MOLECULAR AND CELLULAR BIOLOGY, Jan. 1995, p. 87–93 Vol. 15, No. 1
0270-7306/95/$04.00+0
Copyright © 1995, American Society for Microbiology

A Selector of Transcription Initiation in the Protozoan Parasite *Toxoplasma gondii*

DOMINIQUE SOLDATI AND JOHN C. BOOTHROYD*

Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305-5402

Received 17 May 1994/Returned for modification 7 July 1994/Accepted 27 September 1994

The recent development of an efficient transfection system for the apicomplexan *Toxoplasma gondii* allows a comprehensive dissection of the elements involved in gene transcription in this obligate intracellular parasite. We demonstrate here that for the *SAG1* gene, a stretch of six repeated sequences in the region 35 to 190 bp upstream of the first of two transcription start sites is essential for efficient and accurate transcription initiation. This repeat element shows characteristics of a selector in determining the position of the transcription start sites.

As an obligate intracellular parasite, *Toxoplasma gondii* is an important pathogen of the human fetus and of certain groups of immunocompromised patients. In immunocompetent individuals, the acute infection is controlled by the immune system; however, some rapidly growing tachyzoites are able to escape and survive through differentiation into slowly growing encysted forms called bradyzoites. The cysts remain in the host during its lifetime and are probably responsible for reactivation of infection in immunocompromised hosts. Since the interconversion of tachyzoites to bradyzoites has become a major clinical concern in reactivation of toxoplasmosis in AIDS patients, understanding the mechanisms that trigger cell differentiation is an extremely important issue.

Important morphological and metabolic changes occur during this differentiation, and these changes are accompanied by the stage-specific alteration of gene expression (12). *SAG1* codes for the stage-specific major surface antigen of *Toxoplasma* tachyzoites: anti-SAG1 antibodies detect a protein homogeneously distributed on the surface of both extracellular and intracellular tachyzoites (9, 13), whereas no protein is detectable on bradyzoites or sporozoites (11, 26). Recent studies demonstrated that SAG1 is involved in host cell adhesion and invasion (18). In addition, studies have shown that immunization with SAG1 induces protective immunity in mice (5, 14). Bradyzoite surface antigens have also been identified (11, 26), but the corresponding genes have not yet been reported.

SAG1 is a single-copy gene, containing no introns. Two initiation sites for transcription of this gene, separated by ~35 bp, have been reported (6). The polyadenylated transcript is abundant in tachyzoites (6), but the manner of its developmental regulation has not been determined. As a first step toward understanding the structure and organization of gene transcription and regulation in *T. gondii*, we have chosen to dissect the upstream region of the developmentally regulated *SAG1* gene. The initial cloning and sequencing of the gene did not reveal the presence of the higher eukaryotic TATA promoter element (6). However, an element composed of six tandemly repeated, conserved 27-bp sequences was identified just upstream of the two transcription start sites (35 bp separate the end of the repeats from the first start site, which is in turn ~35 bp upstream of the second site [Fig. 1]) (6). Using recently developed transfection technology (25), we show here that these repeated sequences constitute a promoter element essential for high-level expression of *SAG1* and that they serve as a positioning element, directing the initiation of transcription.

* Corresponding author. Phone: (415) 723-7984. Fax: (415) 723-6853.

MATERIALS AND METHODS

***Toxoplasma* cells.** The parasites used in this study were tachyzoites of the laboratory strain RH of *T. gondii* (4). Parasites were grown in monolayer cultures of human foreskin fibroblasts (HFF) cultured at 37°C in Dulbecco modified Eagle medium (Gibco) containing 10% NuSerum (Collaborative Biomedical Products).

Transfection studies. Transfection of tachyzoites was done essentially as previously described (25). Briefly, the parasites were harvested from freshly lysed cultures of infected HFF, and transfections were performed with 10^7 parasites in a volume of 800 μl of cytomix (27) containing 25 nM plasmid DNA. The DNA concentration used for electroporation was chosen on the basis of the strength of the promoter under study and the tolerance of the electroporator so that the pulse length for each sample would be invariant. Unless otherwise noted, all experiments used supercoiled plasmid DNA. Electroporated parasites were transferred back to fresh HFF culture and incubated at 37°C for 24 h before preparation of whole-cell extracts for assay of chloramphenicol acetyltransferase (CAT) activity.

There is no other readily assayable reporter available for transient expression studies with *T. gondii*. Therefore, to control for the possibility of an error in estimating the DNA concentration (and thus the relative expression level), all transfection experiments were repeated multiple times, using different DNA preparations for each construct. Results from a representative experiment are presented in each figure.

Selection of stable transformants. Parasites (10^7) from strain RH were transfected with 20 μg of supercoiled plasmid DNA and transferred back to HFF culture. At 12 h after electroporation, the parasites were subjected to 20 μM chloramphenicol selection as described previously (15). After 7 to 10 days of selection, stable populations of parasites resistant to chloramphenicol emerged; for all constructs used, high CAT activity was detected in the parasite populations.

DNA sequence analysis. DNA sequence analysis was performed by using a modified T7 DNA polymerase (Sequenase version 2.0) as instructed by the manufacturer (United States Biochemical).

RNA preparation. Parasites that had recently lysed the host culture (one 175-cm^2 flask per experiment) were centrifuged at 400 × *g*, and the culture medium was discarded. After one wash in cold phosphate-buffered saline, the pellet of cells was collected and total RNA was purified with the Ultraspec RNA isolation system (Bioteck Laboratories) as instructed by the manufacturer.

