Expression of a Retroposon-Like Sequence Upstream of the Putative *Trypanosoma brucei* Variant Surface Glycoprotein Gene Expression Site Promoter

MICHAEL J. LODES,¹ BOB L. SMILEY,¹ ANDREW W. STADNYK,^{1†} J. LINDSLEY BENNETT,¹
PETER J. MYLER,^{1,2} AND KENNETH STUART^{1,2*}

*Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, Washington 98109-1651,¹ and
Pathobiology Department, University of Washington, Seattle, Washington 98195²*

Received 24 May 1993/Returned for modification 13 July 1993/Accepted 23 August 1993

We have cloned the region spanning the putative promoter from two variant surface glycoprotein gene expression sites that are at each end of chromosome M4 of *Trypanosoma brucei* IsTat 7. Both expression sites contain a retroposon-like sequence (ESR) pseudogene whose 3' end is approximately 30 bp upstream of the putative expression site promoter. The ESRs from both expression sites share considerable sequence homology and are related to LINE-like elements, especially the *T. brucei* ingi retroposon. Other ESRs are located on large, but not intermediate or mini-, chromosomes in the IsTaR 1 serodeme, and the total copy number is 10 to 20, similar to that estimated for variant surface glycoprotein expression sites. No DNA rearrangements in the vicinity of the ESR and putative expression site promoter were detected following antigenic switches in the IsTaR 1 serodeme. ESR transcripts are present in bloodstream, but not procyclic, forms. Variation in transcript size and sequence between bloodstream variant antigenic types implies that only the ESR from the active expression site is transcribed. This pattern of expression reflects that of sequences downstream of the putative expression site promoter, suggesting that the region of coordinately controlled expression extends upstream of this promoter.

The *Trypanosoma brucei* subspecies cause substantial human suffering and death as well as severe disease in domestic livestock. However, the development of an anti-trypanosome vaccine has been hindered by the parasite's ability to evade the host's immune response. Trypanosomes have evolved an intricate system which allows them to switch their variant-specific surface glycoproteins (VSGs), preventing elimination of the parasite population by host defenses (for reviews, see references 2, 4, 7, and 38). The expressed trypanosome VSG gene is part of a 40- to 60-kb polycistronic transcription unit which is composed of the VSG gene and at least eight expression site-associated genes (ESAGs) (5, 6, 25, 34). VSG gene expression is life cycle stage specific and usually confined to one of several telomeric expression sites in each variant antigenic type (VAT) (4). Inactivation of VSG and ESAG expression in procyclic forms may be due to premature termination of transcription or decreased transcript stability (14, 25). In the bloodstream forms, VSG expression may be regulated at the level of transcript initiation, because only one expression site is usually transcriptionally active (43).

While gene conversion of the VSG gene within an active expression site can bring about an antigenic switch, the mechanism(s) which underlies activation of one expression site and concomitant inactivation of another remains to be elucidated. However, DNA rearrangements in the telomeric region downstream of the VSG (20, 24) and upstream of the putative expression site promoter (10, 41) are often observed following antigenic switches involving such telomeric activation. The latter rearrangements involve recombination in

simple sequences (50-bp repeat) and deletion of an upstream copy of the putative expression site promoter (11).

Because sequences upstream of the putative expression site promoter may be important in the regulation of antigenic switching or undergo DNA rearrangement in bloodstream form trypanosomes, we have cloned and sequenced this region from two expression sites at either end of the 1.4-Mb chromosome M4 from IsTat 7. In this study, we describe a family of expression site-associated retroposon-like sequences (ESRs) which occur approximately 30 bp upstream of the putative VSG expression site promoter. The ESRs are related to LINE elements and especially to the LINE-like *T. brucei* ingi retroposon (15, 19, 33). Members of this ESR family occur in the expression sites of megabase-sized, but not intermediate or mini-, chromosomes. ESR transcripts are detected only when the associated VSG gene is expressed and ESR transcript size is expression site specific. Interestingly, the location of the ESR is within a region of the expression site reported to rearrange during an antigenic switch (11). However, no rearrangements were evident in this area following antigenic switches in the IsTaR 1 serodeme.

MATERIALS AND METHODS

Organisms and nucleic acid. The IsTaR 1 serodeme was developed from *T. brucei* stock 164 (36). VATs used in this study have been described previously (21, 34) (see Fig. 6 for details). Genomic DNA was obtained from bloodstream forms grown in mice or rats and harvested by the methods of Milhausen et al. (17). VAT identity was confirmed by indirect immunofluorescence with VAT-specific monoclonal antibody and fluorescein isothiocyanate-labeled second antibody. Total RNA from procyclic or bloodstream stage trypanosomes was prepared by the guanidinium isothiocy-

* Corresponding author.

† Present address: Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, Canada B3J 3G9.

anate method (3). Poly(A)⁺-selected RNA was purified with Poly(A)Quik columns (Stratagene), using the protocol and buffers supplied by the manufacturer. Chromosome-sized DNA was prepared for pulsed-field gradient electrophoresis (PFGE) by the methods of Scholler et al. (31) except that blocks were prepared at a concentration of 10⁹ cells per ml.

