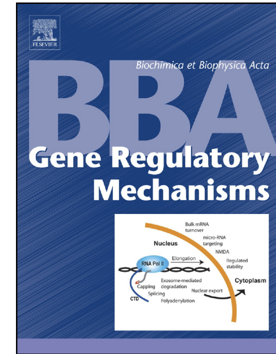
Accepted Manuscript



Characterization of Toxoplasma gondii Spt5 like transcription elongation factor

Pallabi Mitra, Abhijit S. Deshmukh, Rajkumar Gurupwar, Poonam Kashyap

PII: S1874-9399(18)30431-0
DOI: https://doi.org/10.1016/j.bbagrm.2019.01.003
Reference: BBAGRM 94354

To appear in: *BBA - Gene Regulatory Mechanisms*

Received date: 17 October 2018
Revised date: 26 December 2018
Accepted date: 6 January 2019

Please cite this article as: P. Mitra, A.S. Deshmukh, R. Gurupwar, et al., Characterization of Toxoplasma gondii Spt5 like transcription elongation factor, BBA - Gene Regulatory Mechanisms, https://doi.org/10.1016/j.bbagrm.2019.01.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Characterization of *Toxoplasma gondii* Spt5 like transcription elongation factor

Pallabi Mitra[a,*,1], Abhijit S. Deshmukh[b,1], Rajkumar Gurupwar[b] and Poonam Kashyap[b]

[a] *Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad, India*
[b] *National Institute of Animal Biotechnology, Hyderabad, India*

* Corresponding author at: Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad- 500046, Telangana, India

Email: pallabi_scmmjnu@rediffmail.com (Pallabi Mitra)

[1] P.M. and A.S.D. contributed equally to this work

Keywords: Apicomplexa, *Toxoplasma gondii*, Spt5, Spt4, transcription, RNA polymerase II

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Elongation has emerged as a highly regulated step in the multistage process of transcription. Control of gene expression mediated through transcription elongation remains an unexplored area of study in *Toxoplasma gondii* where the demands of complex lifecycle necessitates a regulated transcription program. Here, we elucidate the central role of Spt5 homolog in *T. gondii* mRNA transcription. We demonstrate that TgSpt5 functions in conjunction with a small zinc finger protein TgSpt4. TgSpt5 interacts with TgRpb1, the largest subunit of RNA polymerase II and associates with actively transcribed genes. Enrichment of TgSpt5 towards the 3' end of genes coinciding with P-Ser2 form of RNAPII, a marker of active elongation further underscores its pivotal role in transcription. TgSpt5 undergoes phosphorylation mediated through *Toxoplasma* Cdk9 homolog, TgCrk9, which appears crucial for its function. Inhibition of TgCrk9, which also regulates RNAPII by differential phosphorylation of its C terminal domain, results in loss of TgSpt5 enrichment at 3' sites of the genes and an overall repressive effect on parasite progression. TgSpt5 along with TgSpt4 could successfully complement the loss of function mutations in yeast counterparts emphasizing its functional significance. Together, the results highlight the possible role of TgSpt5 in transcript elongation regulated through phosphorylation by TgCrk9.

Introduction

Toxoplasma gondii, an intracellular protozoan parasite of the phylum Apicomplexa, infects over one billion people worldwide with debilitating consequences for developing foetuses and immunocompromised individuals [1]. It is in fact a common etiology of parasitic abortion in livestock and has serious public health implications due to its potential for zoonotic transmission. *T. gondii* is characterized by its complex lifecycle which involves propagation in different hosts accompanied with considerable changes in gene expression. The lifecycle requires gene expression to be fine-tuned to parasite progression with exquisite temporal regulation in distinct stages and exclusive expression of nearly 5% genes in each developmental stage [2]. This necessitates a robust gene regulation machinery with key role of transcription and mRNA processing. Moreover, targeting transcriptional regulation may potentially intervene parasite progression and provide a foundation to develop anti-parasitic strategies.

Precise control of mRNA levels requires regulation of RNA Polymerase II (RNAPII) mediated transcript synthesis and mRNA processing at multiple steps. It has emerged in

ACCEPTED MANUSCRIPT

recent years that transcription elongation is a highly regulated step in this process [3, 4]. An array of transcription elongation factors has been demonstrated to facilitate this stage [5, 6]. The one transcription accessory protein factor that is conserved across all domains of life is Spt5 or NusG family of proteins [7]. In eukaryotes and archaea, Spt5 functions in complex with another protein factor Spt4. As opposed to Spt5, Spt4 protein conservation is limited to eukaryotic and archaeal lineages [8]. These proteins which were discovered in a yeast genetic screen for mutations that supress the defects caused by insertions of the transposon Ty (Suppressor of Ty, Spt) [9] were subsequently found to regulate transcription elongation [10]. Spt4-Spt5 heterodimer has now emerged as one of the key transcription elongation factors (TEF) which influence the RNAPII activity rendering the elongation complex stable and processive [11].