Primer extension analysis. RNA from ~10^8 parasites was resuspended in water and divided into two aliquots. Each sample was then mixed with 2 pmol of 5′ ^{32}P-labeled primer 1 (5′-ATGTTCTTTACGATGCGATTGG-3′) or primer 2 (5′-AACGGTGGTATATCCAGTGATT-3′). Both primers hybridize within the *CAT* sequence; primer 1 hybridizes 30 bp downstream of primer 2. Samples were denatured by heating to 90°C for 3 min and then placed at 60°C for 15 min and at 37°C for 30 min to anneal. The primer extension was performed at 37°C for 60 min with 200 U of SuperScriptII RNase H$^-$ reverse transcriptase (GIBCO BRL). One half of each reaction mixture was analyzed on a sequencing gel.

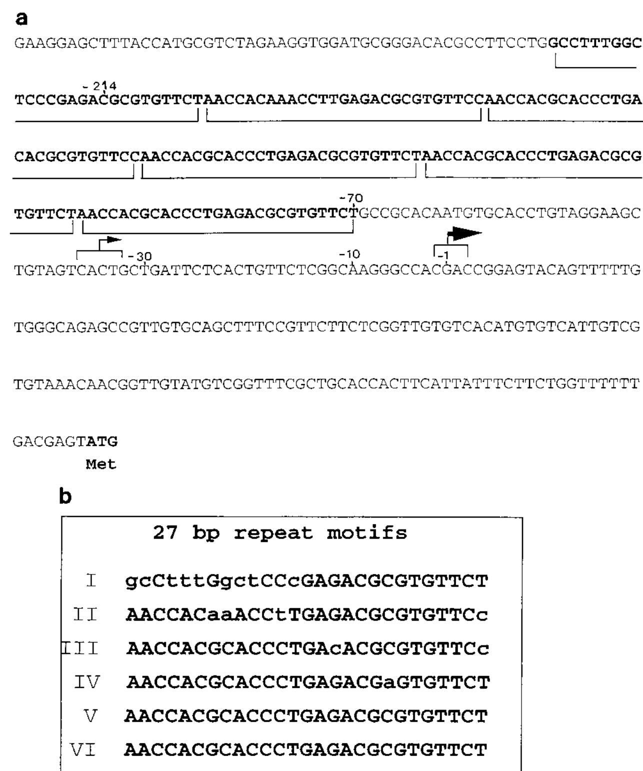

FIG. 1. (a) Nucleotide sequence of the sense strand of *SAG1* 5′ flanking region from 282 bp upstream of the major transcription start site down to the translation initiation codon (from reference 6). The six repeats are in boldface and are underlined. The minor and major transcription start sites are indicated with light and dark arrows, respectively. Numbering is from the middle of the downstream initiation site. (b) Alignment of the six 27-bp repeat sequences present in the *SAG1* promoter. The nucleotides in lowercase do not match the repeat consensus sequence.

Vectors used in transfection studies. (i) 5′ deletion analysis. The plasmid constructs used in all experiments contain the 600-bp coding sequence of *CAT* and 300 bp of the 3′ flanking sequence of the *SAG1* gene, including the polyadenylation site. They are schematically shown in each figure. Plasmid pS/4K contains a 4,000-bp fragment of *SAG1* 5′ flanking sequence extending from the initiation sites upstream to the *Sal*I site in P30.5 COS1 (6). Plasmids pS/800 and pS/214 correspond to the previously described SAG1/2CAT and SAG1′/2CAT, respectively (25). pS/800 contains ~800 bp of 5′ upstream flanking sequences, starting from the major initiation site of *SAG1*. pS/214 contains 214 bp upstream of the major initiation site including five complete repeat motifs.

Deletion of one to four repeats in pS/4R to pS/0R was done by exchanging the *Hin*dIII-*Nsi*I fragment in pS/800 with the fragments obtained from a PCR on pS/800, using primers A (ggcgaagcttGAGACGCGTGTTC) and B (ggcatgCAT ACTCGTCAAAAAACCAGAA) (the bases in lowercase do not correspond to *SAG1* sequences and include restriction sites). The PCR gave rise to six distinct fragments containing 0.5 to 5.5 repeats with *Hin*dIII and *Nsi*I sites at their extremities for cloning. The construct pS/0R contains only half of a repeat motif (14 bp).

Plasmid pS/70 starts downstream of the repeats and was generated by the same PCR strategy, using primers B and D (ggcgaagcttTGCCGCACAATGTGCA CCT).

(ii) Orientation effect analysis. A cassette fragment containing five repeats with a *Hin*dIII restriction site at each end was generated by PCR using pS/800 as the template and primers E (ggcgaagcttAACCACAAACCTT) and F (gctcaagct tGAACACGCGTCTC). This five-repeat cassette was inserted in pS/70 in both orientations, using *Hin*dIII, to generate pS/5Rf and pS/5Rr.

(iii) Distance effect analysis. Constructs pS/D5Rf and pS/D5Rr were generated by cloning the five-repeat cassette (see above) in both orientations into the unique *Sac*I site located downstream of the 300-bp 3′ flanking sequences of *SAG1* in plasmid pS/70. For this purpose, both the vector and the insert were previously treated with T4 polymerase and blunt end cloned. To generate constructs in which the repeat element is moved further upstream relative to the transcription start sites, the five-repeat cassette was first blunt end cloned into the *Kpn*I site of pS/70, positioning the repeats 35 bp upstream from their initial locations to create construct pS/5R35. Subsequently, this plasmid was linearized at the *Cla*I restriction site, located in the polylinker immediately downstream of the repeats, and spacers of different sizes (generated by complete digestion of a Bluescript [Stratagene] plasmid with the restriction enzyme *Taq*I) were inserted. The spacer fragments were identified by sequencing, and their exact sizes were assigned. Plasmids pS/5R35, pS/5R131, and pS/5R292, with 35-, 131-, and 292-bp spacers, respectively, introduced downstream of the repeat element were chosen for promoter analysis.

