Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 9375–9379, December 1989
Genetics

Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes

(trypanosomes/antigenic variation/gene conversion/diversity)

CHARLES ROTH*, FREDERIC BRINGAUD†, RONALD E. LAYDEN‡, THEO BALTZ†, AND HARVEY EISEN*‡§

*Unit of Immunoparasitology, Institut Pasteur, 25, rue du Dr. Roux, 75724, Paris Cedex 15, France; †University of Bordeaux II, Bordeaux, France; and ‡Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104

Communicated by Martin Gellert, July 17, 1989

ABSTRACT **The expression of genes coding for variable surface glycoproteins (VSGs) in *Trypanosoma equiperdum* is linked to duplicative transpositions of silent, basic copy sequences into telomere-linked expression sites. Examination of three independently derived late-appearing trypanosome clones expressing VSG-78 revealed that the expressed gene in all cases is composed of sequences derived from three or four individual silent genes. The 182 base pairs at the 3′ end of the coding sequence are derived from one silent gene, the 3′ donor. The remaining 5′ segment is a mosaic structure containing variable-length segments derived from two, or perhaps three, related silent genes. All of the silent genes that participate in the construction of the VSG-78 expression-linked copy (ELC) genes contain multiple stop codons and are unable to code for VSGs. Individual silent pseudogenes complement one another in the mosaic structure of the 5′ portions of the ELC genes and create functional VSG genes. The joining of the 3′ and 5′ portions of the composite genes occurs in short regions of homology and suggests a mechanism by which the ordered expression of the VSG genes is generated.**

African trypanosomes are extracellular hemoflagellates which are responsible for widespread disease in humans and domestic animals. These parasites evade the host immune system by periodically changing their variable surface glycoproteins (VSGs), which are thought to be the unique parasite antigens presented to the host immune system. While the mechanism of this antigenic variation is not yet understood, it appears that activation of the VSG genes is a complex process involving multiple steps. Silent VSG genes are duplicatively transposed into one of several telomere-linked expression sites where they can be transcribed as expression-linked copies (ELCs) (reviewed in ref. 1). Since transcription of the VSG genes can be regulated within expression sites, their insertion into an expression site, although necessary for transcription, is not sufficient (2–7).

It has been shown for *Trypanosoma equiperdum* and *Trypanosoma brucei* that the variants appear at a predictable time in a loosely defined order (8–10). Furthermore, transfer of any *T. equiperdum* variant to a nonimmune host results in the reappearance of the initial variant, variant antigen type (VAT)-1.

In previous studies we have shown that VSGs appearing early in a *T. equiperdum* infection are derived from complete silent-copy genes (11). On the other hand, the late-appearing VSGs appear to be coded for by composite genes that exist in complete form only in expression sites (11, 12). The 3′-most 200–300 base pairs (bp) of the late-appearing ELCs are donated by one silent-copy gene, the 3′ donor, while the remaining 5′ portion is donated by an unrelated silent gene or family of genes. We have cloned and sequenced all of the component silent-copy genes used to make the VSG-78 family of ELC genes.¶ We report here that the 5′ portions of all the ELCs examined are mosaics of two or possibly three silent-copy genes. Furthermore, three independently isolated VSG-78 ELCs are all composed of different combinations of the donor silent genes. None of the silent genes (either 5′ or 3′) has sufficient open reading frame to encode a VSG on its own. We also demonstrate that there is an 85-bp region of homology between the 5′ and 3′ donor genes and that it is within this region that the recombination occurs to create the active VSG gene. Thus, the complete VSG-78 ELCs are constructed entirely from pseudogenes. These results indicate that the parasites have a mechanism for the formation of late-appearing VSG genes that generates temporary VSG coding genes and that this mechanism leads to antigenic diversification.

MATERIALS AND METHODS

Trypanosomes. All of the trypanosome clones used in this study were derived from Bordeaux trypanozoon antigen type (BoTat)-1 (9). BoTat-78 and -78^{bi} were isolated from independent infections initiated with BoTat-1; BoTat-78^{20} and -1^{78} were isolated from infections initiated with BoTat-20 and -78, respectively. DNA and RNA were isolated from cloned trypanosomes grown in rats (13).

