Genomic environment of variant surface antigen genes of *Trypanosoma equiperdum*

(trypanosome/surface antigens/telomeres/base substitution)

A. RAIBAUD*, C. GAILLARD†‡, S. LONGACRE*, U. HIBNER*, G. BUCK*, G. BERNARDI†, AND H. EISEN*

*Unité d'Immunoparasitologie, Institut Pasteur, 28, Rue du Docteur Roux, 75015 Paris, France; and †Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 75005 Paris, France

ABSTRACT Expression of variant antigen genes in *Trypanosoma equiperdum* is accompanied by the duplication of a silent basic copy gene and the transposition of the copy to an expression site elsewhere in the genome. We have analyzed the genomic locations of both the basic and expression-linked copies of the *T. equiperdum* gene for variable surface glycoprotein VSG-1. Both copies are situated proximal to termini in both extracted DNA and in chromatin. The regions between the VSG-1 genes and the termini have a very high buoyant density in CsCl and contain an unidentified nucleoside that replaces deoxycytidine.

Parasitic African trypanosomes escape the mammalian host's immune response by changing their variable surface glycoprotein (VSG). In this process of antigenic variation, the sequential appearance of clones of trypanosomes expressing different VSGs allows the infection to persist in the immune-competent host. In both *Trypanosoma brucei* and *Trypanosoma equiperdum*, the expression of VSGs is accompanied by genomic rearrangements of the corresponding genes (1–5). In general, a silent basic copy gene is duplicated and transposed to a new genomic environment, and only this new expression-linked copy (ELC) is actively transcribed (4, 5). In other cases however, expression of *T. brucei* VSG-encoding genes, referred to throughout as VSG genes, is not coupled to the duplication of a basic copy but is accompanied by rearrangements downstream from the 3' end of the gene (2). In these cases non-expression-linked 3' rearrangements also occur.

T. equiperdum is spread among horses by venereal contact. Antigenic variation occurs in experimental infections of rabbits, and over 100 antigenically distinct clonally derived variable antigen types (VAT) have been described in such infections (6). The different VATs appear in a loosely defined order in rabbits and have been arbitrarily divided into early, middle, and late groups. Infection of naive rabbits with any VAT invariably results in the appearance of parasites expressing VSG-1 in the first relapse population, thus allowing the isolation and analysis of independent trypanosome clones expressing this antigen.

Using molecular clones of the VSG-1 gene we have analyzed the genomic rearrangements that accompany antigenic variation in *T. equiperdum* (5, 7). Two types of rearrangements have been observed. First, in each of four independently isolated trypanosome clones expressing VSG-1, there are ELCs of the VSG-1 gene. However, the duplicated, transposed copy is located in a different genomic environment, both 3' and 5' to the gene, in each VSG-1-expressing clone. Secondly, the genomic environment adjacent to the 3' end of the silent, basic copy of the VSG-1 gene is different in each of seven clonally derived VATs, independently of which VSG they express. The genome

of *T. equiperdum* also contains a family of apparently silent VSG-1-related sequences that do not undergo genomic rearrangements.

The genomic environments 3' to both the silent basic copy and the ELC of the VSG-1 gene exhibit notable features (5, 7). First, the basic copy and the ELCs of this gene behave as though their 3' ends were proximal to DNA termini. This was suggested by the presence of a "universal" restriction endonuclease site located 2–8 kilobases (kb) downstream from these VSG-1 genes. In addition, there are no restriction endonuclease sites between the genes and this universal site (5, 7). Similar results have been reported for *T. brucei* (2, 8, 9).

In this communication we confirm the presence of DNA termini downstream from the basic copy and ELC of the VSG-1 gene. We also show that these termini probably exist *in vivo*. Finally, it is shown that the DNA between the VSG-1 genes and the termini has an unusually high bouyant density and that deoxycytidine is partially or totally replaced by an as yet unidentified nucleoside in these regions.

MATERIALS AND METHODS

Trypanosomes. All of the *T. equiperdum* VATs used here were of the BoTAT (Bordeaux trypanosoon antigen type) serodeme and were clonally isolated from BoTAT-1. The genealogy of the various VATs has been described (6). Briefly, the number following BoTAT represents the VSG expressed and the superscript represents the parental clone. Trypanosomes were purified and their DNA was extracted as described (5). Nuclei were prepared as described (5) and were resuspended in BAL-31 buffer (see below).

