An upstream element of the TamS1 gene is a site of DNA-protein interactions during differentiation to the merozoite in Theileria annulata

Brian Shiels1,* , Mark Fox2, Sue McKellar1, Jane Kinnaird1 and David Swan1

1Department of Veterinary Parasitology, University of Glasgow, Glasgow, G61 1QH, UK
2Seattle Biomedical Research Institute, Seattle, WA 98109-1651, USA
*Author for correspondence (e-mail: B.Shiels@vet.gla.ac.uk)

SUMMARY

Apicomplexan parasites are major pathogens of humans and domesticated animals. A fundamental aspect of apicomplexan biology, which may provide novel molecular targets for parasite control, is the regulation of stage differentiation. Studies carried out on Theileria annulata, a bovine apicomplexan parasite, have provided evidence that a stochastic process controls differentiation from the macroschizont to the merozoite stage. It was postulated that this process involves the presence of regulators of merozoite gene expression in the preceding stage of the life cycle, and that during differentiation a quantitative increase of these factors occurs. This study was carried out to test these postulations. Nuclear run-on analysis showed that TamS1 expression is controlled, at least in part, at the transcriptional level. The transcription start site showed homology with the consensus eukaryotic initiator motif, and study of the 5’ upstream region by the electrophoretic mobility-shift assay demonstrated that a 23 bp motif specifically bound factors from parasite-enriched nuclear extracts. Three complexes were shown to bind to a 9 bp core binding site (5’-TTTGTAGGG-3’). Two of these complexes were present in macroschizont extracts but were found at elevated levels during differentiation. Both complexes contain a polypeptide of the same molecular mass and may be related via the formation of homodimer or heterodimer complexes. The third complex appears to be distinct and was detected at time points associated with the transition to high level merozoite gene expression.

Key words: Apicomplexan parasite, Stage differentiation, DNA binding factor, Upstream motif

INTRODUCTION

Parasites classified within the subphylum apicomplexa are major pathogens of vertebrates and include the genera Babesia, Eimeria, Plasmodium, Theileria and Toxoplasma (Cox, 1993). A fundamental aspect of apicomplexan biology is the progressive differentiation through a series of cellular stages to generate a life cycle. The molecular mechanisms that regulate this process are not fully understood. This knowledge may be important because experimental data suggests that the transition phase from one stage to another is sensitive to manipulations that can alter the outcome of parasitic infection and control associated disease (Shiels et al., 1998). Differentiation of a number of apicomplexans, in vitro, can be defined as stochastic (Shiels, 1999). In these systems the random probability of a differentiation step occurring can be increased by altering culture conditions. In addition, apicomplexan differentiation is often associated with a reduction in parasite proliferation potential, and continued culture frequently brings about a loss of differentiation capability. These shared characteristics suggest that, for certain apicomplexan differentiation events, a basic mechanism may operate across a range of parasite species (Shiels et al., 1998; Shiels, 1999 and references therein).

The generation of extracellular uninucleated (merozoite) stages from an intracellular multinucleated (schizont) stage is a differentiation step that is common to different apicomplexan life cycles. We have studied this event in Theileria annulata, an apicomplexan parasite of bovines. In vivo the macroschizont stage of Theileria resides within the bovine leukocyte and induces host cell proliferation (Hulliger, 1965). Following a particular time period, host cell division slows down and the macroschizont differentiates into multiple merozoites (Jarret et al., 1969). The merozoites are released as the host cell is destroyed, invade erythrocytes and differentiate into piroplasms. Differentiation to the merozoite of T. annulata in vitro has been analysed using a system consisting of two parasite-infected cloned cell lines, one of which generates merozoites with high efficiency when cultured at 41°C, while the other is severely attenuated with respect to this process (Shiels et al., 1992, 1994). Comparative study of these cell lines provided evidence that the ability to differentiate correlated with the build-up in expression of a major merozoite surface antigen gene, TamS1, which was also expressed at a lower level in the macroschizont stage (Shiels et al., 1994). Low-level TamS1 expression was found to be elevated during an early transitory phase and was reversible until a condition was reached that committed the cell to high-level merozoite
gene expression and merozoite formation. The correlation between upregulation of TamS1 gene expression and merozoite formation was also observed in experiments where the time taken to reach commitment was altered by the addition of specific drugs (Shiels et al., 1997). It was concluded from these studies that the differentiation mechanism may operate on the basis of attaining a particular predetermined condition over time, and that an increase in the ratio of polypeptide factor(s) relative to DNA levels influences the rate of progression of the cell towards this condition. It has been demonstrated, in many eukaryotic systems, that changes to the profile of gene expression following cellular differentiation can be mediated by interactions between polypeptide factors and short nucleotide motifs, normally located upstream of the transcription initiation point (Latchman, 1991; Blau, 1992). Therefore, based on the results outlined above, it was proposed in a *Theileria* model that factors which positively control TamS1 gene expression are present in the preceding stage of the life cycle (the macroschizont), and that elevated expression involves an increase in the level of factor(s) relative to their DNA templates until an unknown commitment threshold is reached (Shiels et al., 1994, 1997). Then, at commitment, a qualitative change in factor regulation was postulated to account for the switch to irreversible high-level TamS1 gene expression. The present study was carried out to test the above model by identifying factors that bind to nucleotide motifs of the TamS1 gene during differentiation to the merozoite.

