MOLECULAR AND CELLULAR BIOLOGY, Sept. 1989, p. 4022–4025
0270-7306/89/094022-04$02.00/0

Vol. 9, No. 9

Trypanosoma brucei: Enrichment by UV of Intergenic Transcripts from the Variable Surface Glycoprotein Gene Expression Site

HÉLÈNE COQUELET, PATRICIA TEBABI, ANNETTE PAYS, MAURICE STEINERT, AND ETIENNE PAYS*

Department of Molecular Biology, University of Brussels, 67, rue des Chevaux, B1640 Rhode-Saint-Genèse, Belgium

Received 21 March 1989/Accepted 14 June 1989

The expression site for the variable surface glycoprotein (VSG) gene AnTat 1.3A of *Trypanosoma brucei* is 45 kilobases long and encompasses seven expression site-associated genes (ESAGs) (E. Pays, P. Tebabi, A. Pays, H. Coquelet, P. Revelard, D. Salmon, and M. Steinert, Cell 57:835–845, 1989). After UV irradiation, several large transcripts from the putative promoter region were strongly enriched. We report that one such major transcript starts near the poly(A) addition site of the first gene (ESAG 7), spans the intergenic region, and extends to the poly(A) addition site of the second gene (ESAG 6), thus bypassing the normal 3′ splice site of the ESAG 6 mRNA. Since this transcript is spliced, we conclude that UV irradiation does not inhibit splicing but stabilizes unstable processing products. This demonstrates that at least some intergenic regions of the VSG gene expression site are continuously transcribed in accordance with a polycistronic transcription model.

In African trypanosomes, antigenic variation is achieved by frequent changes of the variable surface glycoprotein (VSG). Different VSG genes can be expressed, but only one is usually transcribed at a time, and different switching mechanisms have been described (for recent reviews, see references 3, 19, and 26). The transcribed VSG gene is located at the end of a telomeric expression site, which also contains other genes (expression site-associated genes [ESAGs]) and is repeated in 5 to 20 slightly different versions in the genome. This transcription exhibits several characteristics: (i) it resists inhibition by α-amanitin; (ii) it is preferentially stopped by a lowering of temperature; and (iii) it seems to start upstream from the ESAGs, so that several genes may belong to the same transcription unit (1, 10, 19a, 23). The latter characteristic has also been reported for other genes, such as tubulin, calmodulin, and actin genes and genes of the glycolytic pathway, suggesting that in *Trypanosoma brucei*, transcription in general might be polycistronic (2, 5, 8, 9, 25). However, no direct evidence of the existence of polycistronic transcripts has yet been provided. The trypanosome genes appear devoid of introns, but splicing of the RNA precursors occurs; a 39-nucleotide sequence called the miniexon is cleaved from a short precursor and is added at the 5′ termini of all mature mRNAs, irrespective of whether these RNAs are transcribed from the same chromosome as the miniexon (7, 11, 16, 24). This splicing appears necessary for RNA translation (4, 27).

In an attempt to map the transcription promoter of the AnTat 1.3A VSG gene expression site, we have measured the relative sensitivity of its transcription to inhibition by UV irradiation, as was done for another VSG gene expression site by Johnson et al. (9). As reported previously (19a), we found that transcription of a particular region, located about 45 kilobases (kb) upstream from the VSG gene, appears to be selectively and strongly stimulated by UV light. This apparent stimulation also characterizes the promoter region of the rDNA transcription unit (19a). Since it is correlated with an enrichment of large transcripts interpreted as processing intermediates, we concluded that UV irradiation inhibits RNA processing (19a), in addition to its known effect on RNA elongation (22). Near the promoter, where the inhibition of RNA elongation is minimal, the effect of UV irradiation on RNA processing would account for the apparent stimulation of transcription.

In this paper, we detail the effect of UV irradiation on RNA processing, which appears to be a stabilization of intergenic region-specific transcripts. This provides evidence for transcription through intergenic regions in the beginning of the VSG gene expression site.