Recombinant Clones. The cDNAs corresponding to VSG-78 and -78^{20} were prepared by the method of Gubler and Hoffman (14). cDNA clones corresponding to VSG-78^{bi} were made in the same way. The 3′ end of the VSG-78^{bi} gene was not cloned. The 725-bp *Pst* I–*Bam*HI fragment of the VSG-78 cDNA was subcloned in plasmid pGEM2. This plasmid is designated pTe78-100. The genomic clone corresponding to the VSG-78 *BC-B* was described previously (12). The genomic clones corresponding to *BC-A* and *-C* were subclones made from *Hin*dIII genomic fragments isolated from BoTat-1^{78}, the immediate daughter variant of BoTat-78. The genomic clone corresponding to the *BC-3′* silent gene was obtained as a *Dra* I fragment in pGEM2.

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: VSG, variable surface glycoprotein; ELC, expression-linked copy; BC, basic copy.
§To whom reprint requests should be sent at the ‡ address.
¶The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M29497 for VSG-78, M29498 for VSG-78^{20}, M29499 for VSG-78^{bi}, M29501 for *BC-A*, M29502 for *BC-B*, M29503 for *BC-C*, and M29500 for *BC-3′*).

Southern Blot Analysis. Genomic DNA was digested with the appropriate restriction enzymes, electrophoresed in 1% agarose, blotted, hybridized with ^{32}P-labeled probes, and washed as previously described (13).

DNA Sequencing. DNA was sequenced by the dideoxynucleotide chain-termination technique described by Sanger *et al.* (15). Appropriate DNA fragments were cloned in commercially obtained M13 vectors: mp8, -9, -10, and -11 (16). Second-strand DNA synthesis was primed either with universal M13 primer or with specific 15-mer or 18-mer oligonucleotides synthesized on an Applied Biosystems DNA synthesizer. Both strands were sequenced.

Labeled Transcripts. ^{32}P-labeled RNA transcripts corresponding to the sense strand of the VSG-78 ELC were prepared with T7 RNA polymerase from *Bam*HI-digested plasmid pTe78-100 according to the supplier (Stratagene). Template DNA was removed with RNase-free DNase (80 units/ml).

RNase Protection. ^{32}P-labeled RNA (500 pg) was hybridized for 15 hr to various amounts of *Hin*dIII-digested, heat-denatured, cloned or genomic DNA in 30 μl of hybridization buffer [80% (vol/vol) formamide/40 mM Pipes, pH 6.4/0.4 M NaCl/1 mM EDTA] at 45°C. RNase digestion and preparation of samples for electrophoresis were as described by Myers *et al.* (17). Protected products were analyzed on 5% polyacrylamide gels containing 8 M urea.

RESULTS

Restriction Site Polymorphisms in the VSG-78 Genes. We have studied the genes that code for VSG-78, VSG-78^{20}, and VSG-78^{bi}, three independently derived antigenic variants of *T. equiperdum* that are recognized by the same polyclonal and some monoclonal antibodies (ref. 12 and unpublished results). Despite the similarity of the three proteins, there are restriction enzyme site differences in their corresponding genes (see Fig. 1). We have previously shown by Southern blot analysis that there are two or three closely related silent genes that could code for the N-terminal portion of the VSG-78 antigens (11, 12). The simplest hypothesis is that the VSG-78 ELC was constructed by duplication of one of the silent genes, that the VSG-78^{20} ELC was generated by duplication of the second, and the VSG-78^{bi} was generated by duplication of the third silent gene. However, comparison of the restriction maps of the three ELCs with those of the three silent genes (see Fig. 1) reveals that none of the silent genes contains all the restriction sites of any of the ELCs. ELC specific restriction fragments could not be detected in the genomic DNA of trypanosomes expressing VSGs other than 78 (data not shown). This suggests that the simple hypothesis is wrong and that none of the silent genes alone codes for the 5′ end of the ELCs.