**MATERIALS AND METHODS**

**Cell culture and preparation of nuclear extracts**

The uninfected BL20 cell line and the cloned macroschizont-infected cell line, D7, were maintained in culture as described (Shiels et al., 1992). Induction to differentiate at 41°C was performed under standard conditions (Shiels et al., 1994), except in differentiation time-course experiments where the following modifications were introduced. D7 cells were washed in serum-free RPMI (Gibco) medium and seeded at 2×10^6 cells ml⁻¹ in RPMI-20% calf serum (Sigma) plus standard supplements (Shiels et al., 1992) and placed at 41°C. At day 2, the cultures were diluted to 2×10⁷ cells ml⁻¹ in fresh RPMI-cool calf. Following a further 2 days at 41°C the cultures were centrifuged (300 g) and resuspended in RPMI supplemented with 10% foetal calf serum. The cultures were then either replaced at 41°C or at 37°C and cells harvested on day 6. This protocol was adopted because it generated a higher level of cells displaying differentiation morphology than standard conditions (data not shown). Three day-6 time points are represented in this study; D7/day 6I and D7/day 6II are duplicate cultures where the cells were returned at 37°C after 4 days at 41°C. This procedure was previously shown to generate cultures that contained a relatively synchronous population of cells committed to merozoite production and nondifferentiating macroschizont-infected cells (Shiels et al., 1994). The D7/ day 6III culture was a normal asynchronous culture incubated at 41°C throughout.

Parasite- and host-enriched nuclear extracts were generated essentially as described (Swan et al., 1999). To extract nuclear proteins, host-enriched and parasite-enriched nuclear pellets were taken up in 500 μl and 300 μl, respectively, of ice-cold extraction buffer (5 mM HEPES, pH 7.9, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF) and 3 M NaCl added to a final concentration of 300 mM. Extraction was carried out on ice for 30 minutes and the chromatin pelleted by centrifugation (14000 g for 5 minutes). The estimated protein concentration was in the range of 0.4-1.5 μg/ml for parasite-enriched extracts while the host extracts, in general, contained five- to tenfold more protein.

**Isolation and sequence analysis of TamS1, TpmS1 and TsmS1 genomic DNA clones**

A genomic clone containing the TamS1 gene was isolated from a λ DASH library of genomic DNA derived from merozoites of the D7-infected cloned cell line. Screening by hybridisation with a TamS1 cDNA probe was performed using standard protocols (Shiels et al., 1994; Church and Gilbert, 1984). The λ DASH insert (approx. 20 kb) was restriction mapped by Southern blotting (Sambrook et al., 1989) and hybridisation (Church and Gilbert, 1984) using probes representing the 5’ and 3’ ends of the cDNA. A 3.2 kb *HindIII*-SaI fragment and a 4.0 kb *HindIII*-SalI fragment were identified as being represented, respectively, the 5’ and 3’ flanking regions of the TamS1 gene. These fragments were cloned and completely sequenced on both strands using a LiCor automated sequencer and a combination of *EcoRI* deletions (Erase-a-Base®, Promega) plus subcloning of restriction fragments. This generated an 8.0 kb sequence contig, with 3.0 kb of 5’ sequence and 4.1 kb of 3’ sequence flanking the TamS1 open reading frame (accession number: AJ276654 for EMBL, GenBank™ and DDJB data bases).

To isolate a genomic clone of the TpmS1 gene, a minilibrary was constructed with DNA isolated from piroplasms of *T. parva* (Brocklesby et al., 1961). A 6.0 kb *HindIII* fragment was detected on Southern blots with both the 5’ and 3’ TamS1 cDNA probes (Fox, 1997). *HindIII*-digested DNA was then separated by agarose electrophoresis and DNA fragments of approximately 6.0 kb electroeluted from a 1 cm gel slice. Elutip column purified DNA was then ligated into the λ ZAP Express vector and packaged with Gigapack II packaging extract as detailed by the supplier (Stratagene). The *T. parva* minilibrary was screened by hybridisation with the TamS1 cDNA probe at 59°C and positive phage isolated. The pBl-CMV phagemid was excised following the manufacturer’s protocol and contained a *HindIII* insert of approximately 5.5 kb. The TpmS1 flanking regions were then subcloned and sequenced (Fox, 1997). Once completed a 5.6 kb contig containing the protein coding sequence plus 2.08 kb upstream and 2.7 kb downstream sequence was generated (accession number: AF043074 for EMBL, GenBank™ and DDJB data bases).

Attempts were made to PCR amplify the 5’ flanking region of the TsmS1 gene (synonym TS110) from *T. sergenti* genomic DNA using standard protocols (Sambrook et al., 1989; Shiels et al., 1995) and reaction conditions of 95°C for 4 minutes followed by 25 cycles of 1 minute at 95°C (denaturation), 1 minute at 45°C (annealing) and 1 minute at 72°C (elongation). The TsmS1 primer 5’ cttgtctgacactaaggtgtc 3’ located within the protein coding region of the TsmS1 gene (+343 to +320, relative to the transcription start site) and a range of primers located to the 5’ upstream region of TsmS1 were used. Of these only one, CAT-4a (5’acaatttgtagggcga3’, –54 to –49), in combination with TsmS1DB, generated a PCR product. This product was 388 bp in length, and generated an additional 66 bp of sequence upstream of the 5’ end of the published TS110 cDNA sequence (Kawazu et al., 1992).

