A Family of Genes Related to a New Expression Site-Associated Gene in *Trypanosoma equiperdum*

ISABELLE C. FLORENT,^{1,2*} ANNA RAIBAUD,^{2†} AND HARVEY EISEN^{1,2}

Fred Hutchinson Cancer Research Center, M-723 Genetics, 1124 Columbia Street, Seattle, Washington 98104,^{1} and Immunoparasitologie, Institut Pasteur, 75724 Paris Cedex 15, France²*

Received 4 September 1990/Accepted 24 January 1991

Two genes, belonging to a new expression site-associated gene family of six to eight members in *Trypanosoma equiperdum* and *Trypanosoma brucei*, have been cloned from a *T. equiperdum* variant. One of them, called *ESAG-9c*, is contained in the 1.78-C expression site and is found just upstream of the 5' barren region. The other one, called *ESAG-9u*, is unique in the family, is not telomere linked, and apparently is not expression site related. A 2-kb poly(A)⁺ mRNA is detected with probes for this *ESAG-9* family in all *T. equiperdum* variants examined. By using polymerase chain reaction and restriction fragment length polymorphism techniques, it has been possible to distinguish between *ESAG-9c* and *ESAG-9u* and to show that *ESAG-9c* is transcribed in an expression site-specific manner. However, *ESAG-9u* (or another gene in the family having identical characteristics) is transcribed in all variants, regardless of the expression site used by these variants. Thus, this *ESAG-9* family contains at least one gene that is under expression site control but might have other genes that are not. The function of these *ESAG-9* genes is unknown. Transcripts homologous to *ESAG-9* were detected in *T. brucei* bloodstream forms but not in procyclics.

African trypanosomes are extracellular hemoflagellate parasites, responsible for sleeping sickness in humans and other diseases in wild and domestic animals (54). These trypanosomes regularly change their variant surface glycoprotein (VSG) coat, and this phenomenon, called antigenic variation, enables them to escape the immune response of the infected host (for reviews, see references 7, 9, 19, and 40). Despite extensive studies, the molecular basis for antigenic variation is only partially understood. VSGs are encoded by silent genes that must go through several processes before being expressed. The first step in VSG gene activation usually involves the duplication and transposition of the gene into one of several telomere-linked expression sites (16, 32, 35). However, VSG genes present in expression sites can be activated and inactivated *in situ* without detectable gene rearrangements (6) or by telomere exchange (39). Molecular analysis of the transcriptional activation of the VSG gene has focused on determining the start of transcription in the expression site (1, 27, 29, 47). These studies, mainly based on hybridization techniques, have turned out to be very complicated, in part because none of the genetic elements found upstream of the VSG gene is unique to a given expression site (37, 40, 42).

The 5' barren regions, usually found immediately upstream of the VSG gene, contain variable numbers of imperfect 76-bp repeats that are common to all expression sites (10, 22, 40, 46). These 76-bp repeats also occur upstream of some nontelomeric VSG genes and are therefore outside of expression sites (3, 34). Expression site-associated genes, or ESAGs, found upstream of the 5' barren region, are present in 8 to 25 copies per genome, and some are common to different expression sites (1, 2, 4, 13, 23, 30, 37, 40, 41, 42, 48–50, 52). However, it is not clear whether all members of ESAG families are localized to expression sites.

Transcripts homologous to the 76-bp repeats and to the ESAGs have been reported in trypanosome variants (1, 2, 6, 13, 17, 23, 27–29, 37, 40, 41, 47–49, 52). Most of these transcripts seem to be expressed in coordination with VSG transcription. Most are present only in life stages that produce VSG and appear to be synthesized by polymerases that are resistant *in vitro* to elevated levels of α -amanitin; α -amanitin resistance is a property of VSG gene transcription *in vitro* (1, 6, 28, 41, 48). In a few cases, it has been shown that expression site-specific sequences are specifically transcribed when the expression site is active (17, 49, 52). However, it has also been shown that some transcripts homologous to the 76-bp repeats and certain ESAGs are not expressed in coordination with VSG, since they are also found in procyclic forms of trypanosomes or are synthesized by α -amanitin-sensitive RNA polymerases (1, 2).

Our comprehension of expression site transcription is thus ambiguous, because the origin of the transcripts homologous to the expression site is not clear. The start of transcription in the expression site has not yet been clearly defined, though it is probably at least inside or upstream of the 5' barren region (17). However, the recent development of techniques for transforming bloodstream forms of trypanosomes should help in defining the structure of expression site promoter sequences (26).

The role ESAGs play in trypanosomes is not yet understood. It has not been possible to attribute any function to ESAGs, except for *ESAG-4* and its homologs, which have significant homology to the yeast adenylate cyclase (2, 41), and more recently for the trypanosome leucine repeat gene product, which has been proposed to play a role in protein-protein interactions (49). Whatever their roles are, it is also not clear whether ESAGs are specifically required for antigenic variation. In order to better understand this latter point, we felt it was of importance to establish the relationship between ESAG and VSG expression.