MATERIALS AND METHODS

The *T. brucei* AnTat 1.3A gene clone has been characterized elsewhere (12). The procedures for DNA and RNA isolation, Southern and Northern (RNA) blot hybridization, as well as DNA cloning, were as described (1, 18). cDNA libraries were constructed in lambda gt10 according to Gubler and Hoffman (6) with the Amersham cDNA synthesis and cloning kits. The nucleotide sequence of DNA fragments, subcloned in bacteriophage M13 derivatives, was determined on both strands by the method of Sanger et al. (21), with a modified T7 DNA polymerase (Sequenase; United States Biochemical Corp.). Run-on transcription assays were conducted as described by Murphy et al. (15). UV irradiation was performed under the conditions defined by Johnson et al. (9). Briefly, the blood containing trypanosomes was diluted in Baltz medium at 37°C to reach a parasite concentration of 1 to 4.10^7/ml. Samples (125 ml) were irradiated at 254 nm (1 J/s per m^2) in sterile square dishes (22 × 22 cm) (Bio-Assay, Nunc, Roskilde, Denmark) usually for 1 min and under agitation. The irradiated cells were then transferred to culture flasks and kept in the dark until centrifugation.

RESULTS AND DISCUSSION

UV irradiation leads to considerable enrichment of some transcripts from the promoter region of the VSG gene expression site (Fig. 1). We took advantage of this enrichment to clone the corresponding cDNAs. The extent of five such molecules (C1 to C5) is shown under the map in Fig. 1, together with that of the ESAG full-size cDNAs. The different cDNAs partially overlap and cover virtually the whole genomic region, and their sequence perfectly matches the corresponding ones from the cloned expression site, while

* Corresponding author.

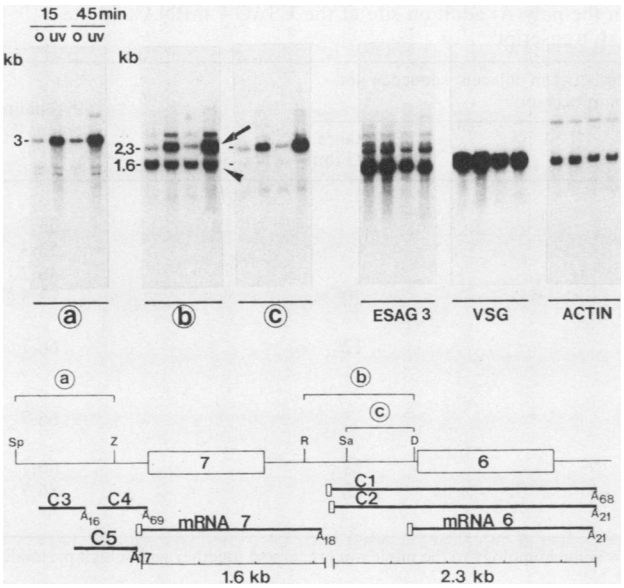

FIG. 1. UV-induced accumulation of some transcripts from the *Sph*I region of the AnTat 1.3A VSG gene expression site. After UV irradiation, unstable transcripts with large nongenic sequences strongly and selectively accumulate. Northern blots of 5 μg of polyadenylated RNA from cloned AnTat 1.3A trypanosomes were hybridized with different probes from the VSG gene expression site and with the actin probe as a control. The trypanosomes were untreated (lane o) or irradiated for 1 min with 1 J/s per m^2 at 254 nm (lane uv), before incubation for 15 or 45 min at 37°C as described previously (9). The order of lanes for each gel is as indicated on top of the first gel. The 2.3-kb transcript (indicated by arrow) is enriched by UV, while the 1.6-kb mature ESAG 6 (and ESAG 7) mRNA (indicated by arrowhead) is not. Both transcripts have been cloned as cDNAs, the former from UV-irradiated trypanosomes and the latter from untreated cells. Two independent 2.3-kb cDNAs (C1 and C2), as well as the 1.6-kb ESAG 7 and ESAG 6 cDNAs (19a) are represented under the map, with open boxes for the miniexon sequence. The extent of each of the three other cDNAs from UV-irradiated trypanosomes (C3 to C5) is also shown below the map. The length of the poly(A) tail of each cDNA is indicated. The open boxes in the map represent the ESAG open reading frames. The ESAG 3 and VSG sequences are respectively about 25 and 45 kb downstream from *Sph*I. The exposure time for the VSG panel is 10-fold less than that for the others. Abbreviations: D, *Dra*I; R, *Rsa*I; Sa, *Sau*3A; Sp, *Sph*I; Z, *Ssp*I; a, b, and c, probes (the extent of each is indicated on the map).