Construction of clones. Genomic lambda clones DTb1.7g-e9 and DTb1.7g-e1 (referred to in this study as e9 and e1, respectively) were isolated as described previously (34, 35). Subclones of e9 and e1, pTb1.7g-e9.1.4, pTb1.7g-e9.1.2, and pTb1.7g-e1.5 (see Fig. 1 for relevant restriction sites) were prepared in pBluescript SK⁻ or pBluescript II SK⁻ (Stratagene). Deletion subclones, used for sequencing, were produced by exonuclease III digestion (Erase-a-base; Promega) following the supplier's instructions. An additional 500-bp clone (pTb1.7g-RT5) was isolated by polymerase chain reaction (PCR) amplification of chromosome M4-enriched DNA (Perkin-Elmer-Cetus, Norwalk, Conn.) with a T5 expression site-specific primer, EX-1-1 (CTTGGAT CCTACGGCGTCTCTGATTAATTC), derived from clone e1 sequence, and a 50-bp repeat primer, EX-50BP (TATGG ATCCTGTACTAAAAGTATTCTCAC) (43), which apparently primed fortuitously. cDNA clones corresponding to the 5' 1.2-kb portion of the ESR transcript from several VATs (5^{A3}, 7, A^{7L}, A¹¹, and 11^{A11}) were prepared by PCR amplification. First-strand cDNA, prepared by XSC-dT₁₇ (GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT) priming of total RNA with Moloney murine leukemia virus H⁻ reverse transcriptase (Bethesda Research Laboratories [BRL]), was amplified with primers specific for the T3 ESR (EX-9-1 [TATGGATCCTGCTACGGTGTCTAATCGC]) (see Fig. 2B) and the trypanosome spliced leader sequence (X-SL24 [TATCTAGAACAGTTTCTGTACTATATTG]). *Taq* polymerase (Boehringer-Mannheim) was added to the reaction mixtures at 94°C, and 24 PCR cycles (one cycle consists of denaturing at 94°C for 1 min, annealing at 48°C for 1 min, and extension at 72°C for 1.5 min) were carried out followed by 10-min extension at 72°C. The resulting PCR products were digested with *Bam*HI and *Xba*I and cloned in *Bam*HI-*Xba*I-digested pBluescript II SK⁻. The VAT 7 telomere 3 cDNA clone is referred to as pTb1.7c-RT3 in this study. Genomic clones that correspond to the 5' region of ESR transcription were produced by PCR amplification from isolated M4 chromosomal DNA, using the specific T3 ESR primer EX-9-1 in combination with primer EX-9-3 (AT TGTCTAGACAGCAGAAGAGTGCCGC) (see Fig. 2B) or EX-9-4 (TCCAGTCTAGATGAAGGATACCGATCAGG). Primer EX-9-3 was developed from the 5' end of VAT 7 cDNA clone pTb1.7c-RT3 and primer EX-9-4 was developed from the 118 expression site in VAT 118a (9a). These primer combinations were designed to produce products of 1.2 and 1.5 kb, respectively. The resulting 1.5-kb genomic clone is referred to as pTb1.7g-RT3. PCR conditions were similar to those listed above but with an annealing temperature of 55°C.

Electrophoresis, transfer, and hybridization of DNA and RNA. Chromosome-sized DNA was separated by PFGE as described by Scholler et al. (31). Separations were performed at 300 V with 90-s pulse time in 0.4% SeaPlaque GTG agarose (FMC BioProducts) at 15°C for 22 h. Chromosome-specific DNA was obtained for restriction enzyme digestion by excising ethidium bromide-stained bands. Agarose plugs containing chromosomal bands M1, M2, M3, and M4 were washed twice with 10 mM Tris, pH 8.0, for 1 h each. Plugs were then equilibrated in restriction enzyme buffer, heated to 65°C for 10 min to melt the agarose, and cooled to 37°C,

and 30 U of restriction enzyme was added. After overnight digestion, DNA was recovered from melted plugs by adding 0.1 volume of buffer (10 mM Tris [pH 6.0], 100 mM EDTA) and then 4 U of β -agarase (New England Biolabs). Agarose was digested for 4 h at 37°C, extracted with equal volumes of phenol, phenol-chloroform-isoamyl alcohol (25:24:1), and chloroform-isoamyl alcohol (24:1), and DNA was precipitated with 0.1 volume of 2 M NaCl and 2.0 volumes of ethanol and washed with 70% ethanol. To ensure complete digestion, the dried DNA pellet was resuspended in 20 μ l of restriction buffer, 10 U of enzyme was added, and the mixtures were incubated for an additional 2 h at 37°C. Digested DNA was separated in 0.6 or 0.8% SeaKem ME agarose with Tris-borate buffer and transferred to Nytran membranes (Schleicher and Schuell) as previously described (29). After RNase-free DNase I treatment, total RNA (10 μ g per lane) was separated in 1.0% SeaKem ME agarose gels containing 2.2 M formaldehyde and 20 mM MOPS (morpholinepropanesulfonic acid), pH 7.0, and transferred as previously described (8). DNA and RNA were cross-linked to Nytran filters by UV irradiation by following the manufacturer's instructions (UV Stratalinker; Stratagene). Filters were prehybridized and hybridized in a mixture of 5 $\times$ SSPE (1 $\times$ SSPE is 0.18 M NaCl, 10 mM NaPO₄, and 1 mM EDTA [pH 7.7]), 1% Sarkosyl, 1 $\times$ Denhardt's solution, 200 μ g of herring sperm DNA per ml, and 90 μ g of heparin per ml for Southern blots with random-primed probe, with the same mixture containing 30% formamide and 100 μ g of tRNA per ml for Southern blots with riboprobe, or with the same mixture containing 50% formamide for Northern (RNA) blots with random-primed probe. Filters were prehybridized for 4 h and hybridized for 18 h at 65°C for Southern blots with random-primed probe, 50°C for Southern blots with riboprobes, and 42°C for Northern blots with random-primed probe. Hybridized filters were washed twice for 20 min at 65°C with 1 $\times$ SSPE and twice for 20 min with 0.1 $\times$ SSPE (Southern blots) or 0.2 $\times$ SSPE (Northern blots).