**Mapping of transcription start site and nuclear run-on sequence analysis**

To map the 5’ terminus of TamS1 mRNA, rapid amplification of cDNA ends (RACE) was employed using a commercial system, following the supplier’s protocol (Gibco-BRL). The piroplasm stage expresses TamS1 mRNA at high levels (Swan et al., 1999). Total RNA was isolated as described (Chomczynski and Sacchi, 1987) and poly(A)+ mRNA purified using a Poly ATtract mRNA Isolation System (Promega). First-strand cDNA was synthesised using a TamS1 gene-specific oligonucleotide primer (GP1: 5’-
cgaattggaagcatctagttccctggeg 3', +448 to +418, relative to translation start site). Single-stranded cDNA was tailored with poly dC and 30 cycles of PCR amplification carried out using the anchor primer in combination with a second nested gene-specific primer (GSB: 5'gctcgagcgaagcaaggagcagcgc 3', +211 to +182). A PCR product which ran just below the 395 bp marker was detected by Southern blotting using the TamS1 cDNA probe. A further round of PCR amplification was followed by ligation of the product into pCR II vector (Invitrogen) and determination of the sequence.

Nuclear run-on analysis was performed on parasite-enriched nuclear fractions isolated from 10^8 parasite-infected cells, as described above. The run-on reactions were performed essentially as outlined (Kooter et al., 1987) except that SDS was omitted from the proteinase K incubation step and RNA was separated from free nucleotides by a NucTrap® purification column (Stratagene). Radiolabelled RNA was then added to hybridisation solution (Church and Gilbert, 1984) and incubated with slot blots of DNA probes for 48 hours at 60°C. Following two washes in 2x SSC (1x SSC: 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS for 30 minutes and one wash in 0.2x SSC: 0.1% SDS for 30 minutes, membranes were hybridised with radiolabelled probes specific for the target sequences using a Hybri-Slot Manifold (Gibco-BRL), after washing in 0.2x SSC: 0.1% SDS for 30 minutes and one wash in 0.2x SSC: 0.1% SDS for 30 minutes. The membranes were then exposed to X-ray film for 1 hour at −70°C. Slices of gel containing the complexes were then excised and exposed to UV illumination (254 nm) for 15 minutes on ice. The gel slices were then dissected into smaller pieces and incubated in 10 mM Tris (pH 7.5), 0.1% SDS at room temperature overnight. Eluted protein was then TCA-precipitated, washed once in acetone, air-dried and redissolved in 1x SDS-PAGE sample buffer plus 0.1 volume of 1 M Tris (pH 6.8). The samples were then analysed by SDS-PAGE and autoradiography, using standard protocols (Sambrook et al., 1989).

RESULTS

Structural analysis of the TamS1 gene locus

To commence studies aimed at identification of nucleotide motifs and polypeptide factors involved in the control of TamS1 gene expression, genomic DNA fragments corresponding to the 5' and 3' regions that flank the known open reading frame were isolated and sequenced. This resulted in a contiguous sequence of over 8 kb that contained, in addition to TamS1 (Shiels et al., 1995), two incomplete small polypeptide open reading frames, at the 5' and 3' termini respectively, and a small internal ORF of 907 bp, that was located between positions 4909-5816 of the contig. As expected the sequence was in general AT-rich (64%). The region (981 bp) flanking the 3' boundary of the Tam1 open reading frame was comparable with respect to the overall AT content (63%), while the 5' flanking region (2.89 kb) was more AT-rich (70%) and contained 18 homopolymeric (dA:dT) tracts >9 bp. A major transcription initiation site was mapped to an adenosine residue 119 bp upstream of the translation start codon (see Fig. 1A) within a motif (TCA(+1)CTT) that has a good match to the consensus sequence (PyPyA(+1)N(T/A)PyPy) established for the eukaryotic initiator (Inr) element (Smale and Baltimore, 1989; Jawahery et al., 1994).

As nucleotide motifs that regulate gene expression are often conserved across related organisms, regions of DNA that flank the TamS1 homologue from Theileria parva (Tmps1; Shiels et al., 1995) were isolated and sequenced. Comparison of the TamS1 and Tmps1 contigs revealed that the 5' and 3' flanking regions showed the same level of identity (79% for 2.1 kb of 5' sequence; 78% for 2.5 kb of 3' sequence) as the protein coding region (78%) across the two species. Identity was higher across sequences proximal to the translation initiation and the mapped transcription initiation site (85% over 207 bp upstream of ATG) and the translation termination codon (91% over 116 bp downstream of TAA).