To verify that the trypanosomes do not contain a single silent copy corresponding to any of the three 78 ELCs, we examined the genomes of BoTat-1 [from which BoTat-78 was derived (9)], BoTat-78, and BoTat-1^{78} (derived directly from BoTat-78) for the presence of sequences completely homologous to the 5′ portion of the 78 ELC. Uniformly ^{32}P-labeled RNA corresponding to the 725 bases 5′ of the *Bam*HI site in the 78 ELC was prepared *in vitro* and hybridized to denatured genomic DNA from variant 78, 1, or 1^{78}. RNase was added under conditions for digestion of single-stranded RNA and mismatched DNA·RNA complexes. Under these conditions, approximately 50% of single base pair mismatches are cleaved by the RNase (17). The undigested RNA was then purified and examined on a polyacrylamide gel. Since there do not appear to be introns in African trypanosomes, the experiment would not detect the results of cis splicing. It can be seen in Fig. 2*b* that only DNA from variant 78 contained sequences completely homologous to the radioactive 78-specific RNA. Thus, we conclude that the silent VSG gene repertoire does not contain a complete colinear copy of the 5′ portion of the 78 ELC.

The VSG-78, -78^{20}, and -78^{bi} ELCs Are Mosaic Genes. To determine in more detail how the 78, 78^{20}, and 78^{bi} ELCs originated, we sequenced cDNAs and compared the sequences to those of the four silent basic copy genes *BC-A*, *BC-B*, and *BC-C*, which make up the 5′ donor family, and *BC-3′*, the 3′ donor (Fig. 3). As predicted from the restriction maps, the 5′-most 1200–1224 bp of the ELCs can be entirely derived from this family of silent genes. We cannot, however, formally rule out contributions from other genes having small regions of homology with the 78 5′ donor family. The remaining portions of the ELC sequences are derived from

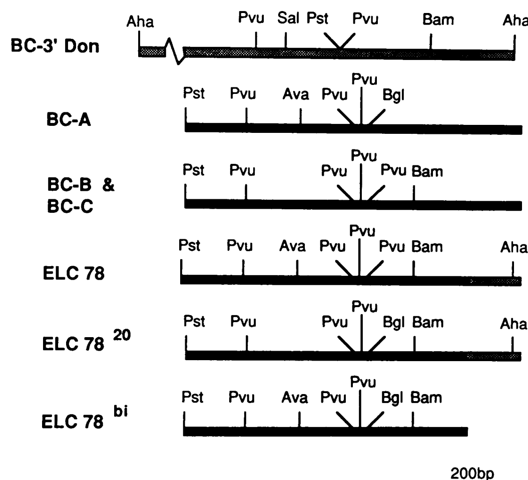

FIG. 1. Restriction maps of VSG-78 ELC and basic copy (BC) genes. Restriction maps of the 5′ portions of VSG-78, -78^{20}, and -78^{bi} ELCs and the silent-copy genes used for their construction. The solid bars represent the VSG coding regions. Restriction enzymes: Aha, *Aha* III; Ava, *Ava* I; Bgl, *Bgl* I; Bam, *Bam*HI; Pvu, *Pvu*II; Pst, *Pst* I; Sal, *Sal* I.

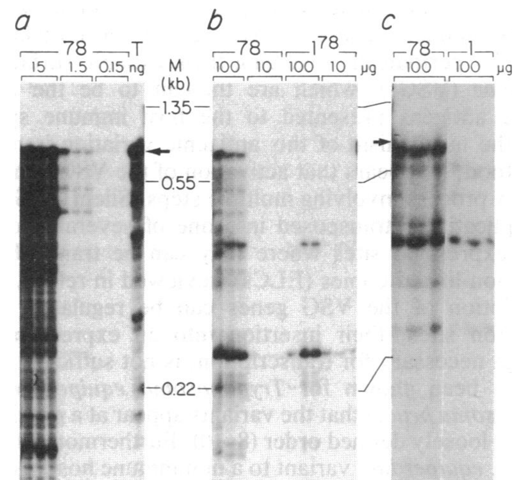

FIG. 2. Lack of colinear silent copy gene homologous to the 5′ end of the VSG-78 ELC. Labeled synthetic RNA transcripts representing the 5′ portion of the VSG-78 ELC sense strand were hybridized with the indicated amounts of *Hin*dIII-digested DNA. Aliquots of the hybrids were then treated with RNase for 15 min, 1 hr, or 2 hr. (*a*) Results with increasing amounts of cloned BoTat-78 cDNA from plasmid pTe 78-100. The lane labeled T shows the untreated transcripts. The arrow shows the full-length 725-bp transcript. M, markers, in kilobases (kb). (*b*) Results with increasing amounts of genomic DNA from BoTat-78 or -1^{78}. (*c*) Results with increasing amounts of genomic DNA from BoTat-78 or -1.