Enzymes. Restriction endonucleases were from commercial sources and were used as directed by the supplier. Exonuclease BAL-31 was purchased from New England BioLabs.

Electrophoresis, Transfer, and Filter Hybridization. After digestion with restriction endonucleases, DNA was electrophoresed in agarose gels as described (5) and transferred to nitrocellulose paper by the method of Southern (10). VSG-1-specific cDNA containing plasmid pTel.2 (5) was ^{32}P-labeled by nick-translation (11) to about 10^8 cpm/μg. Filter hybridizations were done as described (5).

BAL-31 Nuclease Treatment. BAL-31 digestions were carried out at 30°C in 20 mM Tris·HCl, pH 8.0/100 mM NaCl/12 mM CaCl$_2$/1 mM EDTA. Reactions were terminated by addition of EDTA to a final concentration of 100 mM and by heating at 70°C for 15 min. The samples were then dialyzed against

Abbreviations: VSG, variable surface glycoprotein; VAT, variant antigen type; BoTAT, Bordeaux trypanosoon antigen type; ELC, expression-linked copy; kb, kilobase(s).
‡ Present address: Dept. of Biology, Massachusetts Institute of Technology, Cambridge, MA.

the appropriate restriction endonuclease buffer.

CsCl Gradients. *Sal* I-digested trypanosome DNA (120 μg) was brought to 14 ml in a solution containing 10 mM Tris·HCl (pH 7.5), 1 mM EDTA, and CsCl to give a final density of 1.700. Centrifugation was carried out at 20°C at 42,000 rpm for 60 hr in a Ti 50 rotor (Beckman). Fractions (0.2 ml) were collected from the bottom of the tube and dialyzed against 10 mM Tris·HCl, pH 7.5/1 mM EDTA or against 5 mM ammonium formate (pH 4.65) (for base composition analysis).

Analysis of Nucleoside Composition. Nucleoside compositions were determined by chromatography on Aminex A6 (Bio-Rad) of nucleosides generated by digestion of the DNA with hog spleen acid DNase, acid exonuclease, and acid phosphatase (12).

RESULTS

Restriction Maps of Basic Copies and ELCs of the VSG-1 Gene. Restriction endonuclease maps of the DNA located 3′ to the basic copies and ELCs of the VSG-1 gene from three nonexpressing and four expressing *T. equiperdum* clones are shown in Fig. 1. In no case were specific restriction endonuclease sites found downstream from the *Eco*RI site located 70 base pairs 3′ to the cDNA sequence but within the duplicated region. Instead, for each basic copy and ELC, there was a "universal" restriction site where all restriction endonucleases tested appeared to cut. The position of the universal site was different for each basic copy and ELC examined and varied between 3.2 and 8.7 kb from the *Sal* I site in the VSG-1 gene. These results suggest that the VSG-1 genes are located near DNA termini.

BAL-31 Nuclease Sensitivity. To determine whether the universal restriction sites located 3′ to the VSG-1 genes are in fact DNA termini, total DNA extracted from trypanosomes expressing VSG-1 (BoTAT-1[78]) was treated for various times with exonuclease BAL-31. After the BAL-31 treatment, the DNA was digested with *Sal* I and subjected to Southern blot analysis with ^{32}P-labeled VSG-1-specific plasmid pTel.2 as probe. As an internal control, *Eco*RI-digested bacteriophage λ Charon 4A DNA (13) was present during the BAL-31 digestion reaction. The extent of digestion of the phage λ fragments was monitored by agarose gel electrophoresis of an aliquot of each reaction before the restriction endonuclease was added. The results of these experiments are shown in Figs. 2 and 3. BAL-31 treatment resulted in the shortening of two VSG-1-specific bands
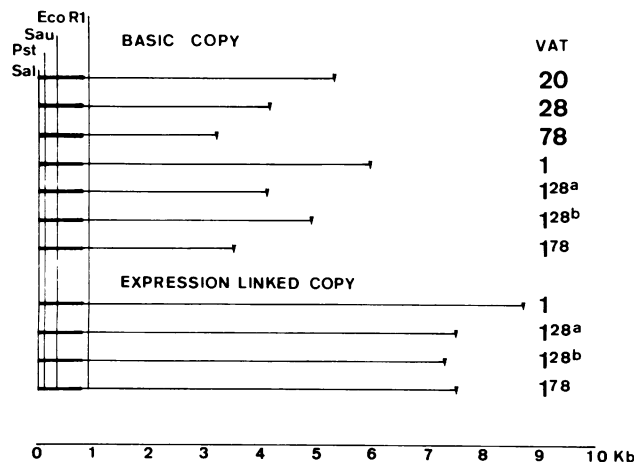