Due to the unexpectedly high level of general sequence conservation between the T. parva and T. annulata contigs, identification of conserved putative regulatory motifs was not possible. Therefore, attempts were made to PCR amplify genomic DNA flanking the 5' region of the TsmS1 gene of T. sergenti, and a comparison of the untranslated 3' region, using available cDNA sequences, was undertaken. Only one PCR product was obtained using a range of TamS1 5' primers in combination with the coding region primer. The alignment shown in Fig. 1A revealed that the sequence proximal to the translation start site of the T. sergenti gene was significantly divergent from the TamS1 and Tmps1 sequences. However, two short stretches of identity were obtained. The first of these is

UV crosslinking and SDS-PAGE

Fixation of protein to oligonucleotide probes was carried out essentially as described (Genersch et al., 1995). Preparative gel-shift reactions were performed with 8 µl of D7/Day 6 nuclear extract and 800 fmol of a radiolabelled CAT1 probe, modified by the incorporation of 5-bromodeoxyuridine during synthesis (MWG-BIOTEK) at position 9, 11 and 13 of the top strand. Nine reactions were run on a polyacrylamide gel and the positions of the CAT1-specific complexes determined by exposure of the wet gel to X-ray film for 1 hour at −70°C. Slices of gel containing the complexes were then excised and exposed to UV illumination (254 nm) for 15 minutes on ice. The gel slices were then dissected into smaller pieces and incubated in 10 mM Tris (pH 7.5), 0.1% SDS at room temperature overnight. Eluted protein was then TCA-precipitated, washed once in acetone, air-dried and redissolved in 1x SDS-PAGE sample buffer plus 0.1 volume of 1 M Tris (pH 6.8). The samples were then analysed by SDS-PAGE and autoradiography, using standard protocols (Sambrook et al., 1989).
Regulation of TamS1 occurs at the transcriptional level

Analysis of nuclear preparations by incorporation of 4-6, diamidino-2-phenylindole (DAPI) and fluorescence microscopy, demonstrated an enrichment of either host or parasite nuclei (see Fig. 2). Parasite-enriched nuclear fractions from time points of a differentiation time course were used in nuclear run-on analysis. The resulting radiolabelled nascent RNA was used to probe cDNA representing TamS1, a rRNA gene previously found to be constitutively expressed (Shiels et al., 1994; Swan et al., 1999) and the plasmid pGEM. The autoradiograph showed that between days 2 and 4 of culture at 41°C a small increase in the level of TamS1 transcript occurred relative to the rRNA control gene (Fig. 3A,B). This was then followed by a more substantial increase between days 4 and 6 (Fig. 3B,C). Thus, control of TamS1 gene expression occurs, at least in part, at the transcriptional level. Further nuclear run-on analysis was carried out utilising the TamS1-P fragment representing genomic DNA positioned −116 to −1 upstream of the mapped transcription start site. No signal relative to the cDNA and pGEM control probes was obtained (see Fig. 3D). This result indicated that nucleotide motifs involved in regulation of transcription are likely to be located within the intergenic DNA region upstream of the TamS1 gene.

Identification of a TamS1 upstream nucleotide motif that binds Theileria nuclear factors

To screen for motifs that bind parasite nuclear factors an electrophoretic mobility-shift assay (EMSA) was developed. The assay used parasite- and host-enriched nuclear extracts and DNA probes of approximately 100 bp, PCR-amplified from the TamS1 3′ flanking region. The primers were designed to generate probes that covered the upstream regions (from position −504 to position −1), which showed the highest level of conservation between TamS1 and TpmS1. Designation of a positive result was based on detection of mobility shifts (after an overnight exposure of the autoradiograph) that were more abundant with extracts derived from parasite-enriched nuclear fractions. Probes covering positions −504 to −120 failed to give a result that met these criteria. A positive result was, however, obtained with the fragments proximal to the transcription start site. Thus a probe, TaGATCAT (covering bases −109 to −16), generated two strong complexes (Fig. 4A). The faster complex of the two was found to be associated with parasite-enriched nuclear extracts, while the second complex predominated when the probe was
incubated with host-enriched nuclear extracts. In addition to
the major complexes, both slower (a) and faster (c) less
abundant complexes were detected in extracts derived from
differentiating cultures. This profile was reproduced with the
TamS1-P fragment (−116 to −1), except resolution of complex-
c was not obtained (data not shown).

Analysis of the TaGATCAT probe sequence showed that it
contained blocks of high identity between TamS1 and TpmS1.
To determine if any of these sequence blocks were responsible
for the mobility shifts, competition experiments were carried
out using two cold double-stranded oligonucleotide probes.
The first probe (CAT1) spanned positions −58 to −38 and
included the primer sequence that resulted in successful PCR
amplification from the T. sergenti genomic DNA. CAT1
specifically competed with a subset of shifts that were
associated with parasite-enriched, but not host-enriched,
nuclear fractions (Fig. 4B); the second double-stranded
oligonucleotide (GATA) spanning nucleotides −84 to −59 did
not result in major competition of the probe/extract complexes.
EMSA using radiolabelled CAT1 probe confirmed that this
sequence could directly bind parasite-associated nuclear
factors (Fig. 4C). Three specific complexes were detected with
extracts of parasite-enriched nuclear fractions. The slower
complexes (a and b) were both present in extracts of
nondifferentiating cells but appeared to be more abundant in
extracts derived from differentiating parasite-infected cultures.
The fastest of the three complexes (c) was present at the lowest
levels and could not be detected in extracts derived from cells
cultured at 37°C. Formation of all three CAT1 complexes was
demonstrated to be specific by competition with cold double-
stranded CAT1, in direct comparison to cold double-stranded
GATA. A complex that migrated slightly slower than complex-a
(Fig. 4C, lane 2) was also observed in shifts performed with
parasite-enriched extracts from 37°C cultures. This band was
not detected consistently and was not due to specific complex formation,
as it was not competed by cold
CAT1.