Insertion of the repeat elements in a heterologous promoter. The five-repeat element was introduced in the pT/230 expression vector derived from the *T. gondii* α-tubulin gene described earlier and previously referred to as TUB1/CAT (25). The cassette was cloned into the *Hin*dIII site 345 bp upstream of the *TUB1* transcription initiation site, to generate construct pT5R/230. Construct pT/70 is a 5′ deletion mutant of pT/230, obtained by substitution of the 500-bp *Hin*dIII-*Nsi*I fragment of pT/230 with a 360-bp *Hin*dIII-*Nsi*I fragment, generated by PCR using the primers G (ggcgaagcttGCCTGCATTGGGTGCGGTTG) and H (ggat gcatTTTGTCGAAAAAGGGAAT), with pT/230 as the DNA template. The repeat cassette was inserted in both orientations in the *Hin*dIII site of pT/70 to generate pT/5Rf70 and pT/5Rr70.

Assay of CAT activity from *Toxoplasma* lysate. CAT assays were performed with the phase partition method described previously (25), and the activities were measured and compared only in the linear range of the assay.

RESULTS

Promoter mapping by 5′ deletion analysis of *SAG1*. To identify the *cis*-acting sequences required for proper functioning of the *SAG1* promoter, we generated a 5′ deletion series and compared them in a transient transfection assay using *CAT* as the reporter gene (Fig. 2). Comparison between different constructs was done on the basis of multiple experiments and with the same molar amount of construct DNA. The relative activity seen with different constructs was constitutively observed. For example, in three experiments, constructs pS/4R and pS/2R gave, respectively, 8 to 18% and 55 to 70% of the activity seen with the full-length construct. A region containing 214 bp upstream of the apparent major transcription start site was determined to contain a functional promoter, whereas the sequences extending further upstream, in constructs pS/800 and pS/4K, did not give any significant increase in CAT activity.

Within the promoter region thus identified is an element composed of six almost identical 27-bp repeats ending 70 bp upstream of the major transcription start site (Fig. 1). Deletion of the entire element in pS/70 results in only minimal basal CAT activity (Fig. 2). As the differences between the constructs all lie outside the transcribed region, changes in CAT activity levels are presumed to reflect relative promoter strength. This point is further addressed below. Analysis of constructs containing different numbers of 27-bp repeats shows that at least two are required to detect a stimulation of transcription over the basal level, as indicated in Fig. 2. Each additional element yields a roughly incremental increase in activity up to six repeats. A further doubling of the activity is obtained when five additional repeats are added tandemly to the five complete repeats present in construct pS/214 (data not shown). Thus, these results suggest a *cis*-acting element in the *SAG1* promoter essential for expression.

To exclude the possibility that the effect seen on deletion of successive repeats was due to bringing some chance negative regulating element in the upstream vector sequence ever closer to the *SAG1* initiator, the experiments were repeated for four of the constructs (pS/2R, pS/3R, pS/4R, and pS/5R), using linearized DNA in the transfections (linearization was at the *Hin*dIII site, just upstream of the repeats). The strict correlation between the number of repeats and CAT activity was retained with the two-, three-, four-, and five-repeat-containing, linearized constructs giving 8, 33, 50, and 93% of the CAT seen with circular pS/5R (data not shown). This result strongly suggests that it is indeed the absence of repeats rather than a juxtaposition of a negative regulator that is responsible for the diminishing activity upon successive deletion.

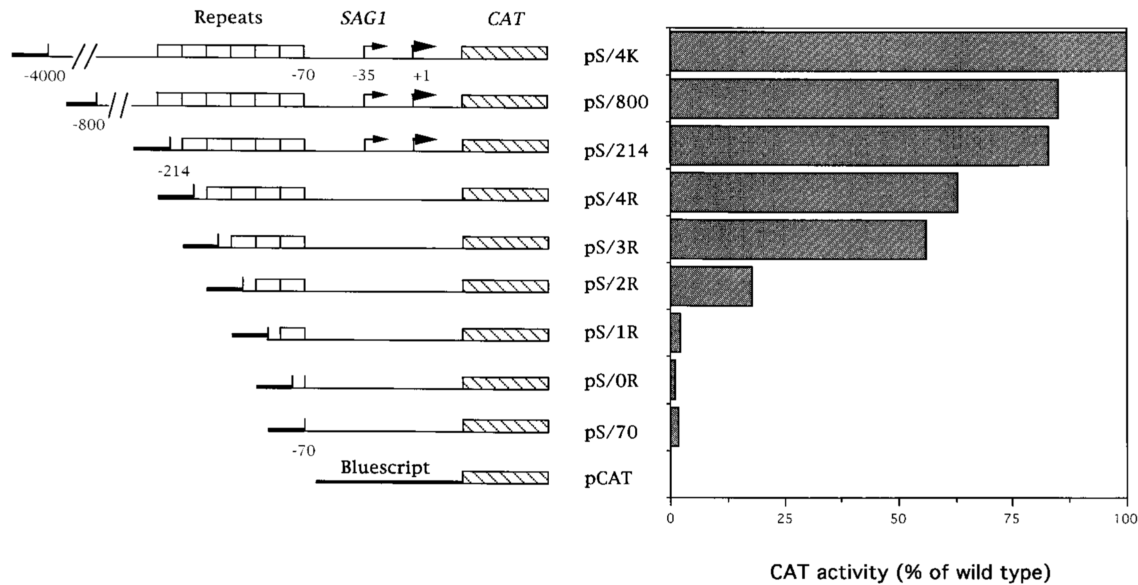

FIG. 2. Transient CAT activity assays driven by the *SAG1* promoter. The plasmids are progressive deletions from pS/4K to pS/70. The construct pS/0R extends to position -83 (Fig. 1) and contains only a partial repeat (14 bp). The control plasmid pCAT does not contain any *SAG1* sequences. The level of CAT activity of each construct is expressed as a percentage of that of pS/4K. Transcription start sites are shown for those constructs that have been directly analyzed (data not shown and Fig. 4). They are presumed to be the same for the remaining constructs but have not been directly determined. The darker arrow indicates the major start site.