FIG. 2. Effects of BAL-31 treatment on VSG-1 genes. (*A*) BoTAT-1[78] DNA (3 μg) was treated sequentially with 0.9 units of BAL-31 (60 min at 30°C) and *Sal* I endonuclease as described (lane 2). The DNA was electrophoresed and subjected to Southern blot analysis with ^{32}P-labeled plasmid pTel.2 DNA. Lane 1 shows the hybridization profile of *Sal* I-digested DNA without prior BAL-31 treatment. The fragments containing the 3′ ends of the VSG-1 ELC and basic copy (BC) genes are denoted. (*B*) BoTAT-78 DNA (3 μg) was treated sequentially with 2 units of BAL-31 (120 min at 30°C) and endonuclease *Eco*RI (lane 2). Lane 1 shows the hybridization profile of *Eco*RI-digested DNA without prior BAL-31 treatment.

of 7.5 and 3.5 kb, representing the fragments containing the 3′ ends of the ELC and basic copy, respectively (Fig. 2*A*). The ELC was identified by its absence in VSG-1 nonexpressing trypanosomes and by its DNase I sensitivity (5, 7). The remaining VSG-1 hybridizing fragments, including those containing the 5′ ends of the ELC and basic copy, were not affected by the BAL-31 treatment. Substitution of the single-stranded endonuclease

FIG. 1. 3′ genomic environments of VSG-1 basic and ELC genes in DNA from four VSG-1-expressing and three nonexpressing trypanosome clones. The heavy lines represent the regions homologous to the cDNA sequence in plasmid pTel.2. Arrowheads show "universal" enzyme cleavage sites.
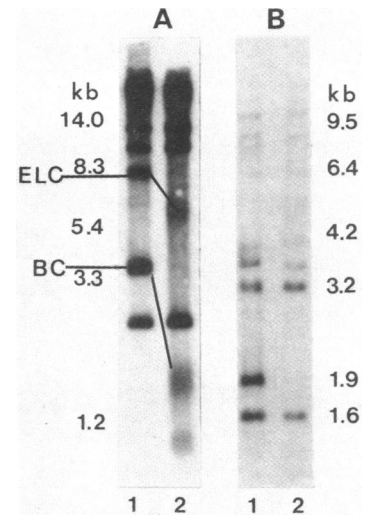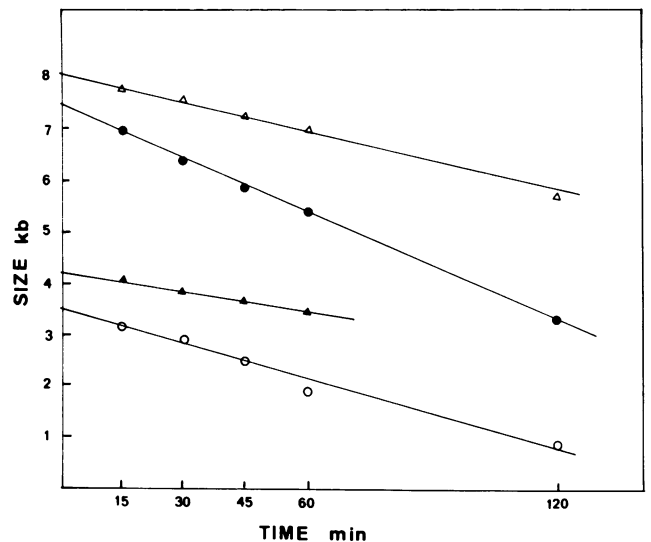