Riboprobe for the T3 expression site was prepared from a deletion clone of pTb1.7g-e9.1.4 (see Fig. 2A). After *Eco*RI digestion and gel isolation of plasmid DNA, [³²P]UTP-labeled probe was synthesized with T7 RNA polymerase by the manufacturer's (Stratagene) instructions. Riboprobe for the T5 expression site was synthesized from *Eco*RI-digested and gel-isolated pTb1.7gRT5 (see Fig. 2A) with T3 RNA polymerase. T3 and T5 expression site-specific probes were also obtained by gel isolation of *Eco*RI-*Pst*I fragments from deletion clones of pTb1.7g-e9.1.4 and pTb1.7g-e1.5, respectively. *T. brucei*-specific α -tubulin probe was prepared from clone pTb α T-cl (37). DNA fragments were labeled with [α -³²P]dATP by mixed hexamer priming (Pharmacia).

DNA sequencing and analysis. Double-stranded DNA obtained from plasmid subclones, PCR-derived clones, and from exonuclease III deletion clones was sequenced with either a Sequenase kit (U.S. Biochemical Corp.), using the dideoxy-chain termination method, or with a *Taq* Dye Primer Cycle Sequencing Kit (Applied Biosystems, Inc.) and a model 373A DNA sequencer (Applied Biosystems, Inc.). Computer analysis of nucleotide sequences was performed with DNASTAR (DNASTAR Inc., Madison, Wis.), ESEE (2a), RDF2 (26), and GCG (Genetics Computer Group, Inc.; version 7) software. Searches of the SWISS-PROT (release 17.0) and GenBank (release 67.0) data bases were performed with FASTA (26).

Nucleotide sequence accession numbers. The GenBank accession numbers for the T3 (pTb1.7g-e9.1.2 and pTb1.7g-

FIG. 1. Physical map of the T3 and T5 expression sites on the 1.4-Mb chromosome M4 in the IsTaR 1 serodeme. The locations of genomic clones DTb1.7g-e9, pTb1.7c-RT3, and pTb1.7g-RT3 (T3 expression site) and DTb1.7g-e1 and pTb1.7g-RT5 (T5 expression site) (see Materials and Methods) are indicated with dashed lines showing the locations of subclones and with arrowheads indicating the direction of exonuclease digestion and sequencing. Open boxes indicate the VSG genes and the ESAGs (ESAGs 1 to 7 and TLR [= ESAG 8]). Solid boxes indicate the T3 and T5 ESRs, and flags denote the putative expression site promoter and/or transcription start sites. Restriction enzyme abbreviations: E, *EcoRI*; S, *Sall*; Ss, *SstI*.

e9.1.4) and T5 (pTb1.7g-e1.5 and pTb1.7gRT5) ESRs are L07804 and L07805, respectively.

RESULTS

Sequence analysis. We have previously described the isolation and partial characterization of expression site clones e9 and e1 (34, 35). e9 and e1 clones are derived from the T3 and T5 expression sites of chromosomal DNA M4, respectively, in VAT 7 (Fig. 1). Comparison of ESAG 6 and 7 sequences from each clone with the sequences from clones obtained by PCR amplification of chromosome M4 DNA and cDNA from VAT 7 (which expresses the T3 expression site) and VAT 5 (which expresses the T5 expression site) confirmed the origin of both genomic clones (unpublished data). Sequence from clones e9 (T3 expression site) and e1 (T5 expression site) extended to an *EcoRI* site 3,984 and 3,982 nucleotides (nt), respectively, 5' to the ATG codon of ESAG 7 (Fig. 1 and 2B). Sequence further 5' to this *EcoRI* site was obtained from clones pTb1.7g-RT3 (1.5 kb) and pTb1.7g-RT5 (0.5 kb) which were generated by PCR amplification of M4 DNA from VAT 7 as described in Materials and Methods. Sequence comparison with the original genomic clones and the cDNA clone pTb1.7c-RT3 (1.2 kb) obtained by PCR amplification from VAT 7 mRNA (see Materials and Methods) confirmed the identities of these clones and extended the sequence of the T3 and T5 expression sites further 5' by 1,400 and 390 nt, respectively (Fig. 1 and 2B).

The T3 and T5 expression site sequences 5' to ESAG 7 have 93% homology overall, while some regions (e.g., 2,914 to 3,279 nt upstream of ESAG 7) show even greater (99%) homology. The T3 and T5 sequences also show over 90% homology to the corresponding regions from the 118 (VATs 118a and 118a') and 221 (VAT 221a) expression sites (10, 11, 43) (Fig. 2A) and AnTat AL and AnTat ES expression sites (23). The region 1.5 kb upstream of ESAG 7 (flags in Fig. 2) has been shown to have promoter function in transient transfection assays (14, 42) and has been proposed as the initiation site for expression site transcription (23, 43). This region is highly conserved (>94%) among all sequences. The T3 and T5 sequences diverge from the 221 and 118a' expression site sequences approximately 2.5 and 3.0 kb, respectively, upstream of ESAG 7, just 3' to the 50-bp repeat sequences in the latter expression sites.