Orientation and position effect. To determine whether the repeat element can function in both orientations, a cassette composed of five repeats flanked by *Hin*dIII restriction sites was used to generate the reverse-orientation mutant, pS/5Rr (plasmid bearing the *SAG1* promoter with five repeats in reverse orientation). This plasmid gave levels of activity essentially similar to those of the comparable forward construct (pS/5Rf), indicating that the element functions in an orientation-independent fashion (Fig. 3). The introduction of the *Hin*dIII site to clone the five-repeat cassette fragment in pS/70 did not significantly affect the CAT activity compared with the wild-type pS/214 construct (data not shown).

Influence of the distance relative to the transcription start sites has been addressed in two ways. First, the five-repeat element was inserted into a vector sequence in construct pS/70, ~1 kb downstream of the transcription start sites, in the forward and reverse orientations to give pS/D5Rf and pS/D5Rr, respectively. In that position, the element is not able to restore any significant *SAG1* promoter activity (Fig. 3).

The second approach was to change the distance of the repeat element relative to the initiation sites in small increments. To do this, random spacer sequences derived from the Bluescript plasmid digested with *Taq*I and of increasing size were inserted immediately downstream of the repeat element. In constructs pS/5R35, pS/5R131, and pS/5R292, in which the spacers inserted were 35, 131, and 292 bp long, respectively, CAT activities dropped to 68, 61, and 8%, respectively, relative to the wild-type pS/5Rf construct (Fig. 3). These values could be due to effects at any of a number of levels (e.g., transcriptional or posttranscriptional). To discriminate between these

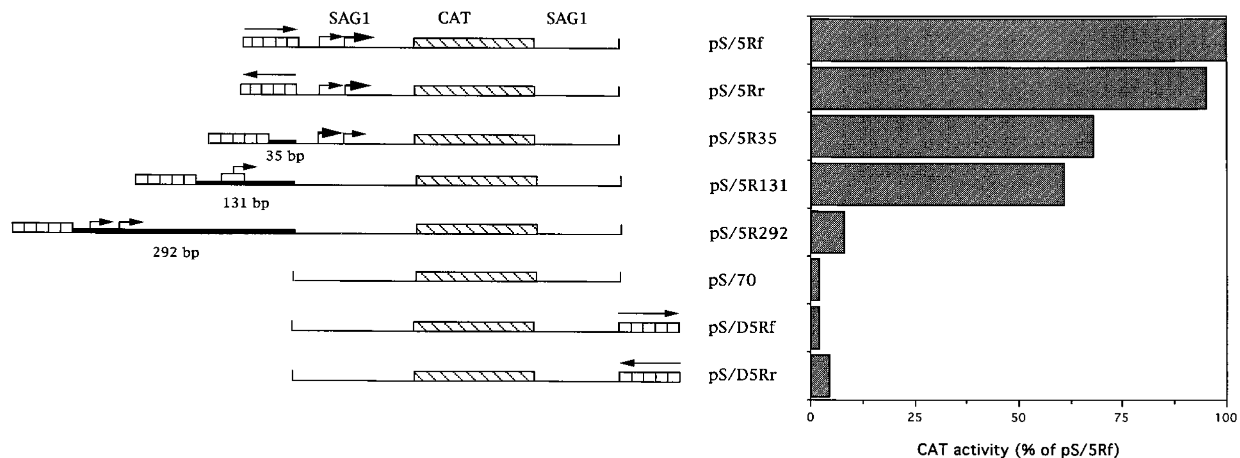

FIG. 3. Transient CAT activity assays driven by *SAG1* mutant constructs, in which the repeat element is positioned in inverted orientation and/or moved either downstream of the gene or further upstream relative to the initiation sites. The spacers introduced downstream of the repeats are indicated by thick lines, and their sizes are indicated below. The level of CAT activity is expressed as a percentage of that of construct pS/5Rf. Transcription start sites as determined by primer extension (Fig. 4 and data not shown for construct pS/5R292) are indicated as in Fig. 2.

alternatives, we mapped the transcription sites used in each construct.

Determination of transcription start sites. Two discrete sites of transcriptional initiation in the *SAG1* gene have been previously determined by nuclease protection and primer extension analysis (6). In the context of *SAG1* promoter analysis, it was necessary to determine first that the same initiation sites were used by the parasites in constructs where the *SAG1* coding sequence was replaced by the *CAT* sequence and second whether the modifications introduced in the mutated promoter constructs affect the positioning of these initiation sites.

Primer extension analysis was performed on total RNA harvested from *T. gondii* clones stably transformed with the pS/4K, pS/5Rr, pS/5R35, and pS/5R131 constructs (Fig. 3; note that pS/5R292 did not generate enough CAT activity to allow selection of stable clones since such selection relies on enough CAT being present to give chloramphenicol resistance [15]). Two primers were used in the analysis to exclude any primer-specific artifact. These hybridize to *CAT* sequences downstream of the *SAG1* promoter, about 30 bases apart. Except for the expected shift in size, exactly analogous results were obtained with the two primers (compare lanes 1 to 4 and 6 to 9 in Fig. 4), indicating that all bands are the result of extension on the target RNA. Similarly, controls using RNA from untransfected parasites gave no signal, again indicating the specificity of the extension products (Fig. 4, lanes 5 and 10). For simplicity, therefore, the results for primer 1 only will be discussed.