FIG. 3. Kinetics of BAL-31 digestion of VSG-1 basic and ELC genes, and of phage λ Charon 4A *Eco*RI fragments. BoTAT-1[78] DNA (30 μg) and *Eco*RI-digested phage λ Charon 4A DNA (11.2 μg) were treated with 9 units of BAL-31 at 30°C. At the indicated times, samples were removed and a portion of each sample was analyzed by agarose gel electrophoresis for determination of the digestion of the phage λ *Eco*RI fragments. The remainder of each sample was digested with *Sal* I, electrophoresed, and subjected to Southern blot analysis as in Fig. 2. ▲ and △, phage λ Charon 4A fragments; ●, 3′ end of the VSG-1 ELC gene; and ○, 3′ end of the VSG-1 basic copy gene.
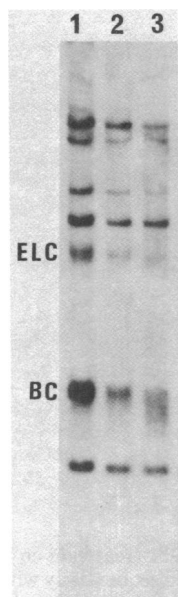
FIG. 4. BAL-31 treatment of BoTAT-1[78] nuclei. BoTAT-1[78] nuclei (2×10^8) were treated for 2.5 hr at 37°C with 0 (lane 1), 1 (lane 2), 9 (lane 3), and 45 (lane 4) units of BAL-31. The DNA was then extracted and treated with endonuclease *Sal* I, electrophoresed, and subjected to Southern blot analysis as described in Fig. 2.

S1 for BAL-31 did not alter any of the bands (not shown), indicating that the BAL-31 was not cutting at single-stranded regions.

We also examined BAL-31-treated trypanosome DNA after cleavage with endonuclease *Eco*RI. Previously, this laboratory has shown that, like the ELC, the putative VSG-1 basic copy gene is included in a unique 1.9-kb *Eco*RI fragment that is situated proximal to a presumptive DNA terminus (5, 7). To verify this, DNA from BoTAT-78 was treated with BAL-31, digested with endonuclease *Eco*RI, and subjected to Southern blot analysis as in Fig. 2A. The 1.9-kb hybridizing fragment was destroyed by BAL-31 treatment (Fig. 2B), demonstrating that this fragment and, thus, the VSG-1 basic copy gene are only present at a DNA terminus.

The kinetics of BAL-31 digestion of the DNA containing the

VSG-1 basic copy and ELC 3' flanking regions is shown in Fig. 3. The digestion rate was compared to that of phage λ *Eco*RI fragments of comparable sizes. Both of the VSG-1 gene-containing DNAs were more sensitive than were the phage λ fragments. If it is taken into account that the phage λ fragments were digested from both ends while the trypanosome DNA was digested only from one end, then the DNA in the 3' termini is 3- to 4-fold more sensitive to BAL-31 than is the phage λ DNA. It is also apparent that the DNA adjacent to the 3' end of the ELC was somewhat more sensitive than was the DNA next to the basic copy 3' end, suggesting that the two termini are not identical. The linear slopes of the digestion curves of the 3' VSG-1-gene-associated DNA, when extrapolated to zero time, showed their original sizes. This again indicates that BAL-31 digested the DNA from termini rather than from internal nicks or gaps.

The preceding results indicate that the 3' ends of the VSG-1 basic copy and ELC genes are proximal to termini in extracted DNA. However, they do not demonstrate that these termini exist *in vivo*. To test this possibility, nuclei were prepared from BoTAT-1[78] and treated with BAL-31. The DNA was then purified, digested with *Sal* I, and analyzed by the Southern blot technique as described above (Fig. 4). As with purified DNA, BAL-31 treatment of nuclei specifically shortened the fragments containing the 3' ends of the VSG-1 basic copy and ELC genes. Comparison of lanes 1 in Figs. 2A and 4 shows that the relative amount of the ELC-containing fragment was lower when the DNA was extracted from nuclei than when it was extracted directly from trypanosomes. This may have resulted from endogenous nucleases because we have shown previously that this fragment is preferentially degraded upon treatment of nuclei with nuclease. Nevertheless, because the size of the ELC-containing fragment derived from DNA extracted from nuclei was identical with that derived from trypanosomes, its shortening by BAL-31 must reflect its localization proximal to a terminus in the chromatin. Thus, both the ELC and the basic copy of the VSG-1 gene appear to be situated near chromosome telomeres.