In this report, we present the characterization of two members of a new ESAG family and the analysis of their expression in various *Trypanosoma equiperdum* variants.

* Corresponding author.

† Present address: Biologie Moléculaire de la Drosophile, Institut Pasteur, 75724 Paris Cedex 15, France.

Using the polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) techniques to distinguish between the transcripts derived from those two genes, we have established that *ESAG-9c*, contained in the 1.78-C expression site, is transcribed when this expression site is active and that *ESAG-9u* (or another gene with identical characteristics), apparently not contained in an expression site, is transcribed regardless of the expression site being used.

MATERIALS AND METHODS

Trypanosomes. All *T. equiperdum* variants are from the Bordeaux trypanosome antigen type (BoTat) serodeme (11). The *Trypanosoma brucei* IsTat-1.1 (Issaquah trypanosome antigen type) is a gift from K. Stuart, Seattle Biomedical Institute, Seattle, Wash.; *Herpetomonas* spp. are a gift from E. P. Camargo, Sao Paulo, Brazil; *Leptomonas* spp. are from the American Type Culture Collection. *T. equiperdum* and *T. brucei* were grown in rats and purified as previously described (33). Total RNA and genomic DNA were isolated by the guanidinium isothiocyanate method (15).

Subclones and fragments. Subclones 1-78-1, 1-78-2, 1-78-3, 1-78-4, and 1-78-6 are genomic BoTat-1⁷⁸ expression site subclones (42). Various genomic fragments were further subcloned into M13mp18, M13mp19, and Bluescribe (+) and Bluescribe (−) (Vector Cloning System) for the sequence analysis (data not shown). A fragment representing the *ESAG-9ci* region was obtained by subcloning the *HaeIII*-*Sau3A* fragment (bp 409 to 595 [see Fig. 2]) from 1-78-6, into Bluescribe (+). A fragment representing the *ESAG-9ui* region was synthesized by PCR with the O20 and O19 oligonucleotides by using *ESAG-9u* cDNA as a template.

Oligonucleotides. Oligonucleotides were synthesized on an Applied Biosystem DNA synthesizer, purified on NAP-25 columns (Pharmacia), and are as follows: O31, 5' GGTCGA CAGTGCCAGAGAGCCGACAGATGCA 3'; O34, 5' GGA ATTCTCTTACCATGTCTCGCCACAACGTTC 3'; O19, 5' GCGCTCGAAGTCGGCCGGC 3'; O20, 5' GGGTGC GAA GGATACTGGG 3'. The 76-bp repeat probe is the 22-mer (5' ACACACTCAACACTCTCCTA 3') homologous to the conserved region of the 76-bp repeat (22).

Probes. Plasmids, inserts, and fragments were purified on agarose gels by electroelution as previously described (36) and were ³²P-labeled by random priming by the method of Feinberg and Vogelstein (20). The tubulin probe is from *T. brucei* (53). The O31-O34 PCR probe was made by amplification of the O31-O34 fragment with the O31 and O34 oligonucleotides by using *ESAG-9u* cDNA as a template by the method of Schowalter and Sommer (44). Oligonucleotides were treated with kinase as previously described (36).

Libraries. The BoTat-1⁷⁸ cDNA library was cloned into the plasmid pUC9 (55) by the method of Gubler and Hoffman (24). The cDNA was inserted into the *PstI* site of pUC9 by G-C tailing. The BoTat-1⁷⁸ genomic library was made in lambda EMBL-4 as previously described (38). It contains partially digested *Sau3A* fragments, selected for sizes of around 15 kb, inserted into the *BamHI* site of EMBL-4. Both libraries were screened by standard procedures (36) with the 1-78-2 probe. Recombinant DNA was isolated and analyzed from the positive clones by standard procedures.

Southern blot and Northern blot analysis. Genomic and recombinant DNA, digested with various restriction enzymes, was electrophoresed in agarose TBE gels and transferred to GeneScreen (NEN Research Products) by the method of Southern (51). Southern blots were hybridized for

16 h at 42°C in 50% formamide–5× SSPE (1× SSPE is 0.18 M NaCl, 10 mM sodium phosphate, 1 mM EDTA [pH 7.4])–5× Denhardt solution–1% sodium dodecyl sulfate (SDS)–100 µg of carrier RNA per ml and then washed at 65°C for 20 min in 2× SSPE–0.1% SDS and twice for 20 min in 0.2× SSPE–0.1% SDS. Total RNA was electrophoresed in formaldehyde gels and transferred to nitrocellulose as suggested by the manufacturer (Schleicher & Schuell). Northern blots were hybridized for 16 h at 42°C in 40% formamide–10% dextran sulfate–5× SSPE–5× Denhardt solution–0.5% SDS–150 µg of carrier RNA per ml and then washed at 55°C for 20 min in 2× SSPE–0.1% SDS and twice for 20 min in 0.2× SSPE–0.1% SDS.