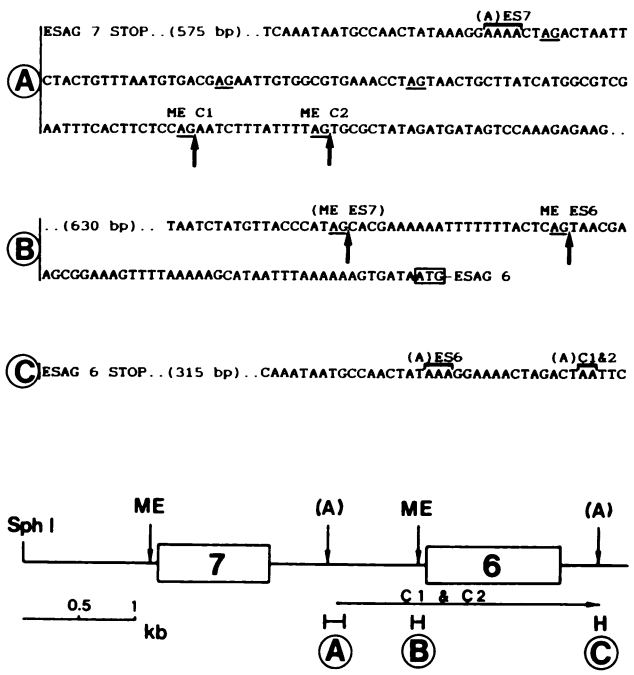

FIG. 2. Nucleotide sequence of two 2.3-kb cDNAs specific to ESAG 6. The sequences of two full-size cDNAs of the 2.3-kb transcript (C1 and C2) were determined and found to be identical to the corresponding one from the expression site (ESAG 6 and 0.7 kb of upstream region). Three relevant portions of the latter sequence are shown and marked A, B, and C; the extent of each is indicated on the map. The polyadenylation sites of ESAG 7 and ESAG 6 mRNAs, as well as those of C1 and C2, are indicated above the sequence (A under the brackets), while the corresponding 3′ splice sites (miniexon addition sites [ME]) are indicated by arrows. The 3′ splice site of the ESAG 7 mRNA could also be indicated for comparison, the nucleotide sequence in this region being the same for ESAG 6 and ESAG 7. The 5′ extremity of the majority of the 2.3-kb transcripts, as determined by S1 mapping (data not shown), is that of C2. The AG dinucleotides referred to in Table 1 are underlined, and the initiation codon of ESAG 6 is boxed. Note that, despite identical sequences downstream from ESAG 7 and ESAG 6, the polyadenylation sites of ESAG 7, ESAG 6, and 2.3-kb RNAs are different. The open boxes in the map represent the ESAG open reading frames.

homologous sequences from other telomeres show differences (Fig. 2; data not shown). This confirms previous information (17, 19a) that the latter appears to be the only telomere transcribed among its 5 to 20 genomic family members and rules out the possibility that the stimulating effect of UV light could be caused by the activation of other potential expression sites. Figure 2 shows the nucleotide sequence of two independent full-size cDNAs (C1 and C2) of a 2.3-kb transcript which becomes particularly abundant after UV irradiation (Fig. 1). This transcript starts about 90 base pairs (bp) downstream from the poly(A) addition site of the first gene (ESAG 7) and extends to the poly(A) addition site of the second gene (ESAG 6). The two 2.3-kb cDNAs analyzed do not start exactly at the same position, but both are capped with a miniexon sequence. That the miniexon can be added at several possible locations of the same sequence has already been shown by Layden and Eisen (13). On the basis of the observations by these authors, the 3′ splice sites observed here appeared to be the first possible ones downstream from the polyadenylation region of ESAG 7 mRNA (Table 1). Thus, the 2.3-kb transcripts are about 0.7 kb larger than the mature ESAG 6 mRNA because of splicing at unusual locations far upstream from ESAG 6 but as close as possible to the end of the ESAG 7 mRNA. Since the 2.3-kb transcripts are natural processing products which can be detected in RNA from unirradiated cells (Fig. 1), and as their accumulation does not occur at the expense of the production of mature mRNA (Fig. 1), we suggest that the effect of UV irradiation is not interference with splicing but the stabilization of unstable processing products. This conclusion is also in agreement with our observation that the effect of UV irradiation cannot be mimicked by heat shock (E. Pays and H. Coquelet, unpublished data), as heat shock is known to inhibit splicing (14, 28).