To delineate the binding site of
the CAT1 probe in more detail, mutagenesis was carried out.
Initially, three mutant double-
stranded oligonucleotides were
generated, each differing from the
CAT1 probe by substitution of a
distinct nucleotide triplet as
outlined in Fig. 5A. Substitution of
the triplet ACA (CAT1.M1) did not
have any effect on the formation of
complex-b or complex-c relative to
the wild-type probe, but did result
in a lower level of complex-a
formation. In contrast, substitution
of the second (TGT: CAT1.M2) and
third triplet (GCG: CAT1.M3)
abolished the ability to form all
three parasite-associated complexes
(Fig. 5B). It was concluded from
this analysis that the sequence 5’
ATTTGTAGGGCG 3’ was critical
for CAT1 complex formation. A
fourth mutant oligonucleotide was
generated where the 5’ adenosine or
3’ cytosine and guanine residues of
this sequence were also altered.
This probe (CAT1.M4) reduced the
quantitative level of all three
complexes relative to the wild-type
CAT1 probe but did not abolish their
formation (data not shown). This
result limited the core motif
responsible for complex formation
to the residues 5’ TTTGTAGGG 3’.

The core binding motif defined for TamS1 is also upstream in the

Fig. 2. Analysis of host and parasite-enriched nuclear fractions by DAPI fluorescence.
(1) unfractionated macroschizont-infected cells (D7) cultured at 37°C (day 0). (2) Host-enriched
nuclear fraction of D7/day 0 cells. (3) Parasite-enriched nuclear fraction of D7/day 0 cells.
(4) Unfractionated cells from differentiating culture of D7 cells (day 9 at 41°C). (5) Host-enriched
nuclear fraction of D7/day 9 cells. (6) Parasite-enriched nuclear fraction D7/day 9 cells. Bar, 6.2
µm. h and p denote, respectively, host and parasite nuclei.
sequence, except the single adenosine residue is substituted by a thymidine. EMSA was performed with a double-stranded oligonucleotide (CA T1-P) incorporating this substitution (Fig. 5C). All three complexes were obtained with radiolabelled CA T1-P probe, and cold CA T1-P competed the three complexes formed by the CA T1 probe. While complexes b and c were generated to comparable levels with both probes, the relative level of complex-a appeared to be reduced for the CA T1-P (T. parva) probe. Thus, while all three complexes are likely to be generated by the Cat1-P probe and T. parva extracts, complex-a may be formed with reduced efficiency.

**Characterisation of nuclear factor-CAT motif complex formation during differentiation to the merozoite**

Previous studies have shown that for the majority of parasites commitment to merozoite formation occurs after incubation at 41°C for 4 days (Shiels et al., 1994). To compare the level of complex formation that occurred during differentiation to the merozoite, the CA T1 probe was incubated with parasite-enriched nuclear extracts derived from different points of a differentiation time course. The level of individual complex formation was then quantified by liquid scintillation counting. As can be seen in Fig. 6 and Table 1, the levels of complex-a and complex-b formation showed a respective increase of up to 3.2-fold and 5.4-fold between the day 2 and day 6 time points. To take into account differences in the load of parasite nuclei between samples, DNA was isolated from the parasite-enriched chromatin pellet generated during preparation of the day 2 and day 6 extracts. The DNA was then analysed by Southern blotting, using a probe (1.2 kb 5’ TamS1 HincII genomic subclone) representing the upstream region of TamS1 that contains the CAT1 motif. The TamS1-specific restriction fragment was detected by autoradiography, excised and counted. This analysis showed an increase of up to 1.6-fold in TamS1 DNA between the day 2 and day 6 time points, indicating respective increases of 2.0- to 3.4-fold in the levels of complex-a and complex-b formation relative to their TamS1 DNA template (Table 1). Complex-c was the least abundant of

---

**Fig. 3.** Analysis of TamS1 transcription. Parasite-enriched nuclear fractions were prepared from day 2 (A), day 4 (B) day 6 (C) and day 7 (D) D7 cells cultured at 41°C. In (A-C), nascent radiolabelled RNA was generated and used to probe a range of Theileria DNA fragments: slot 1, pGEM plasmid vector; slot 2, large subunit ribosomal RNA gene; slot 3, a gene (1C7) encoding a macroschizont polypeptide; slot 4, the merozoite rhoptry gene TamR1; slot 5, an asparagine-rich gene; slot 6, TamS1 cDNA. (D) Nascent radiolabelled RNA was used to probe DNA fragments representing: slot 1, TamR1 cDNA; slot 2, upstream region of TamR1 genomic DNA; slot 3, TamS1 cDNA; slot 4, the TamS1-P PCR fragment representing the upstream region (~116 to ~1) proximal to the transcription start site.