Analysis of the T3 and T5 sequences upstream of the putative expression site promoter showed homology to retroposon sequences at both the nucleotide and predicted amino acid levels. Nucleotide sequence homology of 53% over the 5' 1.4 kb of T5 and 52% over the 5' 2.0 kb of T3 was found with the *T. brucei* ingi retroposon (15, 19, 33). This homology lies within the 5' portion (nt 350 to 2400) of ingi

and does not include the 3' portion (Fig. 2A). A search of the SWISS-PROT data base with the amino acid sequences predicted from this region of the T3 and T5 expression sites revealed significant homology (16 to 20 standard deviations above the mean) with reverse transcriptase (RT) homologs of ingi and other retroposons (LINE elements [12]). However, the absence of an open reading frame from either the T3 or T5 ESR (Fig. 3A) indicates that neither now has protein-coding potential (i.e., they are pseudogenes). Alignment of the RT domain from non-LTR (long terminal repeat) retroposons and LTR transposons with the corresponding region from the T3 and T5 ESRs demonstrates the conservation of five motifs (Fig. 3B, underlined) conserved in all RNA-dependent DNA polymerases (27). The five RT motifs, including four strictly conserved amino acids, are present in the ESRs as part of a 245-amino-acid domain which corresponds to the 3' region of sequence homology with ingi (Fig. 3A). Comparison of the ESRs with ingi-3 and human LINE-1 RT sequences alone reveals the conservation of 20 amino acids (Fig. 3B, asterisks) within this region. The T3 and T5 ESRs are more similar to ingi-3 and human LINE-1 than to the trypanosomatid "siteposons" SLACS, CZAR, and CRE1 (1, 9, 40) and even less similar to the retrovirus-like LTR elements (Gypsy and copia) and human immunodeficiency virus type 1.

Genomic organization of ESRs. The ESRs are present in multiple copies within the trypanosome genome. Southern blot analysis of PFGE-separated chromosome-sized DNA (Fig. 4A) shows that a T3 ESR-derived riboprobe hybridizes to M1 (compression zone), M2, M3, and M4 and large chromosomal DNA (Ch) remaining in the well (>3 Mb) (Fig. 4B), while a T5 ESR riboprobe hybridizes to bands M1 and M4 and Ch chromosomal DNA, but not to M2 or M3 (Fig. 4C). No hybridization to intermediate (I) or mini- (mi) chromosomal DNA was observed with either ESR probe. The genomic environments of ESRs within the megabase-sized chromosomes (M1, M2, M3, and M4) were determined by Southern blot analysis of chromosomal DNA digested with restriction enzymes after isolation by PFGE (Fig. 5). T3 and T5 ESR probes hybridize to single restriction fragments in each digest of M4 chromosomal DNA. For example, in the *SphI* digest of M4 DNA, the T3 ESR probe hybridizes with a 5.2-kb band (Fig. 5C, open arrow), while the T5 ESR probe hybridizes with a 4.5-kb band (Fig. 5C, solid arrow), indicating that the cloned sequences represent the only two sequences closely related to ESR on chromosome M4. Six of the seven restriction digests assayed resulted in fragment polymorphisms between the T3 and T5 ESRs. Thus, although the T3 and T5 ESRs are located in the same area of their respective expression sites, they have substantially different restriction maps (Fig. 5B). Similar analyses with

FIG. 3. ESR is a pseudogene with homology to RTs. (A) Diagrammatic representation of open reading frame analysis (FRAMES; GCG) of the T3 expression site (ES) sequence shown in Fig. 2. Tic marks indicate stop codons, and thick lines indicate the amino acid sequence expanded for comparison in panel B. (B) Partial amino acid sequences from the T3 and T5 ESRs (this paper); the trypanosomatid retroposon ingi-3 (15); siteposons CZAR (40), SLACS (1), and CRE1 (9); human LINE-1 (12); transposons Gypsy (16) and copia (18); and human immunodeficiency virus type 1 (GenBank accession no. P12499). Amino acids identical to those of T5 ESR are shaded, bold and uppercase, while conservative replacements are shown shaded and lowercase. An asterisk (above) indicates identity of the T5 ESR with the human LINE-1 element (boxed), and a bar (below) indicates amino acid identity conserved in all sequences. Arrows indicate frameshifts made in T5 alone (↑) and both T3 and T5 (↓) sequences to improve alignments. Periods indicate stops in the open reading frame. The five polymerase motifs conserved in all RTs (27) are underlined and labeled A to E.

isolated M3 chromosomal DNA revealed a restriction pattern identical to that of M4 chromosomal DNA when a T3 ESR probe is used (Fig. 5C). Thus, the M3 chromosome contains at least one ESR within a genomic environment very similar to that of the T3 ESR. Since the T2 expression site on M3 has been previously shown to be closely related to the T3 expression site on M4 (21), it is likely that the T3-related M3 ESR is located in the T2 expression site. As expected from the results of the PFGE blots (Fig. 4), no hybridization of M3 DNA digests was observed with the T5 ESR probe, indicating that M3 contains no T5-related ESR.

Southern analysis of restriction enzyme-digested M2 chromosomal DNA showed two T3-related ESR sequences with restriction maps which differ from each other and the T3 and T5 expression site. Five of seven M2 digests resulted in different hybridization patterns compared with that of T3 ESR, while digestion of M2 with *HincII* and *ClaI* produced bands of hybridization of the same size as that of T3 ESR. Evidence for at least two T3-related ESR sequences on M2

comes from the *SphI* digest, which gave fragments of 9 and 3 kb (Fig. 5C). No T5-related ESR sequences were detected on this chromosome. Hybridization of *SphI*-digested M1 DNA (PFGE compression zone) with T3 and T5 ESR probes resulted in 10 bands (5 bands unique for each probe) (Fig. 5C). These same bands (in addition to those from M2, M3, and M4) were seen in digestions of total VAT 7 DNA hybridized with a T3 ESR-derived riboprobe (data not shown), suggesting that some, if not all, of the M1 ESR sequences originated from breakage of the large chromosomal DNA. In addition to the major bands described above, faint hybridization signals were also seen with M1, M2, M3, and M4 chromosomal DNA digests (Fig. 5C). These bands could be due to additional ESR-related sequences, more distantly related sequences such as ingi, contamination during chromosome isolation, or DNA degradation. Overall, we have identified at least 12 and possibly 18 ESRs in the trypanosome genome, which approximates the number of predicted expression sites in this serodeme (21).