Interestingly, homolog of Spt4-Spt5 TEF in metazoans were also found to inhibit elongation in apparent contrast to its core function of stimulating RNAPII mediated transcription [12]. The complex is termed as DRB (5, 6-dichloro-1-β-D-ribofuranosylbenzimidazole) -sensitivity inducing factor (DSIF) in organisms where the inhibitory function of Spt4-Spt5 has been demonstrated including mammals and Drosophila [13]. This inhibitory effect manifested as RNAPII pausing at promoter proximal sites, relies on Negative Elongation Factor (NELF), a second multi-subunit complex which is required for DSIF activity [14]. The heterodimer thus regulates transcription elongation both positively as well as negatively in conjunction with other protein partners. Spt5 function appears to be essential across organisms [15-18], whereas Spt4 was found to be dispensable in yeast, however, its importance in eukaryotes is yet to be fully understood [19].

Spt5 is modular in composition consisting of an N-terminal acidic domain, a NusG N-terminal (NGN) domain, multiple KOW (Kyprides, Ouzounis, Woese) domains and a set of short repeats at the C-terminus (C-terminal repeats, CTR). Spt4 is a small zinc finger protein [20] which interact with the NGN domain of Spt5 [20], a domain which is also responsible for interaction with RNAPII. Spt4 does not directly interact with RNAPII, however, it increases the Spt5 binding affinity to RNAPII and the overall stability of the Spt5 protein [21, 22]. The binding of Spt4-Spt5 to RNAPII completely encircles the DNA-RNA hybrid and promotes processivity through allosteric mechanisms [13]. Crystallographic studies have further elaborated on this association revealing that eukaryotic Spt4-Spt5 bound to transcribing RNAPII makes crucial contacts with non-template DNA strand and the newly synthesized transcript exiting the polymerase [23]. The Spt4-Spt5 complex has also been shown to associate with RNA polymerase I, regulating transcript elongation as well as rRNA

ACCEPTED MANUSCRIPT

processing [24, 25]. There is so far no evidence of Spt4-Spt5 involvement in regulation of RNA polymerase III mediated transcription.

Multidomain protein, Spt5, harbours a repetitive motif (consensus= G-S-[R/Q]-T-P) in its C-terminal repeat (CTR) region, similar to the C-terminal domain (CTD) of RNAPII [18]. The sequence of repeats varies across species but a residue that can be phosphorylated (Tyrosine, Serine, Threonine) is typically present. The phosphorylation observed at Spt5 CTR is important for transcript elongation [18, 26-28], and is undertaken by P-TEFb comprising of Cdk9 kinase and its cognate cyclin partner CycT in metazoans or its counterpart Bur1 kinase in yeast [29]. In metazoans, phosphorylated Spt5 is postulated to facilitate release of elongating polymerase II from promoter proximal pause sites [12, 30]. In fact, promoter clearance and promoter proximal pausing are considered critical rate limiting steps in RNAPII mediated transcript elongation where DSIF complex plays an important role along with multiprotein complex NELF and P-TEFb. A downregulation of expression of Cdk9 and Spt5 impair normal transcription by RNAPII [29, 10, 15] and globally alters the distribution of elongating RNAPII [31]. Spt5 also enables co-transcriptional recruitment of factors that modify nascent RNA including 5' capping which may be dependent on the phosphorylation state of its CTR [32, 33].

While transcription elongation factor Spt5 is well studied in bacteria, yeast as well as mammalian systems, it is yet to be characterized in protozoan parasites. In *T. gondii*, where the demands of a complex lifecycle in varied host environments requires a tight transcriptional regulation of mRNA expression, the involvement of Spt5 like proteins and the mechanism remain elusive. Importantly, absence of an important protein complex NELF that essentially assists Spt5 to establish promoter proximal pausing, a rate determining step in RNAPII mediated transcription is very intriguing and points towards a possible presence of non-canonical factors or a unique mechanism in *T. gondii*.

In this study, we identified and reported the presence of homologs of transcription elongation factors Spt5 and Spt4 in the protozoan parasite *T. gondii* and gained insight into how together they impact RNAPII mediated transcription. TgSpt4 and TgSpt5 interact with each other and co-localize in the parasite nucleus. TgSpt5 interacts with TgRpb1, subunit of RNAPII as demonstrated by co-immunoprecipitation reaction. Spt5 associates with RNAPII on actively transcribed genes with the distribution indicating a possible role in transcription elongation. Moreover, Spt5 undergoes phosphorylation which is in agreement with the phosphoproteome data [34]. TgCrk9 is most likely responsible for this phosphorylation event as it phosphorylates Spt5 *in vitro* and *in vivo*. TgCrk9 has previously been demonstrated to be

ACCEPTED MANUSCRIPT

crucial for TgRpb1 phosphorylation at the conserved serine -2 residue in the heptapeptide repeats. The study further highlights TgCrk9 as the key kinase involved in *T. gondii* RNAPII mediated transcription. Targeting of TgCrk9 by inhibitor DRB resulted in reduction of elongating RNAPII marked by serine-2 phosphorylation towards the 3' end of the gene coinciding with Spt5 distribution. TgCrk9 inhibition also affected overall mRNA transcription and stalled parasite progression. Thus, the study elucidated the key role of TgSpt5 in mRNA transcription and its possible regulation by TgCrk9 which was unknown thus far.