DNA sequencing. Nucleotide sequences were determined by the dideoxynucleotide chain termination method of Sanger et al. (43). Commercially available universal primers and specific oligonucleotides were used on various single-stranded and double-stranded DNA templates for the regions of interest. The barren region was sequenced from overlapping unidirectionally deleted clones of the *BamHI*-*EcoRI* fragment, obtained by the method of Henikoff (25). The coding regions and most of the noncoding regions were sequenced on both strands. The sequences were analyzed and compared by using Genepro sequence analysis software, version 4.1 (Riverside Scientific Enterprises).

PCR. The DNA template (10 to 100 pg of plasmid and bacteriophage and 200 ng of genomic DNA) was added to 50 mM KCl–10 mM Tris (pH 8.3)–1.5 mM MgCl₂–0.01% (wt/vol) gelatin–0.25 mM (each) dATP, dCTP, dGTP, and TTP–2 µCi [α-³²P]dCTP (800 Ci/mmol; NEN)–10 µM each oligonucleotide–1.25 U of *TaqI* polymerase (Perkin-Elmer-Cetus) in a final volume of 25 µl. Total RNA was treated with RNase-free DNase (Worthington). For each PCR, 0.5 µg of DNase-treated RNA was reverse transcribed with 10 pM random primers (hexamers; Pharmacia) in 5 µl of 50 mM KCl–20 mM Tris (pH 8.4)–3 mM MgCl₂–100 µg of bovine serum albumin per ml–1 mM (each) dATP, dCTP, dGTP, and TTP–100 U of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories) and incubated for 10 min at 22°C and then for 45 min at 42°C. The reaction was stopped by 5 min of incubation at 95°C. For the amplification reaction, 20 µl of 50 mM KCl–20 mM Tris (pH 8.4)–1.13 mM MgCl₂–0.01% (wt/vol) gelatin–2 µCi [α-³²P]dCTP (800 Ci/mmol; NEN)–10 µM each oligonucleotide–1.25 U of *TaqI* polymerase (Perkin-Elmer-Cetus) was added to the 5 µl of reverse-transcribed RNA described above.

The amplification reactions, 30 cycles of 1 min at 94°C and 5 min at 74°C for the O31-O34 pair and 30 cycles of 1 min at 94°C, 2 min at 63°C, and 3 min at 74°C for the O19-O20 pair, were done in a Perkin-Elmer-Cetus thermocycler. The amplified products were run in nondenaturing acrylamide gels (36) directly after the PCR or after redigestion by chosen restriction enzymes. Redigestion was done by diluting the amplified products 1:10 directly after the PCR in the buffer for the enzyme. After electrophoresis, the gels were dried and autoradiographed.

Controls for absence of contaminant genomic DNA in the total RNA were done by omitting the enzyme in the reverse transcription reaction and amplifying the sample as described above. It was verified by PCR that no bands could be amplified from these controls. The controls, as well as controls testing for the absence of other types of contaminants, have been systematically included. Also, the amplified bands are obtained only when the two oligonucleotides are in the reaction and cannot be amplified when only one oligonucleotide is present.

FIG. 1. Restriction maps, clones, and subclones for *ESAG-9c* and *ESAG-9u* genes. (A) Genomic clone (1-78-11) and subclones for the 1.78-C expression site. (B) Restriction map for the 1.78-C expression site showing the *ESAG-9c* gene and the *VSG-178* gene. (C) Restriction map for the *ESAG-9u* cDNA insert. AAA, Poly(A)⁺ tail. (D) Restriction map for the *Lambda.ESAG-9u* genomic clone. Open reading frames are indicated by a thicker line. Abbreviations: B.R., 5' barren region; B, *Bam*HI; E, *Eco*RI; Hi, *Hind*III; Rs, *Rsa*I; S, *Sal*I; Sa, *Sau*3A; Xh, *Xho*I; t, telomere. (A and B) Information is from Raibaud et al. (42), except for the indication of the *ESAG-9c* open reading frame.

Nucleotide sequence accession number. Nucleotide sequence data reported in this paper have been submitted to the GenBank data base and were assigned accession numbers M37627 and M37628.

RESULTS

***ESAG-9* family.** We have previously described the cloning of a portion of the 1.78-C expression site on a 11.7-kb fragment contiguous with the *VSG-178* gene (42) (Fig. 1). More than 4 kb directly upstream of *VSG-178* was sequenced. The region immediately 5' of the *VSG-178* gene consists of a 500-bp cotransposed region already described (22) and a 2-kb 5' barren region containing 23 imperfect 76-bp repeats (data not shown). Further sequencing revealed a 960-bp open reading frame upstream of the 5' barren region, in the same orientation as the *VSG-178* gene (Fig. 2). This open reading frame represents a potential ESAG. It has homology neither to any of the seven known ESAG families (1, 13, 30, 41, 50, 52) nor to the trypanosome leucine repeat gene (49). Previous work indicated that there are six to eight loci homologous to this open reading frame in the *T. equiperdum* genome (42). Five, designated A, B, C, X, and Y, are telomeric. The three former loci were demonstrated to be in functional expression sites 1.78-A (used in BoTat-1⁴), 1.78-B (used in BoTat-201 and BoTat-20), and 1.78-C (used in BoTat-1⁷⁸), respectively. The 1.78-C expression site appears to be a duplication of the 1.78-B expression site (42). We therefore named this new gene family *ESAG-9*, and the gene contained in the 1.78-C expression site was named *ESAG-9c*. Genes homologous to the *ESAG-9* family were found in *T. brucei* and in *Trypanosoma gambiense* but not in *Trypanosoma cruzi*, *Leptomonas* spp., or *Herpetomonas* spp. (Fig. 3A).