FIG. 4. PFGE-Southern blot analysis of the T3 and T5 ESR shows ESRs on all megabase-sized chromosomes. VAT 7 DNA was separated by PFGE, stained with ethidium bromide (A), transferred to nylon, and hybridized with T3 (B) or T5 (C) ESR-derived riboprobes (see Fig. 2 for location of probes). Chromosomal bands, indicated at the left, are M1 (2.0 Mb), M2 (1.7 Mb), M3 (1.5 Mb), M4 (1.4 Mb), intermediate (I) (375 kb), and minichromosomes (mi) (50 to 150 kb). Large chromosomes (Ch) (>3 Mb) remain in the well.

Southern blot analysis revealed no evidence for DNA rearrangements within a 20-kb region of the expression site (spanning the ESR) following antigenic switches. Total genomic DNA from different VATs representing direct lineages of antigenic switches (Fig. 6A) was digested with several restriction enzymes, both frequently and infrequently cutting, and hybridized with T3 and/or T5 ESR probes. No obvious changes in migration of bands of hybridization were found among different VATs (data not shown).

ESR transcripts. Northern blots of total RNA from several bloodstream form VATs and procyclic forms derived from VATs 7 and 5 showed that stable ESR RNA was present in bloodstream, but not procyclic, forms (Fig. 6B), indicating that ESR transcript accumulation is stage specific. Hybridization with an α -tubulin probe confirmed the presence of undegraded RNA in all VATs and procyclic forms (Fig. 6B). T3 and T5 ESR probes hybridized with a single major transcript in all bloodstream VATs examined, although the size and intensity differed, presumably because of sequence variation of the ESR among expression sites (Fig. 6B). Hybridization of Northern blots with a T3 ESR-specific antisense riboprobe indicated that the transcripts were from the same strand as the ESAGs downstream of the putative expression site promoter (data not shown). Northern blots containing poly(A)⁺-enriched RNA hybridized with a T3 ESR-specific random-primed probe showed an identical hybridization pattern to that obtained with total RNA, indicating that the ESR transcripts are polyadenylated (data not shown).

The size of the ESR transcript varied between VATs, and the ESR transcript size was characteristic of the expression site used. An ESR transcript of ~2.7 kb accumulated in VATs expressing the T3 expression site of chromosome M4 (VATs 7 and 3^{A7}) and the T6 expression site of chromosome M2 (VAT 11). This transcript was also observed in VAT A⁷, which expressed the T3 expression site following a switch involving gene conversion, and in a VAT (11^{A11}) which reexpressed the T6 expression site (data not shown). A transcript of ~3.6 kb was present in VATs (A⁷_L, A⁵, and A¹¹) in which an antigenic switch resulted in telomeric

FIG. 5. ESR-associated restriction fragment length polymorphisms in VSG expression sites. (A) Schematic representation of expression site gene organization. (B) Restriction maps of the T2, T3, and T5 expression sites (ES), 5' to ESAG 7. The proposed expression site transcription initiation site is marked with a flag, and the 5' end of ESAG 7 is boxed. The predicted location of ESAG 7 in the T2 expression site of chromosome M3 (not sequenced) is indicated by a dashed line. Sequenced regions are shown by thick lines; remaining restriction enzyme sites (in regions indicated by thin lines) were derived from genomic Southern analyses. In the latter case, only the most 3' site of each enzyme is shown. Only selected *TaqI* sites are shown. Thick underlines indicate the locations of the T3 or T5 ESR-derived random primed probes (see Materials and Methods) used for Southern blot analysis of chromosomes M1, M2, M3, and M4. Restriction enzyme abbreviations: B, *Bam*HI; Bl, *Bgl*II; C, *Cla*I; E, *Eco*RI; Ev, *Eco*RV; Hc, *Hinc*II; Hd, *Hind*III; R, *Rsa*I; S, *Sal*I; Sp, *Sph*I; Ss, *Sst*I; and T, *Taq*I. (C) Southern blot analysis of chromosomes M1 to M4 (lanes 1 to 4, respectively) with different restriction enzymes and probes. T3- and T5-specific *Sph*I bands from M4 are indicated by open and solid arrows, respectively.

activation (21) of the T2 expression site (Fig. 6A). The minor bands seen in VATs 7 and 3^{A7} may be due to contamination with other VATs or heterogeneity in spliced leader (SL) or polyadenylation sites. Analysis of cDNA clones obtained by PCR amplification of ESR mRNA confirmed the sequence differences in the transcripts from different VATs. Thus, ESR expression appears to be regulated in a VAT-specific fashion, suggesting that it is coordinated with expression of the VSG and associated expression site genes.