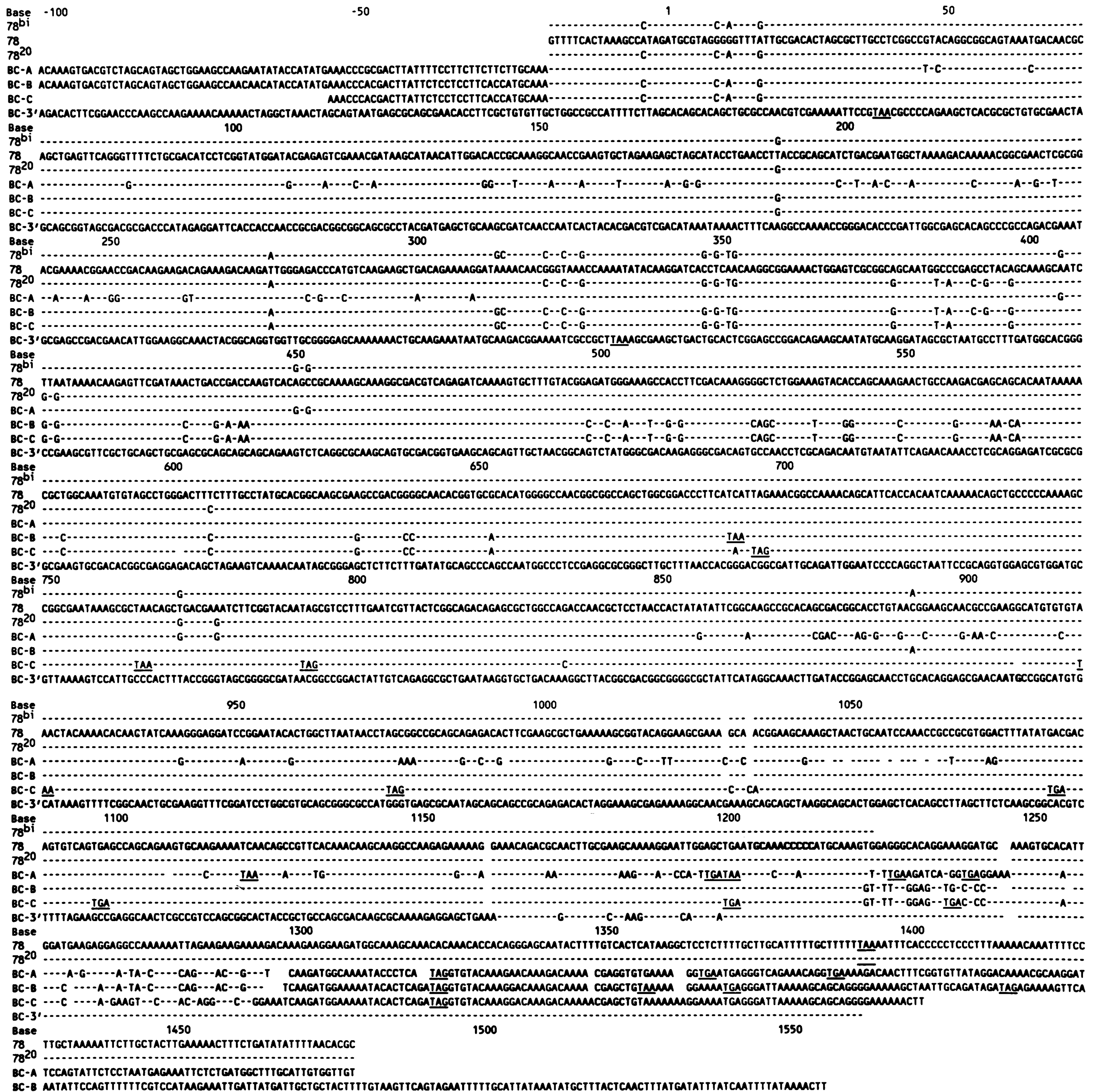

FIG. 3. DNA sequences of ELCs and silent copies. The complete sequence of the VSG-78 cDNA is shown. Nucleotide identity of the other cDNAs or the genomic silent copies *BC-A*, *BC-B*, *BC-C*, and *BC-3'* to VSG-78 is marked by a dash. Nucleotide differences are shown. In-frame chain termination codons in the silent genes are underlined. The first ATG in the reading frame used for the 3' portion of *BC-3'* (bp 915) is in boldface type. Sequences are arranged to give the best alignment.