Northern blots were hybridized with a probe corresponding to the 3' half of the open reading frame. A 2-kb poly(A)⁺ transcript was detected in all the *T. equiperdum* variants examined, whether they utilize a 1.78 expression site or not (Fig. 3B; unpublished data). A transcript of similar size was

detected with this probe in *T. brucei* bloodstream but not procyclic forms (Fig. 3B).

In order to determine which genes of the *ESAG-9* family were transcribed in BoTat-1⁷⁸, a cDNA library, constructed from BoTat-1⁷⁸ poly(A)⁺ RNA, was screened with an *ESAG-9* probe. A positive clone, containing a 1.7-kb insert that corresponds to the 3' end of the 2-kb transcript, was isolated. Its restriction map is very similar, though not identical, to the *ESAG-9c* map. It hybridized to the 1-78-6, 1-78-2, and 1-78-3 subclones of the 1.78-C expression site but not to the 76-bp repeat or the minixon (data not shown). The nucleotide sequence confirmed that it is a truncated cDNA copy of a transcript sharing 80% similarity with *ESAG-9c* (Fig. 2). It must then correspond to another gene of the *ESAG-9* family that we named *ESAG-9u*. Both genes use the same reading frame, for which we have indicated all the in-frame ATGs, and the common TAA (bp 1082 [*ESAG-9c*] and 890 [*ESAG-9u*] [Fig. 2]). The open reading frame is 960 bp for *ESAG-9c* and at least 888 bp for *ESAG-9u*. The potentially encoded peptides are similar, and comparison of the peptides will be discussed later.

Sequence analysis indicated that *ESAG-9c* and *ESAG-9u* are very similar throughout, except for an internal 280-bp region starting about 190 nucleotides in from the first ATG. Specific probes corresponding to these internal regions were derived for both genes and were called *ESAG-9ci* and *ESAG-9ui*. They were used to hybridize genomic Southern blots to determine which subsets of genes they recognize in the *ESAG-9* family (Fig. 4). *ESAG-9ui* turned out to be specific for a single gene, the *ESAG-9u* gene, that runs as a 4-kb *Eco*RI band (Fig. 4B). *ESAG-9ci* recognized all the genes in the *ESAG-9* family except for *ESAG-9u* (cf. Fig. 4A and Fig. 3, lane 2). Thus, while *ESAG-9ui* is specific for the *ESAG-9u* gene, *ESAG-9ci* is not specific for the *ESAG-9c* gene. The *ESAG-9ui* probe was used to hybridize the genomic Southern blot on Fig. 3A and failed to detect any gene in *T. brucei* and *T. gambiense* genomes (data not shown).

Since *ESAG-9u* was expressed in BoTat-1⁷⁸ and since it is not part of the cloned 1.78-C expression site, we wanted to

FIG. 3. (A) Southern blot analysis of *Eco*RI genomic digests, hybridized with the *ESAG-9u* cDNA probe. Lanes: 1, Molecular weight markers; 2, *T. equiperdum* (BoTat-78); 3, *T. brucei*; 4, *T. gambiense*; 5, *T. cruzi*; 6, *Herpetomonas* sp.; 7, *Leptomonas* sp. All lanes were positive on rehybridization with a tubulin probe (data not shown). (B) Northern blot analysis of total RNA (5 μ g per lane), hybridized with O31-O34 PCR probe (top panel) and rehybridized with a tubulin probe (bottom panel). Total RNA is from BoTat-1⁷⁸ (lane 1), BoTat-1 (lane 2), BoTat-28 (lane 3), BoTat-1⁴ (lane 4), BoTat-201 (lane 5), *T. brucei* bloodstream form (lane 6), and *T. brucei* procyclics (lane 7).

determine its genomic environment. The gene was cloned from a lambda EMBL-4 genomic library of BoTat-1⁷⁸ DNA and designated *Lambda.ESAG-9u* (Fig. 1D). Its restriction map was determined and found to be distinct from the restriction maps previously established around the five telomere-linked *ESAG-9* homologs (42). In fact, the bands containing *ESAG-9u* correspond to the ones that are the most resistant to *Bal*31 digestion in the *ESAG-9* family (42; unpublished data). Previously published hybridization analysis indicated that the genomic environments of the five telomere-linked *ESAG-9* homologs were homologous to all

FIG. 4. (A) Southern blot analysis of *Eco*RI digest of BoTat-78 genomic DNA (lane 2) hybridized with the *ESAG-9ci* probe. Lane 1, Molecular weight markers. (B) Southern blot analysis of several digests of BoTat-1⁷⁸ genomic DNA hybridized with the *ESAG-9ui* probe. Lanes: 1, *Eco*RI; 2, *Hind*III; 3, *Sal*I; 4, *Bam*HI.

the probes derived from the 1.78-C expression site (42). However, *Lambda.ESAG-9u* hybridized to 1.78-6, 1.78-2, 1.78-3, and *ESAG-9ui* but did not hybridize to 1.78-4, the 76-bp repeats, or *ESAG-9ci* probes (data not shown). *ESAG-9u* is therefore one of the non-telomere-linked members of the *ESAG-9* family and does not appear to be linked to any of the genes found in the vicinity of *ESAG-9c*. *ESAG-9u* is either further upstream in one expression site or intrachromosomal.