FIG. 6. Northern blot analysis indicates that ESR transcripts are regulated in a VAT- and stage-specific manner. (A) Diagrammatic representation of VAT switches in the IsTaR 1 serodeme (see reference 21 for details). The names and size of the chromosomes involved in antigenic switches are indicated, the telomeric expression sites (ES) are circled, and VATs expressing each expression site are shown to the left and right. Changes in the active expression site after antigenic switches are indicated by solid arrows; dashed arrows represent reexpression of the initial expression site following subsequent switches. For example, the T6 expression site on chromosome M2 is active in VAT 11, and after an antigenic switch, the T2 expression site of chromosome M3 is active (VAT A¹¹). The subsequent switch to VAT 11^{A11} reactivated the T6 expression site. (B) Total RNA (10 μ g) from selected VATs and from VAT 5 and 7-derived procyclic RNA (P5 and P7) was DNase I (RNase-free) treated, separated on a 1.0% formaldehyde gel, transferred to nylon, and hybridized with a T3 ESR-derived random-primed probe (see Materials and Methods). Lanes are identified by their chromosomal (Ch) and telomeric (ES) expression site location and VAT. Major transcripts (3.6 and 2.7 kb) are indicated by arrows. Spurious hybridization caused rRNA shadowing is indicated by a brace. The lower panel shows the same blot after stripping and rehybridizing with an α -tubulin probe to indicate the relative concentration of RNA in each lane. RNA size markers (BRL) are shown in kilobases to the left.

Boundaries of ESR transcripts. The 5' limit of the ESR transcript was determined by reverse transcription of mRNA from VATs 7 (T3 expression site), A⁷_L and A¹¹ (T2), 11^{A11} (T6), and 5^{A3} (T7) with XSC-dT₁₇ and PCR amplification, using primers derived from the T3 ESR sequence and the SL sequence (see Materials and Methods). Hybridization of the PCR products with a T3-derived probe revealed a band of ~1.2 kb in all VATs. Thus, the 5' ends of the ESR transcripts map ~5.1 kb upstream of ESAG 7 (Fig. 2A). Sequence from two T3 ESR cDNA clones (pTb1.7c-RT3a and pTb1.7c-RT3b) was identical to that from the corresponding regions of genomic clones e9.1 and pTb1.7g-RT3, confirming their origin from the T3 expression site. The sequence comparison also eliminated the possibility that the PCR products resulted from fortuitous priming with the SL primer. The SL addition sites in the two cDNA clones differed by 16 nt (Fig. 2B), but both occurred immediately downstream of a SL addition site consensus sequence (AG) (30). The upstream SL addition site was used in the single cDNA clones from VATs A⁷_L and A¹¹ (T2 expression site), and 11^{A11} (T6), while the single cDNA from VAT 5^{A3} used the downstream site. Since the 5' end of both the 2.7- and 3.6-kb ESR transcripts (from the T3 or T6 and T2 expression sites, respectively) map to the same location, variation in transcript length is presumably caused by differences in the polyadenylation sites between transcripts.

Comparison of the ESR with the ingi-3 retroposon (Fig. 2A) suggests that the 3' limit of the ESR may lie more than 1 kb downstream of the region with significant sequence homology to ingi. Oligo(dA)₈ and oligo(dA)₇ tracts located ~0.9 and ~1.4 kb, respectively, downstream of the region of homology with ingi (Fig. 2) are suggestive of the poly(A) tracts usually found at the 3' ends of non-LTR retroposons (28). The oligo(dA)₈ tract corresponds to the 3' end of the 2.7-kb ESR transcript (from the T3 and T6 expression sites) predicted from the SL addition site identified above, while the oligo(dA)₇ tract corresponds to the predicted 3' end of the 3.6-kb transcript (from the T2 expression site). Thus, these tracts probably mark the 3' boundary of the ESR element. The more 3' tract is only ~30 bp upstream of the putative ES promoter region (Fig. 2B). Non-LTR retroposons are usually flanked at their 5' and 3' boundaries by short repeat sequences (28). A pentanucleotide sequence (CGAGT) which occurs immediately downstream of the oligo(dA)₇ tract is also found 90 bp upstream of the more 5' ESR SL addition site (Fig. 2B, shaded). These direct repeats may represent the boundaries of the ESR element.

DISCUSSION

In the present study, we have identified a family of ESRs whose transcripts accumulate in a life cycle stage- and VAT-specific fashion, suggesting that they are expressed as part of the VSG expression site. ESR is most closely related to the ingi retroposon of *T. brucei* (15, 33). Analysis of the RT domain indicates ESR is more closely related to mammalian LINE elements than to the trypanosomatid siteposons CZAR (40), SLACS (1), and CRE1 (9). Not surprisingly, there is less similarity to the Gypsy and copia LTR elements and the human immunodeficiency virus type 1 retrovirus.

The two ESRs studied in greatest detail are from the T3 and T5 expression sites (21). These ESRs are located approximately 30 bp upstream of a sequence proposed as the initiation site for expression site transcription (23, 43) and shown to have promoter function in transient transfection

assays (14, 42). Restriction mapping studies indicate that the T2 expression site (in chromosome M3) also contains an ESR in a location similar to that of the T3 expression site. The ESRs are present on many, if not all, megabase-sized chromosomes, but not on intermediate or minichromosomes in the IsTaR 1 serodeme. Southern analyses detect 12 to 18 ESRs, but other ESRs which were not detected because of fragment comigration may also be present. Thus, the number of ESRs is similar to the number of estimated VSG expression sites and those examined in detail are all upstream of the putative expression site promoter.