Five clusters of strong stops, S1 to S5, are seen with the pS/4K construct which has 4,000 bp of upstream sequence including the repeats in their natural positions (Fig. 4, lane 4). Two of these sites (S1 and S2) correspond to RNA species of the expected size for initiation of transcription occurring at the natural sites previously mapped (6). The third site, S3, maps within *SAG1* sequences downstream of the S1 and S2 sites, but a comparable signal is not seen with a *SAG1* antisense oligonucleotide (6), and thus this signal may represent an unnatural start site derived from the mixing of the *SAG1* promoter with the *CAT* coding region. As discussed further below, the presence of S1 and S2 and absence of S3 signals in lane 2, which corresponds to pS/5Rr (five repeats in reverse orientation), indicates that S3 could not be a result of processing of the S1- and/or S2-derived RNAs (the sequences downstream of the repeats are identical in pS/4K and pS/5Rr, and thus any processing events seen for one should be seen for the other).

The S4 and S5 sites also are not seen when a *SAG1* primer is used (6) (note also that all five clusters are within *SAG1* sequences). Additionally, S4 and S5 are seen in all lanes in which extension products are detected, albeit only after overexposure in some cases, including lanes in which no signal is seen for S1, S2, and S3 (e.g., lanes 3 and 8). Thus, the simplest explanation for these RNAs is that they are a result of degradation of longer transcripts, regardless of where they initiate. Consistent with this conclusion, they were not seen in other experiments in which RNA from transient transfections using identical plasmids was analyzed in an identical manner except for the method of RNA preparation (data not shown).

Extension on RNA from the pS/5Rr construct (lanes 2 and 7) in which the five repeats are in reverse orientation gives initiation at both the S1 and S2 sites in about the same proportion to that seen with the wild-type pS/4K (S1 is clearly visible on longer exposure). Interestingly, S3 is not detectably used in the pS/5Rr construct, indicating it is dependent on a natural orientation of the repeats (but only when juxtaposed to *CAT* sequences). The mechanism for this is unclear.

Primer extension analysis was also applied to parasites stably transformed with constructs pS/5R35 and pS/5R131, in which

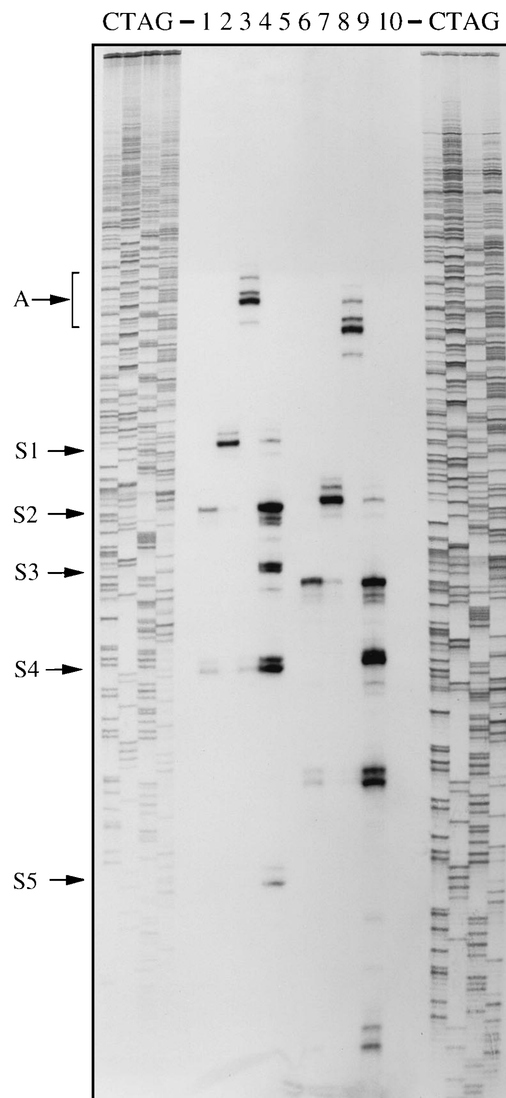

FIG. 4. Mapping of the *SAG1* transcription start sites in *CAT* expression vectors by primer extension using two antisense oligonucleotides specific for the *CAT* sequence. Primers 1 and 2 were used for the extensions shown in lanes 1 to 5 and 6 to 10, respectively. Primer 1 corresponds to *CAT* sequence 30 bp downstream of that corresponding to primer 2. RNA was derived from parasites stably transformed with pS/5Rr (i.e., five repeats in the reverse orientation; lanes 1 and 6), pS/5R35 (i.e., five repeats separated from the initiation sites by insertion of 35 bp of vector sequence; lanes 2 and 7), pS/5R131 (carrying an insertion of 131 bp; lanes 3 and 8), pS/4K (with ~4,000 bp of *SAG1* upstream sequence; lanes 4 and 9), and control, untransformed parasites (lanes 5 and 10). The sequence reactions are derived from pS/4K, using primer 1 on the left and primer 2 on the right, and thus both give an antisense sequence. The primer extension products generated by the transformant harboring pS/4K are labeled S1 to S5; S1 and S2 are the natural *SAG1* initiation sites. The cluster of extensions seen for pS/5R131 is labeled A.

the repeat element was placed progressively further upstream. Primer extension on the transcripts produced from pS/5R35 gave rise to the products seen in Fig. 4, lane 2. For this construct, only the S1 site is appreciably used, with S2 being detectable only on overexposure of the gel (not shown). Thus, shifting the repeats 35 bp upstream apparently restricts initiation to almost exclusively the most upstream natural site.