Purification of the Terminal Fragments. The lack of restriction endonuclease sites and the hypersensitivity of the DNA situated between the VSG-1 genes and the termini suggested that it might have an unusual structure or contain modified bases. Therefore, it was reasoned that restriction endonuclease frag-
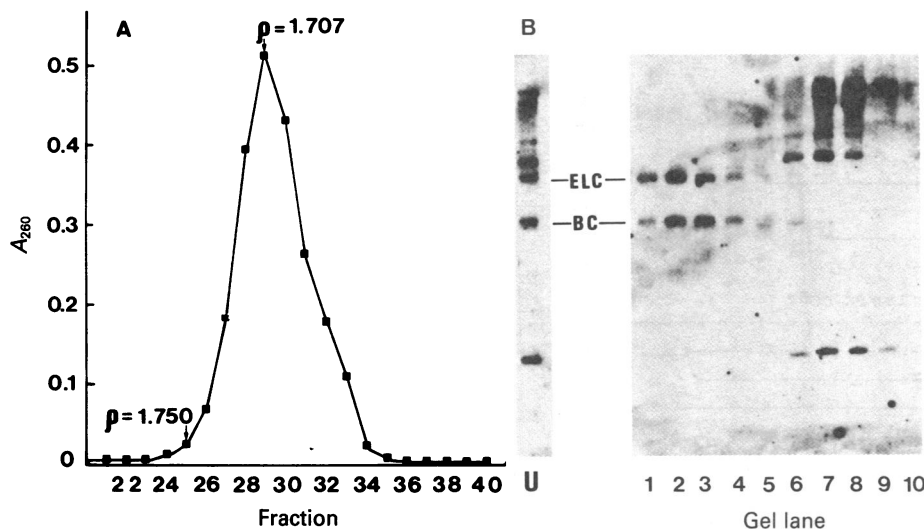

FIG. 5. CsCl density gradient analysis of *Sal* I-generated fragments of BoTAT-1 DNA. (*A*) A_{260} of gradient fractions. (*B*) Southern blot analysis of gradient fractions. U, unfractionated DNA. Lanes 1–10 show fractions 25–34 of the gradient. The fragments containing the 3' ends of the VSG-1 ELC and the basic copy (BC) genes are shown.

Biochemistry: Raibaud et al.
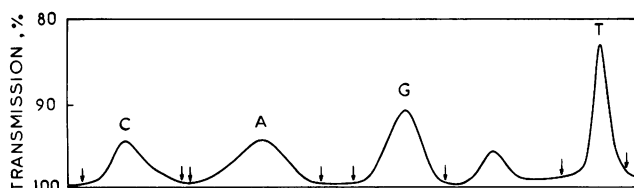
Proc. Natl. Acad. Sci. USA 80 (1983)   4309

FIG. 6. Chromatography of nucleosides obtained by enzymatic digestion from fractions 25–26 of *T. equiperdum* DNA. An unknown nucleoside is eluted between the T and G peaks.

ments containing these regions might be separated from the bulk DNA in isopycnic CsCl gradients. DNA from BoTAT-1 was digested with *Sal* I and centrifuged to equilibrium in CsCl gradients. Fractions were collected, dialyzed, and subjected to Southern blot analysis for VSG-1 hybridizing sequences. The fragments containing the 3' ends of the basic copy and the ELC were found in the most dense fractions of the gradients, separated from all other VSG-1 hybridizing material, including the fragments containing the 5' ends of the same genes (Fig. 5).

To investigate the possibility that the DNA between the VSG-1 genes and the termini contained a modified or substituted base, fractions from three independent CsCl gradients of *Sal* I-digested BoTAT-1 DNA were digested to nucleosides and analyzed. Unfractionated BoTAT-1 DNA also was examined. The results are presented in Fig. 6 and Table 1. The DNA corresponding to fraction 29 of the CsCl gradient in Fig. 5 had a G:C ratio of 0.99. However, the DNA corresponding to fractions 25 and 26 had a G:C ratio of 1.48. Thus, there was a 32% deficit of deoxycytidine in this material. An extra peak was present in the chromatograph of this material in an amount (if one assumes a molar absorption similar to that of deoxycytidine) compatible with the idea that the corresponding, as yet unidentified, nucleoside replaces deoxycytidine in this DNA. There was also a 6% deficit in deoxycytidine in the unfractionated *T. equiperdum* DNA. This was accompanied by the presence of the new nucleoside in equivalent amounts.