From the structure of the 2.3- and 1.6-kb transcripts, we can conclude that at least 1,245 bp from the 1,335-bp region between the ESAG 7 and ESAG 6 open reading frames are transcribed. The apparently untranscribed 90-bp stretch is not likely to be a promoter region; since the 2.3-kb transcript

TABLE 1. Data indicating no potential 3′ splice regions between the poly(A) addition site of the ESAG 7 mRNA and the beginning of the 2.3-kb transcript[a]

AG dinucleotide and miniexon addition sites	Presence of complementarity between adjacent sequence and miniexon precursor[b]:		Pyrimidine content (%)[b]
	Sequence	Distance to AG (bp)	
AG dinucleotides[c]			
+4	None found		33.3
+31	None found		53.3
+51	None found		46.7
C1 (+88)	GCUUAUCAUGGCGUC CGAAGAGUAUGGUUA	20	73.3
C2 (+102)	ACUUCUCCAGAAUGU CGAAGAGUAUGGUUA	12	66.7
Miniexon addition sites			
ESAG 7[d]	AUCUAUGUUACCCAU CGAAGAGUAUGGUUA	5	66.7
ESAG 6[d]	Same as for ESAG 7	28	66.7
ESAG 5[d]	ACUCCGCACAUCUUU CGAAGAGUAUGGUUA	28	86.7

[a] As shown here for C1, C2, ESAGs 7, 6, and 5, the 3′ splice regions generally show some homology to the miniexon precursor, together with a high pyrimidine content (13). C1 and C2 are the two 2.3-kb cDNAs shown in Fig. 1 and 2.

[b] Estimated as in reference 13.

[c] AG sites are given as distances downstream in base pairs from the ESAG 7 poly(A) addition site. Sites +88 and +102 represent sites of cDNAs C1 and C2, respectively.

[d] cDNA sequences presented by Pays et al. (19a).