The active expression site in VAT 118a of the MiTaR 1 serodeme (10, 11) appears very similar throughout the promoter region to the T2, T3, and T5 expression sites of the IsTaR 1 serodeme on the basis of restriction and transcript mapping and sequence comparisons (97 and 95% homology to T3 and T5, respectively, over 450 bp). Analysis of sequence from the 118 expression site in VAT 118a (9a) showed that the ESR occurs in this expression site in the same location as the T3 and T5 expression sites. The 118 expression site in VAT 118a contains a second copy of the promoter 13 kb upstream of the first (11). In the antigenic switch from VAT 118a to VAT 1.8c, a genomic rearrangement leads to the loss of one copy of the promoter and the region between promoters, including the section of the ESR with homology to ingi (Fig. 2A). Thus, the ESR with homology to ingi is no longer found in the 118 expression site in VAT 1.8c or the subsequent VAT 118a' and is also absent in the 221 expression site (Fig. 2A). In these expression sites, the region which corresponds in location to the 5' ESR in the T3, T5, and 118a expression sites contains tandem repeats of a 50-bp sequence (10, 41). The exact site of the 3' end of these simple sequences varies between expression sites (Fig. 2A), suggesting that the region 3' to the region of homology between ESR and ingi is an area of frequent recombination.

Recombination between expression sites can result in antigenic switches by either gene conversion (13, 24) or telomeric exchange (22), and genomic rearrangements, such as changes in telomere size which may involve recombination, are often associated with antigenic switches involving telomeric activation (20, 32, 39). Thus, it is possible that recombination plays a role in all antigenic switches. This hypothesis is attractive since heteroduplexes that are resolved back to the original or nearly original sequences would not be recognized as recombinations but might result in an antigenic switch that would appear to be a telomeric activation. How a recombinational process could result in antigenic switching is unclear in the case of telomeric activation. Subtle sequence changes in an expression site may result in base modification or changes in chromatin structure producing a change in the transcriptional status of the expression site. Alternatively, the recombinational process may result in a change in the nuclear localization of an expression site, possibly affecting its transcriptional status. An attractive feature of a recombination-mediated mechanism is that it could involve the two DNA sequences that are affected in an antigenic switch; one is activated and the other is inactivated. Despite the attractiveness of this model for regulation of VSG expression site transcription, the absence of detectable rearrangements in the IsTaR 1 serodeme suggests that rearrangement may not be essential for antigenic switching. However, we have not eliminated the possibility of sequence rearrangements elsewhere, such as those farther upstream upon activation of the 118 expression site in the VAT 1.8c or VAT 1.8b switch to VAT 118a' or VAT 118b'.

(41). Alternatively, subtle rearrangements, involving very small or no changes in expression site sequence, may also have escaped detection in our analyses.

It seems likely that the different ESR transcript sizes seen in different VATs are based on differences in polyadenylation sites resulting from small sequence variation among expression sites. Thus, it is likely that the larger transcript (3.6 versus 2.7 kb) in VATs (A⁷_L, A⁵, and A¹¹) which express the T2 ESR is caused by sequence divergence in the region of the polyadenylation site in the T2 ESR compared with that of T3, although these sequences appear very similar elsewhere (unpublished data). The accumulation of ESR transcripts in a stage-specific fashion and the production of transcripts of characteristic sizes and sequences in a VAT-specific fashion imply that the expression of the ESR is coordinated with that of the ESAGs and VSG gene. The coordinated expression of the ESR upstream of the putative promoter in the T3 expression site (Fig. 2) with the ESAG and VSG sequence downstream suggests that transcription of this expression site initiates upstream of the ESR, perhaps at a second, upstream copy of the promoter region, as in the case of the 118a expression site (11). This raises the question of whether the downstream promoter is active in this expression site, which has important implications for the mechanisms that regulate expression of the expression site. It is likely that factors other than primary sequence determine promoter activity, since numerous inactive expression sites contain promoter sequences that are identical or nearly identical to that in the active expression site. Thus, it is likely that the genomic context of the promoter sequence is important in determining its activity.

ACKNOWLEDGMENTS

We thank Karyn Carlson, Jeannette Bertles, Tony Morales, and Ken Wilson for DNA sequencing assistance and Augustine Souza for assistance with sequence analysis. We also thank Keith Gottesdiener and Lex van der Ploeg for providing sequence data prior to publication.

This work was supported by Public Health Service grant AI 17375 from the National Institutes of Health. K.S. is a Burroughs Wellcome Scholar in Molecular Parasitology.

REFERENCES

1. Aksoy, S., T. M. Lalor, J. Martin, L. H. T. van der Ploeg, and F. F. Richards. 1987. Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. *EMBO J.* 6:3819-3826.
2. Barry, J. D., and C. M. R. Turner. 1991. The dynamics of antigenic variation and growth of African trypanosomes. *Parasitol. Today* 7:207-211.
- 2a. Cabot, E. L., and A. T. Beckenbach. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Comput. Appl. Biosci.* 5:233-234.
3. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156-159.
4. Cross, G. A. M. 1990. Cellular and genetic aspects of antigenic variation in trypanosomes. *Annu. Rev. Immunol.* 8:83-110.
5. Crozatier, M., L. H. T. van der Ploeg, P. J. Johnson, J. Gommers-Ampt, and P. Borst. 1990. Structure of a telomeric expression site for variant specific surface antigens in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 42:1-12.
6. Cully, D. F., H. S. Ip, and G. A. Cross. 1985. Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. *Cell* 42:173-182.
7. Donelson, J. E., and A. C. Rice-Ficht. 1985. Molecular biology of trypanosome antigenic variation. *Microbiol. Rev.* 49:107-125.
8. Fournay, R. M., J. Miyakoshi, R. S. Day III, and M. C. Paterson.