Analysis of *ESAG-9c* and *ESAG-9u* expression in BoTat-1⁷⁸ and other *T. equiperdum* variants. To determine whether *ESAG-9c* was expressed in BoTat-1⁷⁸, and knowing that *ESAG-9u* was expressed in BoTat-1⁷⁸, we amplified fragments corresponding to *ESAG-9u* and *ESAG-9c* genes from BoTat-1⁷⁸ total RNA. *ESAG-9u* was used as a positive control. To determine whether *ESAG-9c* transcription was specifically linked to expression site activity, we performed the same experiment on BoTat-1. BoTat-1 was chosen because it does not use a 1.78 expression site (35, 42). For amplification we used a pair of oligonucleotides 100% homologous to both genes but which flank regions with different restriction sites. Thus, the products could be distinguished by RFLP. The oligonucleotides O31 and O34 were made, and the restriction sites *Alu*I, *Ava*II, and *Hha*I were chosen for the RFLP studies (Fig. 2). It was predicted that the band amplified with O31-O34 on the *ESAG-9c* gene should be 642 bp long, uncut by *Alu*I, cut into four fragments by *Ava*II (43, 130, 164, and 305 bp), and cut into three fragments by *Hha*I (230, 315, and 97 bp). This will be referred to as the *ESAG-9c* pattern. The O31-O34 band amplified on *ESAG-9u* should be 639 bp long, cut into three fragments by *Alu*I (82, 496, and 61 bp), uncut by *Ava*II, and cut into three fragments by *Hha*I (191, 36, and 412 bp). This is the *ESAG-9u* pattern. It was first verified that the expected patterns could be generated from the clones for the two genes. The *ESAG-9c* pattern was obtained from the cloned 1.78-C expression site as well as from the cloned 1.78-B expression site (data not shown). This was expected since those two expression sites are duplications of each other (42). The *ESAG-9u* pattern was obtained from *ESAG-9u* cDNA and *Lambda.ESAG-9u* (data not shown). Genomic DNAs from BoTat-1⁷⁸ and BoTat-1 were then used as templates. Both types of patterns were obtained (data not shown). Finally, the O31-O34 amplification was conducted on total RNA that was first reverse transcribed with random primer (Fig. 5). Both types of pattern were obtained from BoTat-1⁷⁸ RNA, showing that both genes were expressed in this variant. Only the *ESAG-9u* pattern was seen from BoTat-1 RNA, indicating that *ESAG-9c* was not expressed in this variant, while *ESAG-9u* (or another gene of the family having the *ESAG-9u* pattern) was transcribed in this variant. This result clearly suggests that *ESAG-9c* is expressed in a site-specific manner. To further test this possibility, we examined other *T. equiperdum* variants. We tested BoTat-1⁴ (using the 1.78-A expression site), BoTat-20 (using the 1.78-B expression site), and BoTat-28 (using the 28 expression site) (35, 42). As expected, both patterns were obtained from all genomic DNAs (data not shown). The *ESAG-9c* pattern was observed only in BoTat-20 RNA, while the *ESAG-9u* pattern was obtained from all RNAs. This PCR-RFLP analysis thus indicated that *ESAG-9c* transcription was restricted to active expression site and that *ESAG-9u*-like transcription occurred in all variants independently of the expression site the variants used.

***ESAG-9*-encoded proteins.** *ESAG-9c* and *ESAG-9u* could respectively encode proteins of 320 and at least 296 amino

FIG. 5. PCR-RFLP analysis of the expression of *ESAG-9c* and *ESAG-9u* in BoTat-1 and BoTat-1⁷⁸. Autoradiograph of a 6% acrylamide gel, in which the fragments, amplified with O31 and O34 oligonucleotides, from either BoTat-1 (lanes 1) or BoTat-1⁷⁸ (lanes 78) RNA, have been electrophoresed directly after the amplification (NC) or after redigestion with *AluI*, *AvaII*, or *HhaI* as indicated. Molecular markers are in base pairs.