**Fig. 4.** EMSA of the 5’ flanking region of TamS1 gene with host- and parasite-enriched nuclear extracts. Closed arrowheads denote parasite-associated complexes, open arrowhead denotes host. (A) EMSA using TaGATCAT probe. Probe alone (lane 1); probe plus: BL20 nuclear extract (lane 2); host-enriched nuclear extract of D7/day 0 cells (lane 3); parasite-enriched D7/day 0 nuclear extract (lane 4); host-enriched D7/day 6 nuclear extract (lane 5); parasite-enriched D7/day 6 nuclear extract (lane 6). (B) Competition with cold double-stranded oligonucleotides. Probe plus: cold Cat1 oligonucleotide preincubated with host-enriched D7/day 0 nuclear extract (lane 1); parasite-enriched D7/day 0 nuclear extract (lane 2) and parasite-enriched D7/day 6 nuclear extract (lane 3); cold GATA oligonucleotide preincubated with host-enriched D7/day 0 nuclear extract (lane 4); host-enriched D7/day 6 nuclear extract (lane 5); and parasite D7/day 6-enriched nuclear extract (lane 6). Note that the parasite-enriched day 6 extract used in this experiment was contaminated with host nuclei. The open arrowhead designates the major host-associated complex. (C) CAT1 oligonucleotide probe with parasite-enriched nuclear extracts. Probe plus: BL20 nuclear extract (lane 1); parasite-enriched D7/day 0 nuclear extract (lane 2); parasite-enriched D7/day 6 extract (lane 3); parasite-enriched D7/day 0 extract and cold competitor CAT1 (lane 4); parasite-enriched D7/day 6 extract and cold competitor CAT1 (lane 5); parasite-enriched D7/day 0 extract and cold competitor GATA (lane 6); parasite-enriched D7/day 6 extract and cold competitor GATA (lane 7).
DNA-protein interactions in *T. annulata*

It was undetectable in all extracts derived from cells cultured at 37°C. Clear detection of this complex was not achieved until day 4 of differentiation time courses with a quantitative increase in the intensity of the gel shift between days 4 and 6.

Mutagenesis of the CAT1 probe indicated that sequence upstream to the core motif may be involved in stabilisation of complex-a, but not complex-b or complex-c. EMSA was, therefore, carried out with CAT2-4, which represented CAT1 plus 8 additional 5' nucleotides. The results showed that, compared to the CAT1 probe (Table 1), a change in the affinity of complex-a relative to complex-b occurred when extracts of differentiating cells (day 6) were incubated with CAT2-4 (Table 2). To test whether this change in the relative levels of complex formation was influenced by either protein or probe concentration, titration of both extract and probe was performed (Fig. 7 and Table 2). These experiments showed that the ratio between complex-a and complex-b was influenced by both factors, complex-a being favoured by higher levels of protein and lower levels of probe, although protein concentration appeared to have a greater influence. One possible explanation for these results is that the two complexes

![Fig. 5. Definition of CAT1 core binding site by mutagenesis. (A) Diagram of mutant oligonucleotides, open arrowheads denote nucleotide substitution. (B) EMSA with CAT1 and CAT1Mutant probes incubated with D7/day 6 parasite-enriched nuclear extract. Lane 1, CAT1; lane 2, CAT1M1; lane 3, CAT1M2; lane 4, CAT1M3. It is not known why doublets were resolved at the position of complex-b and complex-c. (C) EMSA with CAT1 and CAT1-P probes incubated with D7/day 6 parasite-enriched nuclear extract. Lane 1, CAT1; Lane 2, CAT1-P; lane 3 CAT1 and cold competitor CAT1-P. For details of CAT1M and CAT1-P probes, see text.](image)

![Fig. 6. EMSA on CAT1 probe with parasite-enriched nuclear extracts of a differentiation time course. Probe alone (lane 1); probe plus: D7/day 0 (lane 2); D7/day 2 (lane 3); D7/day 4 (lane 4); D7/day 6I (lane 5); D7/day 6II (lane 6); D7/day 6III (lane 7).](image)

![Fig. 7. Titration of CAT2-4 probe and D7/day 6 parasite-enriched nuclear extract. (A) Titration of different amounts of nuclear extract relative to a constant amount of probe (40 fmol): lane 1, 3.8 µg; lane 2, 1.9 µg; lane 3, 760 ng; lane 4 380 ng. (B) Titration of CAT2-4 probe relative to constant extract level (3.8 µg). Lane 1, ×10 probe; lane 2, ×5 probe; lane 3, ×2 probe; lane 4, ×1 (40 fmol) probe. Note that lanes 3 and 4 represent a longer exposure of the autoradiograph in order to visualise the complexes formed under the different conditions simultaneously. A fourth complex of slower mobility than complex-a was also obtained. It was not shown to be specific to the core CAT1 motif.](image)
share a common polypeptide and that complex-a is formed by association of this polypeptide with another factor, or with itself. This possibility was tested by UV-crosslinking experiments. Following separation by preparative EMSA, gel slices representing each of the three complexes associated with CAT1 were excised and subjected to UV irradiation. The DNA-protein complexes were then eluted and analysed by SDS-PAGE. For both complex-a and complex-b the predominant UV-fixed polypeptide DNA complex had an estimated molecular mass of 69 kDa, while for complex-c the mass of the fixed polypeptide plus probe was estimated as 57 kDa (Fig. 8). In addition to the major UV-fixed band at 69 kDa, a second minor band was detected at approximately 80 kDa in the tracks representing complex-a and complex-b of the EMSA. However, proof of specificity by removal of the UV-fixed polypeptides through competition with cold CAT1 oligonucleotide was not achieved. This was due to a failure to UV-fix and resolve these polypeptides without prior separation of the complexes by gel-shift electrophoresis (data not shown).