For pS/5R131 (lane 3), in which 131 bp have been inserted between the repeats and the normal initiation sites, none of the

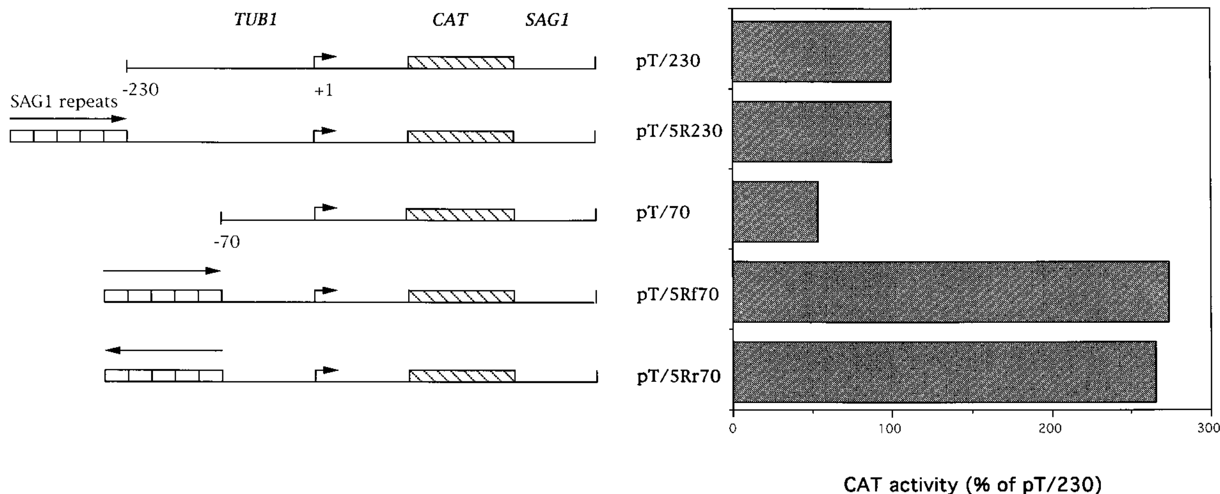

FIG. 5. Modulation of the transient CAT activity driven by *TUB1* promoters with or without the insertion of the *SAG1* repeat element. The level of CAT activity is expressed as a percentage of that of the pT/230 construct.

normal stops (S1 to S3) are detected except S4, which, as already discussed, may be a cleavage site. Instead, a new cluster of bands (marked A in Fig. 4) is seen, indicative of transcripts initiating ~130 bp upstream of the major S2 initiation site seen with the wild-type *SAG1*.

A quantitative analysis of these data is not possible since the efficiency of primer extension is dependent on the length of the product and the recovery of total RNA is variable because of the small number of parasites handled. In fact, analysis of the yield of total RNA from each parasite line showed the same relationship as seen in Fig. 4, i.e., pS/4k < pS/5R35 ~ pS/5R131 < pS/5Rr with respect to both total RNA yield and total primer extension product.

Analysis of RNA from parasites transiently transfected with the pS/5R292 construct similarly showed a shift in the length of the products by about the size of the insert (data not shown; CAT expression in pS/5R292 is insufficient to allow selection of stable transformants).

These results demonstrate that the repeat element determines the position of the transcription start sites. In the light of these data, the CAT activity obtained with constructs pS/5R35, pS/5R131, and pS/5R292 (Fig. 3) could readily be explained by differences in the translation efficiency of the different mRNAs seen for each construct. For example, dramatic differences could result from utilization of cryptic translation start sites present in the spacer sequences.

The repeat element stimulates transcription from a heterologous promoter. We were interested to determine whether this transcriptional element would be capable of stimulating expression from a heterologous promoter; for this purpose, we started with the pT/230 expression vector derived from the α-tubulin gene (*TUB1*) of *T. gondii* and containing 230 bp upstream of the transcription start site (19). The pT/230 construct was previously called TUB1/CAT and used to establish transient transfection protocols (25). A *TUB1* promoter deletion mutant, pT/70, containing 70 bp of 5′ flanking sequence upstream from the transcription start site was also constructed. The pT/70 construct shows 50% less activity than pT/230 but is still very efficiently expressed in *T. gondii* and may constitute the core promoter of *TUB1*.

The repeat element was introduced upstream of the *TUB1* 5′ flanking sequences in both pT/230 and pT/70 to generate plasmids pT/5R230, pT/5Rf70, and pT/5Rr70 (the first two have the five repeats in the forward orientation; the last has them reversed). The potential of the repeat element to modulate the heterologous promoter activity was evaluated in transient transfection assays. A fivefold stimulation in CAT activity was observed when the five-repeat element was inserted in both orientations at a short distance (70 bp) but not at a longer distance (230 bp) upstream of the transcriptional start site (Fig. 5). Primer extension analysis on pT/230, pT/70, pT/5Rf70, and pT/5Rr70 confirmed that the initiation site of transcription (T1) is unaltered; i.e., there is no effect on the start site position of substituting the *TUB1* coding sequence with CAT or of inserting the repeat element (data not shown). Presumably, in these pT/70-based constructs, the repeats are appropriately positioned (i.e., 70 bp upstream) to stimulate expression from the natural initiator.