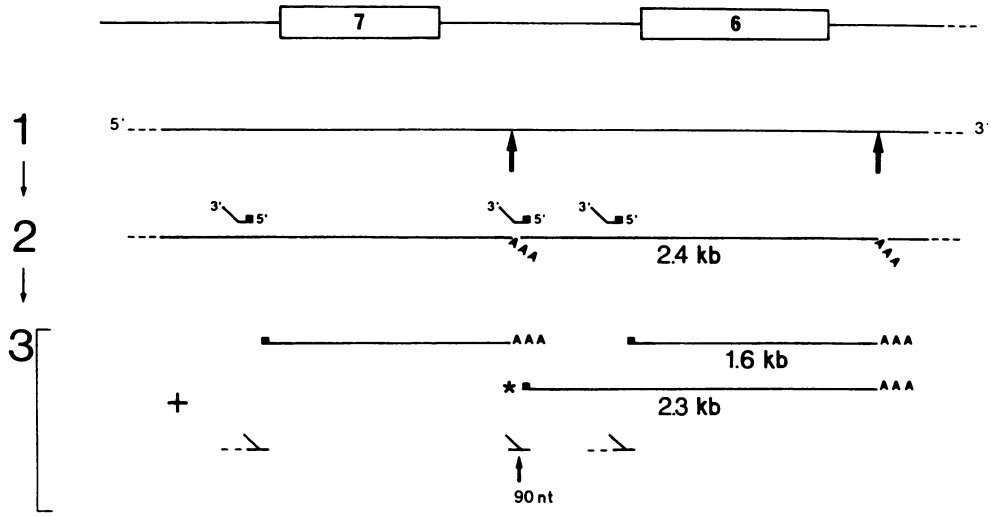

FIG. 3. Model of transcription and processing at the beginning of the VSG gene transcription unit. (1) The primary transcript spans both ESAGs 7 and 6. (1 to 2) The first processing event would be cleavage leading to polyadenylation (arrows). A 2.4-kb transcript is released. (2 to 3) Splicing at two different sites of the 2.4-kb transcript gives rise to the 1.6-kb mature mRNA for ESAG 6 and to an unstable 2.3-kb transcript (★), which accumulates strongly after UV irradiation. This splicing leads to the loss of a 90-nucleotide (nt) branched RNA intermediate (indicated by arrow), which is the only intergenic region we did not find in the cDNAs. The solid boxes represent the 39-nucleotide miniexon sequence. The open boxes at the top represent ESAG open reading frames.

is spliced, any RNA sequence immediately upstream from the 3′ splice site would be degraded as a branched intermediate. The substrate for this splicing could be a 2.4-kb transcript starting at the cleavage site for polyadenylation of ESAG 7 and ending at the polyadenylation site of ESAG 6 (Fig. 3). In support of this model, the 2.3-kb transcript is spliced at the nearest possible location to the poly(A) addition site of ESAG 7, and the length of the transcription gap (90 nucleotides) is within the size range of the branched RNA intermediates released by 3′ splicing (20). This strongly suggests that the primary transcript spans both ESAGs 7 and 6 and supports a model of polycistronic transcription in the VSG expression site (9, 19a).

ACKNOWLEDGMENTS

We acknowledge S. Alexandre and G. Hilgers for help in the S1 mapping and UV irradiation experiments, respectively, and D. Jefferies for comments on the manuscript.

This work has been supported by the Belgian "Fonds de la Recherche Scientifique Médicale" and research contracts between the University of Brussels and Solvay & Cie (Brussels) and between

the University of Brussels and the Commission of the European Communities (TDS-M-023B).

