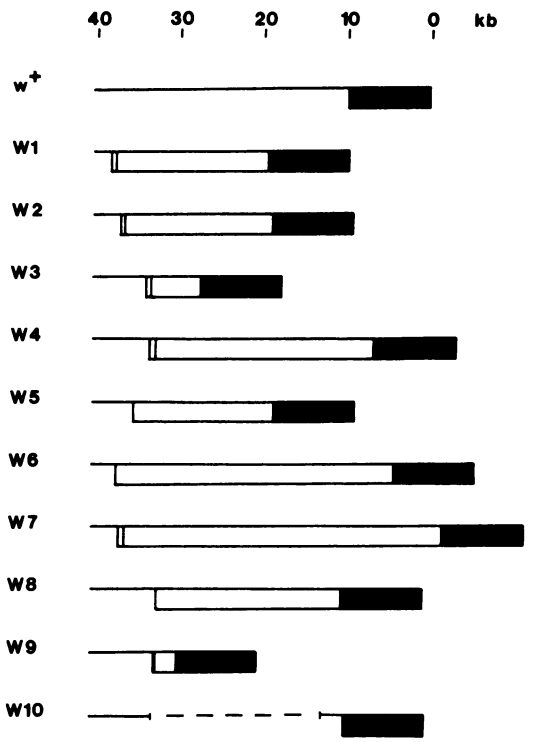

FIG. 5. Structures of region 2 of white-pock variants of CPV-BR. The genomes are aligned at a coordinate that maps at 40 kbp from the left terminus of CPV-BR DNA. Horizontal lines correspond to unrepeated CPV-BR DNA; black rectangles correspond to the ITR of CPV-BR DNA. Open rectangles correspond to region 1 DNA inverted relative to its orientation at the other end of the genome. Double vertical lines at the left-hand ends of open rectangles correspond to the locations of novel junctions between region 1 and 2 DNA sequences; single vertical lines represent novel junctions that were sequenced (see Fig. 6). The broken line in variant W10 corresponds to the sequences deleted in this variant. W^+ indicates the genome of CPV-BR.

FIG. 6. Nucleotide sequences adjacent to the novel junctions that are present in the genomes of white-pock variants. All sequences are oriented as described in Fig. 3. The variant sequences are seen to be composed of sequences from both regions 1 and 2. A single crossover between the DNA of regions 1 and 2 in the locations shown would produce the junction sequences present in the DNAs of variants W5, W6, and W8. Variant W10 is a simple deletion mutant that could have arisen by recombination between two regions in region 2 DNA.

nonhomologous, the locations of the recombination sites were determined (Fig. 2). The regions around the novel junctions in four variants were then sequenced, as were the corresponding regions of the DNA of CPV-BR. The four variants chosen for this study were deletion/duplication variants W5, W6, and W8 and the simple deletion variant W10.

For each variant, a single crossover event between region 1 and region 2 DNA (or, in the case of variant W10, between two regions in region 2) of CPV-BR would produce the novel junction sequence (Fig. 6). No statistically significant direct or staggered sequence homologies are discernible in any variant between the two regions of DNA at the junction sites, nor is there any evidence for the existence of "filler" DNA sequences (23). Further, there are no significant alternate purine/pyrimidine stretches indicative of an ability to assume the Z configuration. Finally, the sequences were examined for dyad symmetry. Although the region 2 sequence at the junction site in variant W8 possesses dyad symmetry, none of the other regions do so.

Much larger regions, up to 1000 bp long, around the junction regions of several of these variants were also sequenced. These sequences also fail to reveal significant homologies, dyad symmetries, or other notable features.

The conclusion from inspection of all these sequences is that homologous recombination does not appear to be involved in the generation of the genomes of white-pock CPV variants.