DISCUSSION

Within the group of protozoan parasites, analysis and characterization of promoters have been limited. The lack of DNA transfection in apicomplexans (e.g., *Toxoplasma* and *Plasmodium* species) previously hampered determination of the elements involved in their gene transcription. In members of the Kinetoplastida (e.g., *Leishmania* and *Trypanosoma* species), for which transfection is available, promoter analysis is complicated by the fact that transcription involves the synthesis of large polycistronic precursors from which mature RNAs are produced by *trans* splicing (reviewed in reference 8). Those promoters that have been well characterized in trypanosomes drive genes that code for surface antigens variant surface glycoprotein and procyclic acidic repetitive protein) and are unusual in apparently being transcribed by RNA polymerase I (Pol I) (7, 21, 30). (Although in most systems Pol I transcribes only the rRNA genes, the trypanosome Pol I can be used to transcribe protein-coding genes, because a 5′ cap, essential for mRNA functioning, can be added posttranscriptionally by *trans* splicing to the pre-mRNA.) A putative actin gene promoter is the only Pol II promoter found so far in trypanosomes, but in experiments reported to date it shows only very weak activity (2), and no detailed characterization has yet been performed.

There is no evidence of polycistronic transcription or *trans* splicing in *T. gondii*, facilitating the identification of promoters and rendering unlikely the possibility that a polymerase other

than Pol II is responsible for the transcription of protein-coding genes in this organism. Until recently, analysis of promoter function in *T. gondii* has been restricted to sequence comparison, and such analyses do not reveal any obvious common conserved elements that might potentially interact with the transcription machinery and none of the well-characterized proximal/distal elements identified in higher eukaryotic promoters, including the TATA box, seemed to be present. TATA-less promoters have generally been found associated with housekeeping genes (reviewed in reference 23), and multiple sites of transcription initiation are often seen (reviewed in reference 20).

A highly divergent TATA-binding protein (TBP) has been characterized in the related apicomplexan parasite *Plasmodium falciparum* (17). Thus, it is very likely that a TBP exists in *T. gondii* and that if it is involved in *SAG1* promoter function, it recognizes a noncanonical TATA box or works via indirect interaction with another, DNA-binding protein. Several cases in which, in the absence of a TATA box, regulatory factors are believed to tether TFIID (of which TBP is part) or other components of the preinitiation complex to the promoter and participate in start site selection have been reported. For example, the Sp1 site directs initiation of transcription in the hamster dihydrofolate reductase promoter to about 45 bp downstream (3), as do four binding sites for the glucocorticoid receptor (22, 29). The activator proteins recognizing these sites are believed to determine the specific initiation start site through protein-protein interactions (16). Our results are consistent with the hypothesis that the *SAG1* repeat element functions in a similar manner, serving as a selector to position the RNA polymerase around 35 to 70 bp downstream of the element. The fact that the repeats can be inverted and still drive transcription of *CAT* suggests that they naturally may drive divergent transcription of sequences upstream of *SAG1*. This possibility has not yet been tested, and there are no data on the possible coding function of this upstream region.

In most eukaryotes, transcription typically starts at an A within several pyrimidine nucleotides called the initiator (10). It is generally accepted that TBP cooperates with the initiator and plays a central role in transcription from yeasts to humans (28). A hierarchy of steps in the initiation process is believed to take place: the selector binds or tethers TFIID to the promoter in the first step of initiation and controls the position of the preinitiation complex, and then the initiator interacts with a common component of the complex and controls the exact start site. The results obtained here suggest that the *SAG1* repeat element plays a predominant role in the initiation process but that there is considerable flexibility in the initiator, since some cryptic sequences at the proper distance are apparently found in all of the random spacer sequences introduced and analysis of sites S1 to S3 did not reveal any obvious similarity even with respect to the nucleotide where initiation apparently occurs at each.

Study of the developmental regulation of gene expression in *T. gondii* has previously been extremely difficult because of the obstacle of obtaining significant amounts of bradyzoite-containing tissue cysts from infected animals. The concomitant development of both a transfection system and a protocol for in vitro differentiation of tachyzoites to bradyzoites (e.g., see reference 24) provides the ingredients necessary to investigate the mechanisms involved in the expression of stage-specific genes. Also, the fact that this *SAG1* repeat element is able to stimulate transcription by using a heterologous promoter provides the ideal constructs to address the role of the repeats in stage-specific control of gene expression.

ACKNOWLEDGMENTS

We extend special thanks to S. Tomavo and K. Kim for sharing materials and helpful discussions and to Edward Mocarski and Brendan McCormack for critical reading of the manuscript.

This work was supported by the National Institutes of Health (AI21423), the John D. and Catherine T. MacArthur Foundation, and the Burroughs Welcome Fund and by fellowships from the European Molecular Biology Organization and the Swiss National Science Foundation.