LITERATURE CITED

1. **Alexandre, S., M. Guyaux, N. B. Murphy, H. Coquelet, A. Pays, M. Steinert, and E. Pays.** 1988. Putative genes of a variant-specific antigen gene transcription unit in *Trypanosoma brucei*. Mol. Cell. Biol. **8**:2367–2378.
2. **Ben Amar, M. F., A. Pays, P. Tebabi, B. Dero, T. Seebeck, M. Steinert, and E. Pays.** 1988. Structure and transcription of the actin gene of *Trypanosoma brucei*. Mol. Cell. Biol. **8**:2166–2176.
3. **Borst, P.** 1986. Discontinuous transcription and antigenic variation in trypanosomes. Annu. Rev. Biochem. **55**:701–732.
4. **Cornelissen, A. W. C. A., M. P. Verspieren, J. J. Toulme, B. W. Swinkels, and P. Borst.** 1986. The common 5′ terminal sequence on trypanosome mRNAs: a target for anti-messenger oligodeoxynucleotides. Nucleic Acids Res. **14**:5605–5614.
5. **Gibson, W. C., B. W. Swinkels, and P. Borst.** 1988. Posttranscriptional control of the differential expression of phosphoglycerate kinase genes in Trypanosoma brucei. J. Mol. Biol. **201**: 315–325.
6. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
7. **Guyaux, M., A. W. C. A. Cornelissen, E. Pays, M. Steinert, and P. Borst.** 1985. A surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. EMBO J. **4**:995–998.
8. **Imboden, M. A., P. W. Laird, M. Affolter, and T. Seebeck.** 1987. Transcription of the intergenic regions of the tubulin gene cluster of Trypanosoma brucei: evidence for a polycistronic transcription unit in a eukaryote. Nucleic Acids Res. **15**:7357–7368.
9. **Johnson, P. J., J. M. Kooter, and P. Borst.** 1987. Inactivation of transcription by UV irradiation of Trypanosoma brucei provides evidence for a multicistronic transcription unit that includes a variant surface glycoprotein gene. Cell **51**:273–281.
10. **Kooter, J. M., H. J. van der Spek, R. Wagter, C. E. d'Oliveira, F. van der Hoeven, P. J. Johnson, and P. Borst.** 1987. The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in Trypanosoma brucei. Cell **51**:261–272.
11. **Laird, P. W., J. C. B. M. Zomerdijk, D. de Korte, and P. Borst.** 1987. In vivo labelling of intermediates in the discontinuous synthesis of mRNAs in Trypanosoma brucei. EMBO J. **6**: 1055–1062.
12. **Laurent, M., E. Pays, E. Magnus, N. Van Meirvenne, G. Matthyssens, R. O. Williams, and M. Steinert.** 1983. DNA rearrangements linked to the expression of a predominant surface antigen gene of trypanosomes. Nature (London) **302**: 263–266.
13. **Layden, R. E., and H. Eisen.** 1988. Alternate *trans* splicing in *Trypanosoma equiperdum*: implications for splice site selection. Mol. Cell. Biol. **8**:1352–1360.
14. **Muhich, M. L., and J. C. Boothroyd.** 1988. Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in messenger RNA synthesis. Mol. Cell. Biol. **8**:3837–3846.
15. **Murphy, N. B., A. Pays, P. Tebabi, H. Coquelet, M. Guyaux, M. Steinert, and E. Pays.** 1987. Trypanosoma brucei repeated element with unusual structural and transcriptional properties. J. Mol. Biol. **195**:855–871.
16. **Murphy, W. J., K. P. Watkins, and N. Agabian.** 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. Cell **47**:517–525.
17. **Pays, E., H. Coquelet, A. Pays, P. Tebabi, and M. Steinert.** 1989. *Trypanosoma brucei*: posttranscriptional control of the variable surface glycoprotein gene expression site. Mol. Cell. Biol. **9**:4018–4021.
18. **Pays, E., M. Delronche, M. Lheureux, T. Vervoort, J. Bloch, F. Gannon, and M. Steinert.** 1980. Cloning and characterization of DNA sequences complementary to messenger ribonucleic acids coding for the synthesis of two surface antigens of Trypanosoma brucei. Nucleic Acids Res. **8**:5965–5981.
19. **Pays, E., and M. Steinert.** 1988. Control of antigen gene expression in African trypanosomes. Annu. Rev. Genet. **22**:107–126.

19a. **Pays, E., P. Tebabi, A. Pays, H. Coquelet, P. Revelard, D. Salmon, and M. Steinert.** 1989. The genes and transcripts of an antigen gene expresson site from T. brucei. Cell **57**:835–845.

20. **Ralph, D., J. Huang, and L. H. T. Van der Ploeg.** 1988. Physical identification of branched intron side-products of splicing in Trypanosoma brucei. EMBO J. **7**:2539–2545.
21. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
22. **Sauerbier, M., and K. Herculez.** 1978. Genes and transcription unit mapping by radiation effects. Annu. Rev. Genet. **12**: 329–363.
23. **Shea, C., M. G. S. Lee, and L. H. T. Van der Ploeg.** 1987. VSG gene 118 is transcribed from a co-transposed pol-I like promoter. Cell **50**:603–612.
24. **Sutton, R. E., and J. C. Boothroyd.** 1986. Evidence for trans splicing in trypanosomes. Cell **47**:527–535.
25. **Tschudi, C., and E. Ullu.** 1988. Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. EMBO J. **7**: 455–463.
26. **Van der Ploeg, L. H. T.** 1987. Control of variant surface antigen switching in trypanosomes. Cell **51**:159–161.
27. **Walder, J. A., P. S. Eder, D. M. Engman, S. T. Brentano, R. Y. Walder, D. S. Knutzon, D. M. Dorfman, and J. E. Donelson.** 1986. The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNA's. Science **233**:569–571.
28. **Yost, H. J., and S. Lindquist.** 1986. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell **45**:185–193.