DISCUSSION

We have reported results that suggest that the 3' ends of the VSG-1 basic copy and ELC genes of *T. equiperdum* are proximal to termini in purified DNA and *in vivo*. Similar results have been presented for *T. brucei* (14, 15). Both the kinetic analysis of BAL-31 digestion and the finding that the single-strand-specific endonuclease S1 did not shorten the termini suggest that, if these structures have single-stranded nicks or gaps as has been shown for other telomeric structures (16), then the nicks or gaps are located very close to the termini. We cannot, however, rule out the possibility that endonuclease S1 does not digest DNA containing the unidentified base.

The DNA located between the VSG-1 genes and the termini has unusual properties. First, no restriction endonuclease sites

Table 1. Nucleoside composition of CsCl fractions of *Sal* I-treated BoTAT-1 DNA

DNA	dT, %	dG, %	dA, %	dC, %	X,* %	dG+dC+X, %
Unfractionated	28.3	21.8	28.1	20.5	1.3	43.6
Fractions 25 and 26	26.5	24.8	25.5	16.8	8.0	49.6
Fraction 27	26.7	24.3	26.1	21.1	3.2	48.6
Fraction 29	28.9	22.2	27.0	21.9	0	44.1

*Unidentified nucleoside: percentage estimated by assuming a molar absorption similar to that of deoxycytidine.

were found in these regions even though they are up to 8.7 kb long. Second, these regions, which contain an average of 51% A+T, have a buoyant density in CsCl of approximately 1.75. Finally, base composition analysis of the most enriched fractions revealed that 32% of the expected deoxycytidine was replaced by an as yet unidentified nucleoside. Because the purity of the VSG-1 gene-associated terminal fragments could not be determined, we cannot estimate the extent of replacement of deoxycytidine in these regions. However, the lack of restriction endonuclease sites in these regions would suggest that a high percentage of deoxycytidine is replaced. If the lack of restriction endonuclease sites in the 3' terminal regions is an indication of the presence of the modified nucleoside, then the substitution is not present in the VSG-1 gene coding sequences because all of the endonuclease cleavage sites predicted from the cDNA sequence were found to be sensitive to digestion. This suggests that the unidentified nucleoside is formed by postreplicative modification of deoxycytidine.

The CsCl gradient fractions with the most purified VSG-1-associated DNA termini (25 and 26) contained approximately 1.25% of the DNA put on the gradient (120 μg). The total nuclear DNA content of *T. equiperdum* is about 7×10^4 kb (17). Because the fragments containing the 3' ends of the VSG-1 gene have an average length of 6.5 kb, they represent approximately 0.01% of the total DNA. Assuming that fractions 25 and 26 contain 50% of the VSG-1-associated ends, we estimate that these fragments account for approximately 1.0% of the DNA in the purified fractions. Because this does not account for all of the substituted bases in these fractions, it is probable that other substituted base-containing sequences are present. These could include DNA termini associated with other VSG genes. It is also possible that all of the telomeres of *T. equiperdum* contain the substituted base.

Although the nucleoside that replaces deoxycytidine in the terminal fragments is as yet unidentified, it is certainly not deoxymethylcytidine, which elutes from Aminex A6 at a position close to that of deoxycytidine. Also, the presence of glycosyl groups is unlikely because these would render the DNA resistant to the spleen exonuclease used for the digestion of the DNA. Although some *Bacillus subtilis* phages have been shown to contain deoxyuracil or deoxyhydroxyuracil, both of which increase the buoyant density of the DNA, these nucleosides always replace thymidine (18). It is highly unlikely that these nucleosides would pair with deoxyguanosine in naturally occurring DNA. The finding that the DNA between the VSG genes and the termini is hypersensitive to exonuclease BAL-31 suggests that the novel nucleoside might form a two-hydrogen-bond pair with deoxyguanosine because it has been shown that BAL-31 has a strong preference for A·T base pairs (19). The presence of the unknown base might not be the only factor responsible for the extremely high buoyant density of the VSG-1-associated termini. Other factors, such as the presence of short repeated sequences, have been shown to influence the buoyant density of DNA (20, 21).