REFERENCES

1. **Bangs, J. D., P. F. Crain, T. Hashizume, J. A. McCloskey, and J. C. Boothroyd.** 1992. Mass spectroscopy of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J. Biol. Chem. **267:**9805–9815.
2. **Ben Amar, M. F., D. Jefferies, A. Pays, N. Bakalara, G. Kendall, and E. Pays.** 1991. The actin gene promoter of Trypanosoma brucei. Nucleic Acids Res. **19:**5857–5862.
3. **Blake, M. C., R. C. Jambou, A. G. Swick, J. Kahn, and J. C. Azizkhan.** 1990. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. Mol. Cell. Biol. **10:**6632–6641.
4. **Boothroyd, J. C., K. Kim, E. R. Pfefferkorn, L. D. Sibley, and D. Soldati.** Forward and reverse genetics in the study of the obligate intracellular parasite *Toxoplasma gondii*. Methods Mol. Genet., in press.
5. **Bulow, R., and J. C. Boothroyd.** 1991. Protection of mice from fatal *Toxoplasma gondii* infection by immunization with P30 in liposomes. J. Immunol. **147:**3496–3500.
6. **Burg, J. L., D. Perelman, L. H. Kasper, P. L. Ware, and J. C. Boothroyd.** 1988. Molecular analysis of the gene encoding the major surface antigen of Toxoplasma gondii. J. Immunol. **141:**3584–3591.
7. **Clayton, C. E., J. P. Fueri, J. E. Itzakhi, V. Bellofatto, D. R. Sherman, G. S. Wisdom, S. Vijayasarathy, and M. R. Mowatt.** 1990. Transcription of the procyclic acidic repetitive protein genes of *Trypanosoma brucei*. Mol. Cell. Biol. **10:**3036–3047.
8. **Cross, G. A. M.** 1990. Cellular and genetic aspects of antigenic variation in trypanosomes. Annu. Rev. Immunol. **8:**83–110.
9. **Dubremetz, J. F., C. Rodriguez, and E. Ferreira.** 1985. Toxoplasma gondii: redistribution of monoclonal antibodies on tachyzoites during host cell invasion. Exp. Parasitol. **59:**24.
10. **Grosschedl, R., and M. L. Birnstiel.** 1980. Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of the H2A histone gene transcription *in vivo*. Proc. Natl. Acad. Sci. USA **77:**7102–7106.
11. **Kasper, L. H.** 1989. Identification of stage specific antigen of *Toxoplasma gondii*. Infect. Immun. **57:**668–672.
12. **Kasper, L. H., M. S. Bradley, and E. R. Pfefferkorn.** 1984. Identification of stage-specific sporozoite antigens of *Toxoplasma gondii* by monoclonal antibodies. J. Immunol. **132:**443–449.
13. **Kasper, L. H., J. H. Crabb, and E. R. Pfefferkorn.** 1983. Purification of a major membrane protein of Toxoplasma gondii by immunoadsorption with a monoclonal antibody. J. Immunol. **130:**2407–2412.
14. **Khan, I. A., K. H. Ely, and L. H. Kasper.** 1991. A purified parasite antigen (P30) mediates CD8+ T-cell immunity against fatal *Toxoplasma gondii* infection in mice. J. Immunol. **147:**3501–3506.
15. **Kim, K., D. Soldati, and J. C. Boothroyd.** 1993. Gene replacement in Toxoplasma gondii with chloramphenicol acetyltransferase as selectable marker. Science **262:**911–914.
16. **Kollmar, R., and P. J. Farnham.** 1993. Site-specific initiation of transcription by RNA polymerase II. Proc. Soc. Exp. Biol. Med. **203:**127–139.
17. **McAndrew, M. B., M. Read, P. F. Sims, and J. E. Hyde.** 1993. Characterization of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T-rich human malaria parasite Plasmodium falciparum. Gene **124:**165–171.
18. **Mineo, J. R., R. McLeod, D. Mack, J. Smith, I. A. Khan, K. H. Ely, and L. H. Kasper.** 1993. Antibodies to *Toxoplasma gondii* major surface protein (SAG1, P30) inhibit infection of host cells and are produced in murine intestine after peroral infection. J. Immunol. **150:**3951–3964.
19. **Nagel, S. D., and J. C. Boothroyd.** 1988. The a- and b-tubulins of *Toxoplasma gondii* are encoded by single copy genes containing multiple introns. Mol. Biol. Parasitol. **29:**261–273.
20. **Pavé-Preux, M., M. Aggerbeck, C. Veyssier, B. Bousquet-Lemercier, J. Hanoune, and R. Barouki.** 1990. Hormonal discrimination among transcription start sites of aspartate aminotransferase. J. Biol. Chem. **265:**4444–4448.
21. **Pays, E., H. Coquelet, A. Pays, P. Tebabi, and M. Steinert.** 1990. Trypanosoma brucei: constitutive activity of the VSG and procyclin gene promoters. EMBO J. **9:**3145–3151.
22. **Schatt, M. D., S. Rusconi, and W. Schaffner.** 1990. A single DNA-binding transcription factor is sufficient for activation from a distant enhancer and/or from a promoter position. EMBO J. **9:**481–487.
23. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57:**103–113.
24. **Soete, M., D. Camus, and J. F. Dubremetz.** In vitro expression of bradyzoite

specific proteins and cyst formation by the RH strain of Toxoplasma gondii in the absence of immunomodulators. Exp. Parasitol., in press.
25. **Soldati, D., and J. C. Boothroyd.** 1993. Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii.* Science **260:**349–352.
26. **Tomavo, S., B. Fortier, M. Soete, C. Ansel, D. Camus, and J. F. Dubremetz.** 1991. Characterization of bradyzoite specific antigens of *Toxoplasma gondii*. Infect. Immun. **59:**3750–3753.
27. **Van den Hoff, M. J. B., A. F. M. Moorman, and W. H. Lamers.** 1992. Electroporation in intracellular buffer increases cell survival. Nucleic Acids Res. **20:**2902.
28. **White, R. J., and S. P. Jackson.** 1992. the TATA-binding protein: a central role in transcription by RNA polymerases I, II, and III. Trends Genet. **8:** 284–287.
29. **Wieland, S., M. D. Schatt, and S. Rusconi.** 1990. Role of TATA element in transcription from glucocorticoid receptor-responsive model promoters. Nucleic Acids Res. **18:**5113–5118.
30. **Zomerdijk, J. C. B. M., M. Ouelette, A. L. M. A. Ten Asbroek, R. Kieft, M. M. A. 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.