1988. Northern blotting: efficient RNA staining and transfer. *Focus* 10:5-7.
9. Gabriel, A., and J. D. Boeke. 1991. Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. *Proc. Natl. Acad. Sci. USA* 88:9794-9798.
- 9a. Gottesdiener, K., et al. Unpublished data.
10. Gottesdiener, K., H.-M. Chung, S. D. Brown, M. G.-S. Lee, and L. H. T. van der Ploeg. 1991. Characterization of VSG gene expression site promoters and promoter-associated DNA rearrangement events. *Mol. Cell. Biol.* 11:2467-2480.
11. Gottesdiener, K. M., L. Goriparthi, J. P. Masucci, and L. H. T. van der Ploeg. 1992. A proposed mechanism for promoter-associated DNA rearrangement events at a variant surface glycoprotein gene expression site. *Mol. Cell. Biol.* 12:4784-4795.
12. Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki. 1986. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature (London)* 321:625-628.
13. Hoijmakers, J. H. J., A. C. C. Frasch, A. Bernards, P. Borst, and G. A. M. Cross. 1980. Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. *Nature (London)* 284:78-80.
14. Jefferies, D., P. Tebabi, and E. Pays. 1991. Transient activity assays of the *Trypanosoma brucei* variant surface glycoprotein gene promoter: control of gene expression at the posttranscriptional level. *Mol. Cell. Biol.* 11:338-343.
15. Kimmel, B. E., O. K. ole-MoiYoi, and J. R. Young. 1987. Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. *Mol. Cell. Biol.* 7:1465-1475.
16. Marlor, R. L., S. M. Parkhurst, and V. G. Corcos. 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol. Cell. Biol.* 6:1129-1134.
17. Milhausen, M., R. G. Nelson, M. Parsons, G. Newport, K. Stuart, and N. Agabian. 1983. Molecular characterization of initial variants from the IsTaT1 serodeme of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 9:241-254.
18. Mount, S. M., and G. M. Rubin. 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol. Cell. Biol.* 5:1630-1638.
19. 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.
20. Myler, P. J., R. F. Aline, Jr., J. K. Scholler, and K. D. Stuart. 1988. Changes in telomere length associated with antigenic variation in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 29:243-250.
21. Myler, P. J., R. F. Aline, Jr., J. K. Scholler, and K. D. Stuart. 1988. Multiple events associated with antigenic switching in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 29:227-241.
22. Pays, E. 1985. Gene conversion in trypanosome antigenic variation. *Prog. Nucleic. Acid Res. Mol. Biol.* 32:1-26.
23. Pays, E., H. Coquelet, P. Tebabi, A. Pays, D. Jefferies, M. Steinert, E. Koenig, R. O. Williams, and I. Roditi. 1990. *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. *EMBO J.* 9:3145-3151.
24. 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.
25. Pays, E., P. Tebabi, 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.
26. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85:2444-2448.
27. Poch, O., I. Sauvaget, M. Delarue, and N. Tordo. 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 8:3867-3874.
28. Rogers, J. H. 1985. The origin and evolution of retroposons. *Int. Rev. Cytol.* 93:187-279.
29. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1988. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
30. Scholler, J. K., R. F. Aline, Jr., and K. D. Stuart. 1988. Variant specific transcripts from the co-transposed segments of variant surface glycoprotein genes in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 29:89-103.
31. Scholler, J. K., S. G. Reed, and K. Stuart. 1986. Molecular karyotype of species and subspecies of *Leishmania*. *Mol. Biochem. Parasitol.* 20:279-293.
32. Shea, C., D. J. Glass, S. Parangi, and L. H. T. van der Ploeg. 1986. Variant surface glycoprotein gene expression site switches in *Trypanosoma brucei*. *J. Biol. Chem.* 261:6056-6063.
33. Smiley, B. L., R. F. Aline, Jr., P. J. Myler, and K. Stuart. 1990. A retroposon in the 5' flank of a *Trypanosoma brucei* VSG gene lacks insertional terminal repeats. *Mol. Biochem. Parasitol.* 42:143-152.
34. 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.
35. 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*. Wiley-Liss, Inc., New York.
36. Stuart, K., E. Gobright, L. Jenni, M. Milhausen, L. Thomashow, and N. Agabian. 1984. The IsTaR 1 serodeme of *Trypanosoma brucei*: development of a new serodeme. *J. Parasitol.* 70:747-754.
37. Thomashow, L. S., M. Milhausen, W. J. Rutter, and N. Agabian. 1983. Tubulin genes are tandemly linked and clustered in the genome of trypanosome brucei. *Cell* 32:35-43.
38. van der Ploeg, L. H. T. 1991. Control of antigenic variation in African trypanosomes. *New Biol.* 3:324-330.
39. Van der Werf, A., S. Van Assel, D. Aerts, M. Steinert, and E. Pays. 1990. Telomere interactions may condition the programming of antigen expression in *Trypanosoma brucei*. *EMBO J.* 9:1035-1040.
40. Villanueva, M. S., S. P. Williams, C. B. Beard, F. F. Richards, and S. Aksoy. 1991. A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. *Mol. Cell. Biol.* 11:6139-6148.
41. Zomerdiijk, J. C. B. M., R. Kieft, M. Duynham, P. G. Shiels, and P. Borst. 1991. Antigenic variation in *Trypanosoma brucei*: a telomeric expression site for variant-specific surface glycoprotein genes with novel features. *Nucleic Acids Res.* 19:1359-1368.
42. Zomerdiijk, J. C. B. M., R. Kieft, P. G. Shiels, and P. Borst. 1991. Alpha-amanitin-resistant transcription units in trypanosomes: a comparison of promoter sequences for a VSG gene expression site and for the ribosomal RNA genes. *Nucleic Acids Res.* 19:5153-5158.
43. Zomerdiijk, J. C. B. M., M. Ouellette, A. L. M. A. ten Asbroek, R. Kieft, A. M. M. Bommer, C. E. Clayton, and P. Borst. 1990. The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *EMBO J.* 9:2791-2801.