acids that are 67% similar (Fig. 6). The open reading frames were confirmed from the *HindIII* site downstream, by expression in *Escherichia coli* (41a). We do not know where the *ESAG-9u* putative protein starts, since we have no data for its N terminus. The two putative proteins are highly related, with a divergent region of 95 amino acids between the homologous domains. Despite this difference, both putative proteins have a similar amino acid content, hydropathy profile (Fig. 6), and predicted structure (45). The proteins are very rich in glycines (12.8 and 13.2% for *ESAG-9c* and *ESAG-9u*, respectively) and serines (11.6 and 11.1% for *ESAG-9c* and *ESAG-9u*, respectively). The proteins are primarily hydrophilic with predicted globular structure (β -turns). The *ESAG-9c* putative protein has an N-terminal

signal peptide, with a cleavage site predicted between amino acids 20 and 27 (56), and a potential glycosylation site (at amino acid 224). Both *ESAG-9c* and *ESAG-9u* putative proteins have a C-terminal hydrophobic domain, which suggests that they might be glycosyl-phosphatidylinositol anchored (8, 12, 21). However, this domain, which is only 13 amino acids long, might be too small to serve as a signal for proper anchoring (5). Data base searches failed to detect any significant homology between the two putative proteins and known proteins. However, the abundance of glycine residues led us to look for the presence of sequences matching the consensus proposed for GTP-binding sites (18). Good matches to such sequences were found in the C-terminal domain common to both putative proteins, tentatively suggesting that *ESAG-9* family members could encode some GTP-binding proteins. However, the function of these proteins is not yet known.

DISCUSSION

We have found a new *ESAG* family that has six to eight members in *T. equiperdum* and *T. brucei*, five of which are telomere linked in *T. equiperdum*. We have cloned two members of this *ESAG-9* family from a *T. equiperdum* variant; one is contained in a telomeric expression site (*ESAG-9c*) and the other is not telomere linked (*ESAG-9u*). The two genes are highly related but diverge in a 280-bp internal region.

Hybridization studies using the 280-bp internal region from either gene indicated that the four other telomere-linked members of the *ESAG-9* family are similar to *ESAG-9c*. However, the 280-bp region of *ESAG-9u* was specific for *ESAG-9u*, and our Southern blot data, combined with data comparing band intensity between members of the *ESAG-9* family (42), suggested that *ESAG-9u* was a single-copy gene. While the five telomere-linked members of *ESAG-9* family are encoded in similar genomic environments (42), *ESAG-9u* was found to be encoded in an unrelated environment. These observations suggested that *ESAG-9u* is probably a single-copy gene which is either far upstream in an expression site, beyond *Bal31* sensitivity, or intrachromosomal. If the former were true, *ESAG-9u* would be linked to an expression site and its transcription would be expected to be dependent on the activity of this expression site. If the latter were true, *ESAG-9u* could be encoded outside of expression sites and its transcription could be independent of which expression site is active.

The Northern blot analysis of *ESAG-9* family expression indicated that at least one gene of the family is expressed in all *T. equiperdum* variants. The PCR and RFLP analysis detected *ESAG-9c*-like transcripts only in variants having the *ESAG-9c* gene in their active expression site and *ESAG-9u*-like transcripts in all the variants. This indicated that *ESAG-9c* transcription was under expression site regulation and that, in addition, another gene having *ESAG-9u* characteristics was always expressed in any variant. The simplest explanation for this latter observation is that *ESAG-9u* is encoded outside of expression sites and expressed in all variants, regardless of the expression site the variants use. Alternatively, the *ESAG-9u*-like transcripts could be encoded by other genes in the *ESAG-9* family that are not yet characterized. However, it is interesting to note that *ESAG-9u* is the only member of the family containing the *ESAG-9ui* region. We do not know whether this region confers a unique function to the gene. However, one could speculate that *T. equiperdum* must express this unique

brane associated, but their functions have not yet been uncovered (1, 13, 14). Both predicted proteins contain amino acid motifs which suggest that they could bind GTP. They are currently being expressed in *E. coli* in order to be analyzed biochemically. The possibility that *ESAG-9c* could encode a GTP-binding protein is extremely interesting, especially since this gene is found, in our expression site, downstream of an *ESAG-4* homolog that encodes a protein having adenylate cyclase activity (42a). Our trypanosome leucine repeat gene homolog is found in between these two. Whether *ESAG-9u* could also encode a GTP-binding protein that could have a different role because of its unique domain (encoded by *ESAG-9ui*) is also under investigation.

Our analysis indicates that African trypanosomes contain expression site-linked ESAGs that are regulated like VSG genes. They probably also contain non-expression site-linked ESAGs that are transcribed constitutively in blood-stream forms in a non-expression site-regulated manner. This suggests that ESAG function might not be exclusively related to antigenic variation.

ACKNOWLEDGMENTS

We thank Andrew Strand for critical reading of the manuscript and Elisabeth Khouri for technical assistance.

This research was supported by grants from the Centre National de la Recherche Scientifique, European Community, and by Public Service Health grant AI 23010 from the NIH.