the 3' donor. To account for the sequence arrangement in the 5' 1200 bases of the three ELC genes, regions must be taken from at least two of the silent genes of the 5' donor family. In most of the cases, sequences derived from *BC-B* and *BC-C* cannot be distinguished. There are no regions in any of the ELCs that can unambiguously be identified as having been derived from *BC-C* although the C residue at position 892 in the 78 and 78^{20} ELCs is likely to be *BC-C* derived. The boundaries between patches derived from the different silent copy genes were estimated as the midpoints between nonidentities in the silent genes and are shown in Fig. 4. Different patterns of silent gene segments are used to form the 5' portions of all three mosaic VSG-78 ELC genes. Since there are numerous regions of identity in the three silent genes, unseen patches cannot be ruled out. The simplest interpretation of these results is that the 5' ends of the ELCs are mosaics, constructed *ad hoc* with blocks of sequence from two or possibly three of the silent genes.

The Silent Genes Are Incomplete and Defective. We have previously shown that the silent 78 genes *BC-A* and *-B* are incomplete genes that do not have competent 3' ends and must depend on interaction with a 3' donor gene for the missing information (12). This is also the case for *BC-C* (Fig. 3). Examination of the sequences of the silent 78 *BC* genes reveals that all are pseudogenes and incapable of individually coding for the 5' end of a VSG gene. *BC-B* has an in-frame

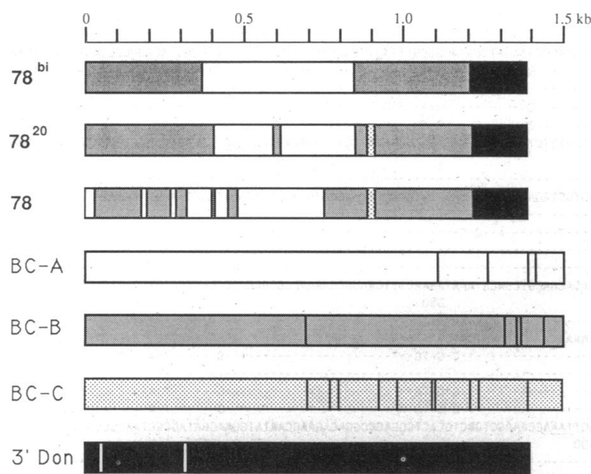

FIG. 4. Structure of the VSG-78 silent gene family and the -78, -78[20], and -78[bi] ELCs. The 78 silent-copy genes and ELCs are shown schematically. The positions of in-frame stop codons in the silent genes *BC-A*, *BC-B*, and *BC-C* are shown as dark vertical lines. The segment boundaries of the ELCs derived from different silent copies were estimated. 3′ Don indicates a 3′ donor gene.

stop codon at position 693 (see Figs. 3 and 4). *BC-C* differs very little from *BC-B*. However, a single base deletion at position 601 leads to numerous stop codons in all reading frames. *BC-A* has the longest open reading frame of the three 5′ silent genes. Nevertheless, the presence of several frameshifts results in an in-frame TAA at position 1123.

We have also cloned and sequenced the 3′ basic copy of the 78 gene (Figs. 3 and 4). Like the 5′ silent copy genes, it lacks sufficient open reading frame to code for a VSG. Although it can code for the 3′ end of the 78 ELC genes, it cannot code for the 5′ end in the same reading frame because there are TAA codons at bases 36 and 333. Also, the first methionine codon is at base 915. This gene is not a member of the family of silent genes that codes for the 5′ region of the ELC. Rather, between bases 1164 and 1268, it shares 77% identity with *BC-B*. It is in this short homology region that the recombination event that created the 78 ELC genes must have occurred.