DISCUSSION

The genome rearrangements that lead to the formation of the white-pock variants of CPV possess several remarkable features: (*i*) they occur frequently, (*ii*) the sequences where arrangements occur are nonhomologous and devoid of unusual features, (*iii*) the rearrangements are always clustered in one particular region in one parent, and (*iv*) deletion/duplication variants are about 10 times more frequent than simple deletion variants. The evidence for the last feature derives not only from the work reported here but also from other reports concerning the nature of the genomes of RPV, MPV, CPV, and vaccinia virus variants (8–11, 22, 24–26).

The simplest way to account for the generation of the genomes of the white-pock variants of CPV is by a single crossover recombinational event (8), which would generate deletion/duplication variants or simple deletion variants depending on whether recombination occurs between sequences in regions 1 and 2 aligned in opposite directions or between two region 2 sequences aligned in the same direction. One problem with this mechanism is the nonhomology of the sequences where recombination occurs. However, evidence that joining of nonhomologous sequences occurs frequently is accumulating rapidly. Examples are the highly efficient manner in which fragments of retrovirus proviral DNA are reconstituted (27, 28) and the ends of unrelated DNAs are joined after transfection into eukaryotic cells (29). Another problem is that this mechanism would yield roughly similar numbers of deletion/duplication and simple deletion variants.

Another model for the generation of white-pock variant genomes was proposed by Moyer and Graves (30), who postulated that the duplications result from very large deletions (about 1 unit genome length) that occur between "recognition sequences" within putative concatameric head-to-head and tail-to-tail replication intermediates. While there is no hard evidence against this model, the "recognition sequences" that would specify sites of recombination are not apparent in CPV-BR DNA. Therefore, there is no reason why approximately unit genome-length deletions should predominate, and consequently, this model would suffer from the same problems as the first model discussed above.

A third model is one that involves the nonreciprocal transfer of genetic information between terminal regions of different genomes or of the same genome. For example, a double-stranded break or single-stranded nick [perhaps one produced during DNA replication (3)] near one terminus might enable a free 3′ single-stranded end to invade the duplex at the other end of the genome and either repair or replace the gapped end with a newly synthesized copy of the intact end. The loop structure at the DNA terminus would facilitate the synthesis of a double-stranded replica of the template terminus. Intramolecular exchanges of this type would produce

deletions/duplications, while intermolecular exchanges would produce both deletions/duplications and simple deletions, depending on the ends involved. This mechanism would tend to yield an excess of deletions/duplications over simple deletion variants; therefore, it may resemble more closely the actual mechanism that generates white-pock variant genomes.

The distribution of deletion/duplication end points, scattered throughout region 1 but clustered in region 2, deserves comment. This distribution is characteristic not only of the 10 variants studied here but also of mutants recently examined by Archard *et al.* (11): the inner ends of deletions in region 2 are all located between 38 and 32 kbp from the right-hand end. It has been suggested that this region of the DNA possesses unusual/special features (11); but this is not apparent in the immediate vicinity of the junctions, nor is there any evidence for such features in the longer sequences that we established for some of the variants. Therefore, the reason why the inner ends of the deletions are limited to the region between 38 and 32 kbp is more likely to be due to the nature of the information encoded in this region. It is interesting that no attempt has yet been made to map the position of the "red-pock" gene(s) in MPV, RPV, or CPV. We have found that in CPV the information for red-pock phenotype is located at a position that maps at about 32 kbp from the terminus of region 2: the insertion of an *Eco*RI fragment of DNA that spans the region from 29 to 34 kbp from the terminus of region 2 of CPV-BR DNA into the genomes of white-pock variants results in the production of recombinant virus that produces red ulcerated pocks (unpublished results). Therefore, the reason why the inner ends of the deletions are no closer to the genome end than 32 kbp is because that is where the "red-pock" gene is located; and the reason why they are no further in than 38 kbp is most probably because a gene that is essential for virus multiplication is located at this position.