REFERENCES

- 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.
- Alexandre, S., P. Paindavoine, P. Tebadi, A. Pays, S. Halleux, M. Steinert, and E. Pays. 1990. Differential expression of a family of putative adenylate/guanylate cyclase genes in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 43:279-288.
- Aline, R., Jr., G. MacDonald, E. Brown, J. Allison, P. Myler, V. Rothwell, and K. Stuart. 1985. (TAA)_n within sequences flanking several intrachromosomal variant surface glycoprotein genes in *Trypanosoma brucei*. *Nucleic Acids Res.* 13:3161-3177.
- Barnes, D. A., J. C. Mottram, and N. Agabian. 1990. Blood-stream and metacyclic variant surface glycoprotein gene expression sites of *Trypanosoma brucei gambiense*. *Mol. Biochem. Parasitol.* 41:101-114.
- Berger, J., A. D. Howard, L. Brink, L. Gerber, J. Haubert, B. R. Cullen, and S. Udenfriend. 1988. COOH-terminal requirements for the correct processing of a phosphatidylinositol-glycan anchored membrane protein. *J. Biol. Chem.* 263:10016-10021.
- Bernards, A., J. M. Kooter, and P. Borst. 1985. Structure and transcription of a telomeric surface antigen gene of *Trypanosoma brucei*. *Mol. Cell. Biol.* 5:545-553.
- Boothroyd, J. C. 1985. Antigenic variation in African trypanosomes. *Annu. Rev. Microbiol.* 39:475-502.
- Boothroyd, J. C., G. A. M. Cross, J. H. J. Hoeijmakers, and P. Borst. 1980. A variant surface glycoprotein of *Trypanosoma brucei* synthesized with a C-terminal hydrophobic "tail" absent from purified glycoprotein. *Nature (London)* 288:624-626.
- Borst, P. 1986. Discontinuous transcription and antigenic variation in trypanosomes. *Annu. Rev. Biochem.* 55:701-732.
- Campbell, D. A., M. P. Van Bree, and J. C. Boothroyd. 1984. The 5'-limits of transposition and upstream barren region of a trypanosome VSG gene: tandem 76 base-pair repeats flanking (TAA)₉₀. *Nucleic Acids Res.* 12:2759-2774.
- Capbern, A., C. Giroud, T. Baltz, and P. Mattern. 1977. *Trypanosoma equiperdum*: etude des variations antigeniques au cours de la trypanosome experimentale du lapin. *Exp. Parasitol.* 42:6-13.
- Caras, I., and G. N. Weddell. 1989. Signal peptide for protein secretion directing glycopospholipid membrane anchor attachment. *Science* 243:1196-1198.
- Cully, D. F., S. I. Hon, and G. A. M. Cross. 1985. Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. *Cell* 42:173-182.
- Cully, D. F., C. P. Gibbs, and G. A. M. Cross. 1986. Identification of proteins encoded by variant surface glycoprotein expression site-associated genes in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 21:189-197.
- Davis, L. G., M. D. Dibner, and J. F. Battey. 1986. Basic methods in molecular biology. Elsevier Science Publishing, Inc., New York.
- DeLange, T., and P. Borst. 1982. Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. *Nature (London)* 299:451-453.
- DeLange, T., J. M. Kooter, J. Luirink, and P. Borst. 1985. Transcription of a transposed trypanosome surface antigen gene starts upstream of the transposed segment. *EMBO J.* 4:3299-3306.
- Dever, T. E., M. J. Glynias, and W. C. Merrick. 1987. GTP-binding domain: three consensus sequence elements with distinct spacing. *Proc. Natl. Acad. Sci. USA* 84:1814-1818.
- Donelson, J. E., and A. C. Rice-Ficht. 1985. Molecular biology of trypanosome antigenic variation. *Microbiol. Rev.* 49:107-125.
- Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6-13.
- Ferguson, M. A. J., and A. F. Williams. 1988. Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. *Annu. Rev. Biochem.* 57:285-320.
- Florent, I., T. Baltz, A. Raibaud, and H. Eisen. 1987. On the role of repeated regions 5' to VSG genes in African trypanosomes. *Gene* 53:55-62.
- Gibbs, C. P., and G. A. M. Cross. 1988. Cloning and transcriptional analysis of a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 28:197-206.
- Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28:351-359.
- Jefferies, D., P. Tebadi, and E. Pays. 1991. Transient activity of the *Trypanosoma brucei* variant surface glycoprotein gene promoter: control of the gene expression at the posttranscriptional level. *Mol. Cell. Biol.* 11:338-343.
- 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 including a VSG gene. *Cell* 51:273-281.
- Kooter, J. M., and P. Borst. 1984. Alpha-amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Res.* 12:9457-9472.
- Kooter, J. M., H. J. van der Spek, R. Wagter, C. E. de Oliveira, F. Van der Hooven, P. J. Johnson, and P. Borst. 1987. The anatomy and transcription of a telomeric expression site for variant-specific surface antigen in *Trypanosoma brucei*. *Cell* 51:261-272.
- Kooter, J. M., A. J. Winter, C. de Oliveira, R. Wagter, and P. Borst. 1988. Boundaries of telomere conversion in *Trypanosoma brucei*. *Gene* 69:1-11.
- Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157:105-132.
- Laurent, M., E. Pays, E. Magnus, N. Van Meirvenne, G. Matthyssens, R. O. Williams, and M. Steinert. 1983. DNA rearrangements linked to expression of a predominant surface antigen gene of trypanosomes. *Nature (London)* 302:263-266.
- Layden, R. E., and H. Eisen. 1988. Alternate *trans* splicing in *Trypanosoma equiperdum*: implications for splice site selection.