The Silent Pseudogenes Evolve from Complete Genes. There are two highly conserved sequences, referred to as the 8-mer and the 14-mer, located at the 3′ ends of all complete VSG genes (18-20). As shown in Fig. 5, the 8-mer and 14-mer sequences are present between bases 1444 and 1476 of the VSG-78 ELC genes. The silent *BC-B* gene sequence between bases 1515 and 1544 aligns with bases 1443–1469 of the VSG-78 cDNA (24 out of 30 bases), and remnants of the 8- and 14-mer sequences can be found (Fig. 5). There is also a 8 of 8 match to the 8-mer starting at position 1473 of *BC-B*. These results strongly suggest that the *BC-B* silent gene has evolved from a complete gene. The apparent lack of homology between *BC-B* and the VSG-78 cDNA or 3′ donor genes between bases 1268 and 1444 suggests that *BC-B* evolved from a complete gene by using a 3′ donor other than the one

```
                     8mer                    14mer
78 cDNA   1443  TTGCT  ACTTGAAAAACTTTCTGATATATTTT AAC
BC-B      1473   TGCT  ACTT
   "      1515   TGCTTTACT C    AACTTTATGATAT TTATCAA
```

FIG. 5. Comparsion of 3′ ends of VSG-78 cDNA and silent *BC-B*. The sequence in the noncoding region of the VSG-78 ELC between nucleotides 1443 and 1476 was aligned with two sequences in the silent *BC-B* gene, nucleotides 1469–1477 and 1515–1544. The sequences were aligned with gaps to provide maximum identity. The highly conserved 8-mer and 14-mer sequences found in all ELC genes are indicated above the VSG-78 cDNA sequence.

used in the 78 and 78[20] ELCs. Although the *BC-A* sequence does not contain the conserved 8-mer and 14-mer in the region that we sequenced, its high degree of homology with *BC-B* suggests that the two have the same origin.

DISCUSSION

We have analyzed the structures of three VSG-78 ELC genes coding for *T. equiperdum* VSGs. These genes were found in independently derived variants from separate infections. All appear to have used the same blueprint for the construction from silent-copy genes, as they are all pieced together from silent copies that are themselves pseudogenes. All three VSG-78 ELC genes make use of two kinds of chimeric structure. First, we have defined the hybrid nature of the ELCs. The 5′ 1200 bases are derived from a family of closely related silent genes, while the 3′ portion is derived from another silent gene that shares only an 85-bp region of homology with the 5′ donor family.

The second type of chimeric structure used for the construction of the VSG-78 ELCs is the mosaic patchwork found in the 5′ segments of the genes. In each case the mosaics are made from highly related silent-copy genes. For the three independently derived VSG-78-expressing isolates we examined, the mosaics all differed, strongly suggesting that the use of incompetent silent-copy genes to create mosaics results in the generation of diversity.

Remarkably, not one of the silent genes used to construct the VSG-78 ELCs is itself capable of coding for a VSG. Despite this, they are able to contribute to the construction of complete VSG genes. This seemingly complex event is not unique since BoTat-20, -20[bis], and -20[ter], other late-appearing variants, use the same mechanism for the construction of their ELC genes (ref. 21, unpublished data). Variants expressing VSG-78 and -20 appear in most *T. equiperdum* infections (9), suggesting that the parasites possess an efficient mechanism for the construction of composite ELC genes. Although we do not know the precise mechanism of this process, our analysis suggests that a gene conversion-like mechanism generates the 5′ mosaics, since the donor genes do not appear to be altered by the process. Similar mechanisms have been proposed for the generation of *Neisseria gonorrhoeae* pilin genes (22, 23) and the chicken immunoglobulin genes (24). It is possible that the 5′ donors first form an intermediate extrachromosomal mosaic structure, which then recombines with the appropriate resident gene within the expression site. It is also possible that there are multiple rounds of conversion of an initial hybrid gene by silent 5′ donor genes.