**DISCUSSION**

Previous work carried out on *T. annulata* stage differentiation postulated that factors controlling expression of merozoite genes were likely to be present in the preceding stage of the life cycle, the macroschizont, and that these factors increased relative to their DNA templates during the differentiation process (Shiels et al., 1994, 1997). This postulation was tested in the present study by screening for parasite nucleotide motifs that bind nuclear polypeptides before and during merozoite formation.

The region of the genome flanking the *TamS1* gene was cloned and sequenced and the general AT content between the protein coding and intergenic regions was found to be very similar. However, stretches of AT-rich sequence >9 bp in length were frequently observed within the 5’ intergenic region (18 times), and six tracts were found to be conserved between *T. parva* and *T. annulata* within 1.1 kb of sequence proximal to the ATG translation initiation codon. Homopolymeric (dA:dT) tracts have also been shown to be overrepresented within *Plasmodium* intergenic regions (Horrocks et al., 1998). As homopolymeric (dA:dT) tracts are involved in gene expression in lower eukaryotes (Hori and Firtel, 1994; Struhl, 1985; Lue et al., 1989), it was proposed that they may modulate transcriptional activity in this parasite. The (dA:dT) tracts within the *TamS1* upstream intergenic region could, therefore, be involved in controlling transcription of this gene.

A major start site of transcription was mapped to 121 bp upstream of the ATG translation initiation. Interestingly, the sequence flanking the mapped site showed a good match to the consensus (PyPyA(+1)N(T/A)PyPy) initiator (Inr) motif defined for a number of eukaryotic promoters (Smale and Baltimore, 1989; Javahery et al., 1994), whereas a consensus TATAAA box sequence could not be identified within close proximity upstream of the transcription initiation site. Inr has been shown to constitute a simple functional promoter and an Inr motif has been shown to be required for expression of the NTP3 gene in *T. gondii* (Nakaar et al., 1998).

The Inr sequence was found to be perfectly conserved between the *T. parva* and *T. annulata* upstream regions, although the start site for *TpmS1* has yet to be defined. Additional short conserved motifs could not be identified as the level of sequence identity between *TamS1* and *TpmS1* was too great. However, within the upstream PCR product obtained for the more distantly related species, *T. sergenti*, three possible conserved motifs were found (see Fig. 1). The first of these motifs, derived from the primer sequence, needs to be

---

**Table 1. Quantification of CAT1 EMSA complex formation during differentiation to the merozoite**

(A) Incorporation of radiolabelled CAT1 probe (c.p.m.) into complex-a and complex-b

<table>
<thead>
<tr>
<th>Complex</th>
<th>Day 0</th>
<th>Day 2</th>
<th>Day 4</th>
<th>Day 6 I</th>
<th>Day 6 II</th>
<th>Day 6 III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex-a</td>
<td>200</td>
<td>367</td>
<td>412</td>
<td>1181</td>
<td>980</td>
<td>1138</td>
</tr>
<tr>
<td>Fold increase day 2:day 6</td>
<td>3.2</td>
<td>2.7</td>
<td>4.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex-b</td>
<td>584</td>
<td>1107</td>
<td>1124</td>
<td>6003</td>
<td>5217</td>
<td>4886</td>
</tr>
<tr>
<td>Fold increase day 2:day 6</td>
<td>5.4</td>
<td>4.7</td>
<td>4.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex-a/b</td>
<td>0.36</td>
<td>0.33</td>
<td>0.36</td>
<td>0.20</td>
<td>0.19</td>
<td>0.23</td>
</tr>
</tbody>
</table>

(B) Analysis of *TamS1* DNA levels (c.p.m.) between day 2 and day 6 time points

<table>
<thead>
<tr>
<th>Hybridised probe</th>
<th>Day 2</th>
<th>Day 6 I</th>
<th>Fold increase</th>
<th>Day 6 II</th>
<th>Fold increase</th>
<th>Day 6 III</th>
<th>Fold increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1827</td>
<td>2911</td>
<td>1.6</td>
<td>2473</td>
<td>1.4</td>
<td>2763</td>
<td>1.5</td>
<td></td>
</tr>
</tbody>
</table>

Note: Computation of fold increase between day 0 and other time points was not feasible due to the low protein concentration in the day 0 extra.
confirmed. The second motif was located 50 bp downstream of the mapped transcription start site of TamS1 and is a 10 bp homopolymeric (dA:dT) motif with 100% identity. The third motif was located at either side of the translational start site and showed 88% identity over a 17 bp stretch. Given the sequence diversity between *T. annulata/parva* and *T. sergenti* (Shiels et al., 1995), it is possible that these three motifs have a functional role. As the second and third motifs are downstream of the mapped transcriptional start site, they could be involved in regulating gene expression at the transcriptional, post-transcriptional or translational level.

Nuclear run-on analysis indicated that TamS1 expression is regulated during differentiation at the transcriptional level, although additional post-transcriptional control may operate. A motif that can be predicted to be involved in regulating transcription was identified by the ability of the CAT1 double-stranded oligonucleotide to display specific mobility shifts when incubated with parasite-enriched nuclear fractions. Under the conditions tested, this motif was the only region in over 500 bp of sequence analysed which demonstrated significant shifts that could be attributed to specific binding of factors associated with parasite-enriched nuclear extracts. Extracts derived from uninfected BL20 failed to generate them (Fig. 4), whereas extracts from infected TaBL20 cells reproduced the D7 shifts (data not shown), which is further evidence that these shifts were associated with parasite fractions.