- Mol. Cell. Biol. 8:1352–1360.
34. Liu, A. C. Y., L. H. T. Van der Ploeg, F. A. M. Rijsewijk, and P. Borst. 1983. The transposition unit of variant surface glycoprotein gene 118 of *Trypanosoma brucei*: presence of repeated elements at its border and absence of promoter-associated sequences. J. Mol. Biol. 167:57–75.
 35. Longacre, S., U. Hibner, A. Raibaud, H. Eisen, T. Baltz, C. Giroud, and D. Baltz. 1983. DNA rearrangements and antigenic variation in *Trypanosoma equiperdum*: multiple expression-linked sites in independent isolates of trypanosomes expressing the same antigen. Mol. Cell. Biol. 3:399–409.
 36. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 37. Murphy, N. B., M. Guyaux, E. Pays, and M. Steinert. 1987. Analysis of VSG expression site sequences in *Trypanosoma brucei*, p. 449–469. In N. Agabian, H. Goodman, and N. Nogueira (ed.), Molecular strategies of parasitic invasion. Alan R. Liss, Inc., New York.
 38. Murray, N. E. 1983. Phage lambda and molecular cloning, p. 395–432. In R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 39. Pays, E., M. Guyaux, D. Aerts, N. Van Meirvenne, and M. Steinert. 1985. Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. Nature (London) 316:562–564.
 40. Pays, E., and M. Steinert. 1988. Control of antigen gene expression in African trypanosomes. Annu. Rev. Genet. 22:107–126.
 41. Pays, E., P. Tebadi, A. Pays, H. Coquelet, P. Revelard, D. Salmon, and M. Steinert. 1989. The genes and transcripts of an antigen gene expression site from *T. brucei*. Cell 57:835–845.
 - 41a. Porter, S. Personal communication.
 - 41b. Raibaud, A. Personal communication.
 42. Raibaud, A., G. Buck, T. Baltz, and H. Eisen. 1986. Cloning and characterization of a variant surface glycoprotein expression site from *Trypanosoma equiperdum*. Mol. Cell. Biol. 6:2950–2956.
 - 42a. Ross, D. T., A. Raibaud, I. C. Florent, S. Sather, M. K. Gross, D. R. Storm, and H. Eisen. Submitted for publication.
 43. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
 44. Schowalter, D. B., and S. S. Sommer. 1989. The generation of radiolabelled DNA and RNA probes with polymerase chain reaction. Anal. Biochem. 177:90–94.
 45. Schulz, G. E., and R. H. Schirmer. 1979. Principles of protein structure. Springer-Verlag KG, New York.
 46. Shah, J. S., J. R. Young, B. E. Kimmel, K. P. Iams, and R. O. Williams. 1987. The 5' flanking sequence of a *Trypanosoma brucei* variable surface glycoprotein gene. Mol. Biochem. Parasitol. 24:163–174.
 47. Shea, C., M. G.-S. Lee, and L. H. T. Van der Ploeg. 1987. VSG gene 118 is transcribed from a Poll-like promoter. Cell 50:603–612.
 48. Shea, C., and L. H. T. Van der Ploeg. 1988. Stable variant-specific transcripts of the variant cell surface glycoprotein gene 1.8 expression site in *Trypanosoma brucei*. Mol. Cell. Biol. 8:854–859.
 49. Smiley, B. L., A. W. Stadnyk, P. J. Myler, and K. Stuart. 1990. The trypanosome leucine repeat gene in the variant surface glycoprotein expression site encodes a putative metal-binding domain and a region resembling protein-binding domains of yeast, *Drosophila*, and mammalian proteins. Mol. Cell. Biol. 10:6436–6444.
 50. Son, H. J., A. C. Cook, T. Hall, and J. E. Donelson. 1989. Expression site associated genes of *Trypanosoma brucei rhodesiense*. Mol. Biochem. Parasitol. 33:59–66.
 51. Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–518.
 - 51a. Stadnyk, A. Personal communication.
 52. Stadnyk, A. W., J. K. Scholler, P. J. Myler, and K. D. Stuart. 1990. Ribonuclease protection determines sequences specific to a single variable surface glycoprotein gene expression site, p. 99–109. In N. Agabian and A. Cerami (ed.), Parasites: molecular biology, drug and vaccine design. Alan R. Liss, Inc., New York.
 53. Thomashow, L. S., M. Milhausen, W. J. Rutter, and N. Agabian. 1983. Tubulin genes are tandemly linked and clustered in the genome of *Trypanosoma brucei*. Cell 32:35–43.
 54. Vickerman, K. 1985. Developmental cycles and biology of pathogenic trypanosomes. Br. Med. Bull. 41:105–114.
 55. Vieira, J., and J. Messing. 1982. The pUC plasmids, and M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.
 56. Von Heijne, G. 1983. Patterns of amino acids near signal sequences cleavage sites. Eur. J. Biochem. 133:17–21.