For both the VSG-78 and VSG-20 families of ELCs, the junction between the 5′ and 3′ donor genes occurs within a short (80 bp for VSG-20, 85 bp for 78) region of homology (21). This strongly supports our earlier model proposing that the 5′ donor sequences for late-expressed VSG genes must find homology with the resident VSG gene in an expression site for integration into the site (11, 12, 25). Furthermore, the finding that the VSG-20 and VSG-78 partners use different homology sequences suggests that this homology might play a role in the generation of ordered VSG gene expression. For example, different families of late 5′ donor genes might share homology with the same 3′ donor. The region of homology with one 5′ donor gene family, family "y," might be located on the 5′ donor segment of the previously expressed family, family "x." Thus, the y family of 5′ donor genes could not be used until after a member of the x family was used in an expression site.

Although the homologies between the 5′ and 3′ donors differ for the VSG-20 and -78 genes, they do share an 11-bp-region identity. Each has the sequence TGCAAAC-CCCC (at position 1205–1215 in Fig. 3) in either the indicated

or opposite orientation (marked in bold type in Fig. 3). In both cases, the actual recombination must have occurred within or very near this sequence. It is thus possible that this 11-mer is important in the recombination process.

We propose that African trypanosomes use mosaic gene formation, or gene conversion, as a primary mechanism for the generation of VSG diversity. This is consistent with the repeated findings that many VSG genes are members of related, but not identical, families (26, 27). The use of such a mechanism for the generation of VSG diversity would allow the parasites to generate large numbers of different VSGs with a rather small repertoire of silent genes. The three ELCs that we have characterized code for very similar VSGs. However, we selected for this by choosing variants that are recognized by the same antisera. It is probable that the same silent genes, used in different configurations, could code for immunologically different VSGs.

The formation of mosaic genes would also amplify diversity generated by base-substitution mutations. Mutations that accumulated randomly in the individual silent copies would be distributed differentially each time an ELC was constructed. Furthermore, since the ends of individual patches in the mosaic genes might occur within codons, a given base substitution could lead to more than one amino acid change. Frameshift mutations could also expand the available amino acid sequence diversity.

The early appearing VSGs are coded for by silent copy genes that contain complete open reading frames and presumably those sequences necessary to allow them to replace any VSG gene in an expression site. It is probably this property and the efficiency with which it can be used that allow them to be used early in the infection. On the other hand, the incomplete, late-appearing VSG silent genes, which do not have competent 3′ ends, would be dependent on finding homology with the resident VSG gene and on other seemingly complex mechanisms for their successful integration into expression sites. This situation raises the question of why the partial genes have been maintained in the VSG gene repertoire. In other words, why aren't all the silent VSG genes complete? One reason for the maintenance of the incomplete silent copy genes might be that the use of mosaic genes permits the accumulation of mutations in the silent genes and also offers the selective advantage of generating more diversity per nucleotide than does the use of complete genes. Also, it is possible that genes that possess the sequences necessary for their integration into expression sites might be lost from the genome more rapidly than the incomplete genes due to accidents during variation events. Whatever the reason, it is clear that there has been a selection for the maintenance of incomplete silent VSG genes in *T. equiperdum*, since there appear to be many more late genes than early genes (9).

The use of complete genes for the early variants also offers a selective advantage to the parasites. In a natural infection, the initial inoculum is relatively small and the infected animal is healthy. Since the use of the complete genes is presumably more efficient (i.e., carried out with fewer steps) than that of the partial genes, a high proportion of variation events would be successful, thus allowing for the survival of the parasites in the infected host. At later stages of the infection, when the parasite load is high, and the host immune system more compromised, the inefficiency inherent in the use of the partial genes (i.e., the high probability of composing pseudogenes in the expression site) would no longer pose a threat to the survival of the parasites in the infected host.

Finally, our results offer a plausible explanation for African trypanosomes' use of the cassette mechanism for VSG gene activation. This is based on the similarity of the late-expressed VSG gene construction to that of immunoglobulin genes in lymphoid cells of vertebrates. In both cases, a relatively constant C-terminal region is joined to a variable N-terminal region. Also, in both cases the genes coding for the two regions are maintained in separate regions of the genome and brought together only for construction of the complete gene. In the immunoglobulin case this is done in specialized cells by deletion, and less frequently, by inversion. However, since the trypanosomes are unicellular, deletion would be difficult or impossible since genetic information would be lost at each variation event. Formally, duplication achieves the same purpose.