Finally, like the genome of CPV-BR, the genomes of white-pock variants possess ITRs; however, theirs are much larger, some up to 50 kbp long. We have sequenced both ITR–unique sequence junction regions in CPV-BR DNA (Fig. 3) and find them to be as devoid of unusual structural features as are the junction regions of the four white-pock variants that we examined. Presumably, the ITRs in CPV-BR arose by a mechanism similar to that postulated above for the white-pock variants. As for the significance of the ITRs, they clearly could increase the genetic potential of orthopoxviruses: large segments of DNA could be deleted as well as added, the relationship of control sequences to coding sequences could be altered, and novel coding sequences could be created. Clearly, variation in flanking regions, which appear to encode information that is not essential for actual virus multiplication but rather relates to the nature of the virus–cell interaction, could play a major role in the generation of new orthopoxvirus strains. These regions of orthopoxvirus DNAs may provide a useful and experimentally accessible model system for studying virus evolution in particular and genome modification in general.

We gratefully acknowledge the expert technical assistance of Sherry Larson and Rodger Pryzant. This work was supported by Research Grant 1 RO1 08909 and Program Project Grant 2 PO1 CA 30246 from the National Institutes of Health.

1. Mackett, M. & Archard, L. C. (1979) *J. Gen. Microbiol.* **45,** 683–701.
2. Wittek, R. & Moss, B. (1980) *Cell* **21,** 277–284.
3. Baroudy, B. M., Venkatesan, S. & Moss, B. (1982) *Cell* **28,** 315–325.
4. Pickup, D. J., Bastia, D., Stone, H. O. & Joklik, W. K. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 7112–7116.
5. Baroudy, B. M. & Moss, B. (1982) *Nucleic Acids Res.* **10,** 5673–5679.
6. Fenner, F. (1958) *Virology* **5,** 502–529.
7. Gemmel, A. & Fenner, F. (1960) *Virology* **11,** 219–235.
8. Moyer, R. W., Graves, R. L. & Rothe, C. T. (1980) *Cell* **22,** 545–553.
9. Moyer, R. W. & Rothe, C. T. (1980) *Virology* **102,** 119–132.
10. Esposito, J. J., Cabradillo, C. D., Nakano, J. H. & Obijeski, J. F. (1981) *Virology* **109,** 231–243.
11. Archard, L. C., Mackett, M., Barnes, D. E. & Dumbell, K. R. (1984) *J. Gen. Virol.* **65,** 875–886.
12. Joklik, W. K. (1962) *Virology* **18,** 9–18.
13. Nevins, J. R. & Joklik, W. K. (1977) *J. Biol. Chem.* **252,** 6930–6938.
14. Lu, A.-L., Clark, S. & Modrich, P. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4639–4643.
15. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
16. Marko, M. A., Chipperfield, R. & Birnboim, H. C. (1982) *Anal. Biochem.* **121,** 382–387.
17. Pickup, D. J., Bastia, D. & Joklik, W. K. (1983) *Virology* **124,** 215–217.
18. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398.
19. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
20. Sanger, F., Coulson, A. R., Barell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
21. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
22. Archard, L. C. & Mackett, M. (1979) *J. Gen. Virol.* **45,** 51–63.
23. Anderson, R. A., Kato, S. & Camerini-Otero, R. D. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 206–210.
24. McFadden, G. & Dales, S. (1970) *Cell* **18,** 101–108.
25. Drillien, R., Koehren, F. & Kirn, A. (1981) *Virology* **111,** 488–499.
26. Panicali, D., Davis, S. W., Mercer, S. R. & Paoletti, E. (1981) *J. Virol.* **37,** 1000–1010.
27. Kopchick, J. J. & Stacey, D. W. (1983) *J. Biol. Chem.* **258,** 11528–11534.
28. Bandyopadhyay, P. K., Watanabe, S. & Temin, H. M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3476–3480.
29. Wilson, J. H., Berget, P. B. & Pipas, J. M. (1982) *Mol. Cell. Biol.* **2,** 1258–1269.
30. Moyer, R. W. & Graves, R. L. (1981) *Cell* **27,** 391–401.