The regions flanking the 3' ends of the VSG genes in trypanosomes may play a role in regulating expression of the genes. Expression of some VSG genes in *T. brucei* is accompanied by rearrangements of DNA near the 3' ends of the genes without duplication of the basic copy genes (8). For these VSGs no rearrangements of sequences at the 5' ends of the genes have been described. In *T. equiperdum* we have observed that after direct variation of VAT-1 to other early VATs, the ELC of the previously expressed VSG-1 gene is not destroyed, although it is no longer expressed (unpublished data). Rather, its 3' flanking region is changed. No changes in the positions of restriction endonuclease sites located 5' to the gene were found. If 5' rear-

rangements undetected by restriction endonuclease mapping are not present, these findings indicate that, in some manner, the 3' flanking regions may play a role in controlling the expression of these VSG genes.

All of the ELCs of trypanosome VSG genes thus far described appear to be located near telomeric structures, even when the corresponding basic copy genes are not near telomeres (2, 5, 8, 9). This finding raises the possibility that the telomeric sequences could also play a role in the duplication–transposition event leading to the formation of the ELC. The presence of the new nucleoside in these sequences could be important in determining the specificity of the enzymes that carry out this process.

1. Williams, R., Young, J. & Majiwa, P. (1979) *Nature (London)* **282**, 847–849.
2. Hoeijmakers, J., Frash, A., Bernards, A., Borst, P. & Cross, G. (1980) *Nature (London)* **284**, 78–80.
3. Pays, E., Van Meirvenne, N., LeRay, D. & Steinert, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2673–2677.
4. Pays, E., Lheureux, M. & Steinert, M. (1981) *Nature (London)* **292**, 265–267.
5. Longacre, S., Hibner, U., Raibaud, A., Eisen, H., Baltz, T., Giroud, C. & Baltz, D. (1983) *Mol. Cell. Biol.* **3**, 399–409.
6. Capbern, A., Giroud, C., Baltz, T. & Mattern, P. (1977) *Exp. Parasitol.* **42**, 6–13.
7. Longacre, S., Raibaud, A., Hibner, U., Buck, G., Eisen, H., Baltz, T., Giroud, C. & Baltz, D. (1983) *Mol. Cell. Biol.* **3**, 410–414.
8. Williams, R., Young, J., Majiwa, P., Doyle, J. & Shapiro, S. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 945–950.
9. Pays, E., Lheureux, M. & Steinert, M. (1982) *Nucleic Acids Res.* **10**, 3149–3163.
10. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
11. Rigby, P., Diekemann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
12. Thiery, J., Ehrlich, S., Devillers-Thiery, A. & Bernardi, G. (1973) *Eur. J. Biochem.* **38**, 434–442.
13. Blattner, F., Williams, B., Blechl, A., Denniston-Thompson, K., Faber, H., Furlong, L., Grunwald, D., Kiefer, D., Moore, D., Schumm, J., Sheldon, E. & Smithies, O. (1977) *Science* **196**, 161–169.
14. Williams, R., Young, J. & Majiwa, P. (1982) *Nature (London)* **299**, 417–421.
15. De Lange, T. & Borst, P. (1982) *Nature (London)* **299**, 451–453.
16. Blackburn, E. & Gall, J. (1978) *J. Mol. Biol.* **120**, 33–53.
17. Riou, G., Baltz, T., Gabillot, M. & Pautrizel, R. (1980) *Mol. Biochem. Parasitol.* **1**, 97–105.
18. Takahashi, M. & Marmur, J. (1963) *Nature (London)* **197**, 794–796.
19. Katinka, M. & Yaniv, M. (1982) *Mol. Cell. Biol.* **2**, 1238–1246.
20. Wells, R. & Blair, J. E. (1967) *J. Mol. Biol.* **27**, 273–288.
21. Corneo, G., Ginelli, E., Soave, C. & Bernardi, G. (1968) *Biochemistry* **7**, 4373–4379.