We thank Armand Berneman, Catherine Jacquemot, Tom St. John, and Janet Ajioka for expert technical assistance; Genevieve Thon for sharing unpublished data; and C. Nottenburg, John Swindle, Sue Sather, M. Groudine, H. Weintraub, and S. Brown for valuable discussions. This work was supported by grants from the Centre National de la Recherche Scientifique and the Institut National de la Santé et de la Recherche Médicale (France), the European Community, and the National Science Foundation. R.E.L. is supported by a fellowship from the Medical Research Council, Canada.

1. Donelson, J. E. & Rice-Ficht, A. C. (1985) *Micobiol. Rev.* **49**, 107–125.
2. Williams, R. O., Young, J. R. & Majiwa, P. A. O. (1979) *Nature (London)* **282**, 847–849.
3. Pays, E., Van Meirvenne, N., Le Ray, D. & Steinert, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2673–2677.
4. Young, J. A., Miller, N., Williams, R. & Turner, M. (1983) *Nature (London)* **306**, 196–198.
5. Bernards, A., De Lange, T., Michels, P., Liu, A., Huisman, M. J. & Borst, P. (1984) *Cell* **36**, 163–170.
6. Buck, G., Jacquemot, C., Baltz, T. & Eisen, H. (1984) *Gene* **32**, 329–336.
7. Buck, G., Longacre, S., Raibaud, A., Hibner, U., Giroud, C., Baltz, T., Baltz, D. & Eisen, H. (1984) *Nature (London)* **307**, 563–566.
8. Van Meirvenne, N., Janssens, P. G. & Magnus, E. (1975) *Soc. Belge Med. Trop.* **55**, 409–423.
9. Capbern, A., Giroud, C., Baltz, T. & Mattern, P. (1977) *Exp. Parasitol.* **42**, 6–13.
10. Myler, P., Allen, A., Agabian, N. & Stuart, K. (1985) *Infect. Immun.* **47**, 684–690.
11. Longacre, S. & Eisen, H. (1986) *EMBO J.* **5**, 1057–1063.
12. Roth, C., Longacre, S., Raibaud, A., Baltz, T. & Eisen, H. (1986) *EMBO J.* **5**, 1065–1070.
13. Longacre, S., Hibner, U., Raibaud, A., Eisen, H., Baltz, T., Giroud, C. & Baltz, D. (1983) *Mol. Cell. Biol.* **3**, 399–409.
14. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
15. Sanger, F., Coulson, A. R., Barell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
16. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–275.
17. Myers, R., Larin, Z. & Maniatis, T. (1985) *Science* **230**, 1242–1246.
18. Lui, A. Y. C., Van der Ploeg, L., Rijsewijk, F. & Borst, P. (1983) *J. Mol. Biol.* **167**, 57–75.
19. Lui, A. Y. C., Michels, P., Bernards, A. & Borst, P. (1985) *J. Mol. Biol.* **175**, 383–396.
20. Merritt, S. C., Tschudi, C., Konigsberg, W. H. & Richards, F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1536–1540.
21. Thon, G., Baltz, T. & Eisen, H. (1989) *Genes Dev.* **3**, 1247–1254.
22. Haas, R. & Meyer, T. (1986) *Cell* **44**, 107–115.
23. Swanson, J., Bergstrom, S., Robbins, K., Barrera, O., Corwin, D. & Koomey, J. M. (1986) *Cell* **47**, 267–276.
24. Raynaud, C.-A., Anquez, V., Grimal, H. & Weill, J. C. (1987) *Cell* **48**, 379–388.
25. Eisen, H., Longacre, S. & Buck, G. (1985) in *The Impact of Gene Transfer Techniques in Eukaryotic Biology*, ed. Starlinger, P. (Springer, Berlin), pp. 49–53.
26. Pays, E., Van Assel, S., Laurent, M., Darville, M., Vervoot, T., Van Meirvenne, N. & Steinert, M. (1983) *Cell* **34**, 371–381.
27. Pays, E., Houard, S., Pays, S., Van Assel, S., Dupont, F., Aerts, D. & Steinert, M. (1985) *Cell* **42**, 821–829.