Selected mutagenesis of the CAT1 probe revealed that a core sequence of TTTGTAGGG is essential for specific binding, while the nucleotides flanking this region probably function to stabilise factor-motif association. We postulate that control of TamS1 transcription during differentiation is mediated, at least in part, by an Inr motif in combination with the upstream (TTTTAGGG) motif positioned at −42. Our data also suggests that the related gene in *T. parva* has a similar promoter structure. The core motif was not found in known eukaryotic promoter motifs, indicating it could be *Theileria*-specific. Homology was found between the 5′ region of CAT1 with the MREe motif (Labbe et al., 1991) and included the sequence CACACA that was shown to be involved in stabilising complex-a formation. Sequence stretches similar to the reverse complement (GTGTGT) are also present in the SV40 core enhancer region (Weihler et al., 1983), and related motifs have been found in upstream regions of *Plasmodium* (Lanzer et al., 1993) and *Toxoplasma* genes (Gross et al., 1996). Whether these motifs play a significant role in regulating apicomplexan transcription is not clear: deletion analysis demonstrated that the TGCTGTGTC motif had no apparent influence on expression of the *BAG1* gene in *T. gondii* (Bohne et al., 1997).

Three complexes were found to bind specifically to the CAT1 probe. All showed elevated levels of formation in extracts derived from differentiating cultures, and the two main complexes were clearly detected in extracts of nondifferentiating cultures. The level of increase in complex-a and complex-b formation was greater than that observed between TamS1 DNA template at days 2 and 6 of differentiation time courses. Thus, it appears, as postulated previously, that factors with the potential to regulate merozoite TamS1 gene expression are observed in the preceding phase of the life cycle and show a quantitative increase relative to their DNA template(s) during the differentiation process.

Compared to the original CAT1 probe, experiments performed with the longer CAT2-4 probe and day 6 extracts generated a higher level of complex-a relative to complex-b formation. Titration experiments showed that the relationship between these two complexes was influenced by the concentration of nuclear extract relative to that of DNA probe, the upper complex being favoured by higher levels of protein relative to DNA template. These results bear similarity to data from other systems, where the ability of transcription factors to form protein-protein interactions results in the formation of monomer, homodimer and heterodimer DNA binding complexes in a concentration-dependent manner (Demczuk et al., 1993; Gabrielsen et al., 1991). The results from the UV-fxation experiments indicated that the major polypeptide associated with complex-a and complex-b has the same estimated molecular mass (Fig. 8). Therefore, the quantitative increase in the level of the DNA binding factor(s) relative to DNA template observed during differentiation could result in formation of a dimer (or higher order complex). It is well known that homodimer and heterodimer formation of DNA binding factors can modulate (Latchman, 1991; Jones, 1990) and elevate (Polly et al., 1996; Feuerstein, 1996) target gene expression. However, as complex-a was also observed in extracts of nondifferentiating cells it cannot be concluded that this complex is associated with the commitment step of differentiation.

The formation of multiple complexes with the core CAT1 DNA binding motif could also be due to independent recognition of the same (or closely related) binding site by different factors. This may be the case for complex-c, as the UV fixation experiment showed a distinct profile for this complex. Complex-c was detected in extracts of cells close to (day 4) or during (day 6) merozoite production, and it is possible that formation of this complex is the result of a qualitative changeover of factors in a subpopulation of cells that are committed to differentiation. Alternatively, as this complex was the least abundant, it may be that a low level of its associated factor(s) in nondifferentiating cells was beyond the limit of detection by EMSA. It has yet to be demonstrated conclusively, therefore, that a qualitative changeover in factors which bind to the CAT1/CAT2-4 motifs occurs at the switch to irreversible merozoite production.

### Table 2. Modulation of CAT2-4 complex formation (c.p.m.) by titration of nuclear extract and DNA template

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Extract</th>
<th>Complex-a</th>
<th>Complex-b</th>
<th>Complex-a/b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.8 μg</td>
<td>1007</td>
<td>1061</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>1.9 μg</td>
<td>423</td>
<td>972</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>760 ng</td>
<td>151</td>
<td>696</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>380ng</td>
<td>30</td>
<td>87</td>
<td>0.34</td>
</tr>
</tbody>
</table>

*Standard binding reactions contained 40 fmol of probe and 2-4 μg of extract. See Materials and Methods for details.*
In addition to the CAT1 motif of the TamSL gene, formation of multiple complexes on a 20 bp DNA motif has been described for the KAHRP gene of P. falciparum (Lanzer et al., 1992). Two of these complexes were associated with expression of KAHRP in a stage-specific manner, while the third showed a quantitative difference in levels between stages. Recognition of core nucleotide upstream motifs by multiple factors may not be unique to Theileria, and could be a component of a basic mechanism that regulates apicomplexan gene expression during stage differentiation.

Thanks to Afshan Fairley and Alan May for help in Table construction and photographic printing. This work was supported by a grant from The Wellcome Trust.

REFERENCES