2. Materials and methods

2.1. Parasite culture

Toxoplasma gondii RH strain parasites were maintained in primary human foreskin fibroblasts (HFF, ATCC) cells in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum, 10 µg/mL Gentamicin and 2 mM $_{L}$-glutamine at 37°C and 5% CO_2 in a humidified incubator. Freshly harvested parasites were purified by filtration through 3.0 µm filters (Nucleopore).

2.2. Cloning, expression and purification of Spt5, Spt4 and IMC1

Sequences encoding TgSpt5 (32-1169aa, excluding signal sequence), TgSpt5N (32-230aa), TgSpt5M (231-830aa), TgSpt5C (831-1169aa), TgSpt4 (1-146aa) and TgIMC1 (1-609aa) were amplified from *T. gondii* cDNA with specific primers (Supplemental Table 1) and cloned into pET-21a expression vector (Invitrogen) between *Nde*I-*Xho*I sites. Proteins were expressed in *E. coli* as C-terminally His_6-tagged recombinant protein and purified on a nickel-nitrilotriacetic acid-agarose resin column as described previously [35]. TgCrk9, TgCrk7, TgCrk9-T1 & T2, TgCycL, TgCycH and TgRpb1-CTD proteins were used from the previous studies [35, 36].

2.4. Antibodies

Mouse polyclonal antibodies to recombinant TgSpt5, TgSpt4 and TgIMC1 were produced by primary injection with 30 µg of purified protein in Freund's complete adjuvant (Sigma) followed by four boosts of 30 µg each in Freund's incomplete adjuvant (Sigma) at 2-weeks intervals. Rabbit polyclonal antibodies were raised against TgSpt4 and TgIMC1 by

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

immunizing rabbits with 200 µg of recombinant protein mixed with appropriate Freund's adjuvant. Serum was collected after three weeks of immunization.

In addition, we used the following commercial antibodies against RNAPII-CTD-P-Ser5 (ab5131, Abcam), RNAPII-CTD-P-Ser2 (ab5095, Abcam), Histone 3 (ab1791, Abcam), Phospho-(Ser/Thr) antibody (ab117253, Abcam), GAPDH (ab9485, Abcam), Histidine (H1029, Sigma), Alexa flour 488 (mouse, A11001, Thermo Fisher Scientific), Alexa flour 594 (mouse, A11037, Thermo Fisher Scientific), rabbit secondary HRP (31460, Thermo Fisher Scientific) and mouse secondary HRP (31430, Thermo Fisher Scientific).

2.5. Immunofluorescence assay

HFF cells were grown in confluent monolayers on coverslip in 6-well plates and infected with *T. gondii* RH parasites. The infected cells were fixed post 10 to 18 h of infection in 4% paraformaldehyde, followed by permeabilization with 0.25% Triton X-100. The samples were further blocked in 1% BSA and incubated sequentially with primary and secondary antibodies. The following primary antibodies were used at the indicated dilutions: mouse anti-Spt5 (1:500), mouse anti- Spt4 (1:500), rabbit anti-Spt4 (1:500), mouse anti-IMC1 (1:1000) and rabbit anti-IMC1 (1:1000). Alexa-conjugated secondary antibodies (Alexa Fluor-488 or Alexa Fluor-594) were used at a dilution of 1:1000. The coverslips were mounted with Vectashield medium (Vector, Laboratories) with DAPI (4′,6-diamidino-2-phenylindole) on glass slide and images were captured with a Leica Confocal microscope with 100X objective. Images were processed using LAS X software.

2.6. Yeast two-hybrid assay

Yeast two-hybrid assay was performed as per the manufacturer's guidelines (Clontech). Bait and pray plasmids were constructed by cloning full length ORF of TgSpt5 (excluding signal sequence) and TgSpt4 into either pGBKT7 or pGADT7 vectors. TgCrk7-BD and TgCycH-AD constructs were used from the previous studies [35]. Yeast strain Y2HGold was co-transformed with Bait and pray plasmids and plated onto SD plates lacking Leu (leucine) and Trp (tryptophan). The colonies were subsequently transferred to medium lacking Histidine, Leucine and Tryptophan (-His/-Leu/-Trp) and Adenine, Histidine, Leucine, Tryptophan (-Ade/-His/-Leu/-Trp). For spot assays, serial dilutions were prepared (OD_{600}) as 1, 10^{-1}, 10^{-2}, and 10^{-3} and plated on appropriate plates.

2.8. Immunoprecipitation assay

ACCEPTED MANUSCRIPT

The IP from parasite proteins was undertaken using IP kit (Thermo Fisher Scientific) following protocol detailed by the manufacturer. Filter purified parasites were collected, washed in PBS and lysed in PBS with 0.5% NP40, 400mM NaCl and protease inhibitor on ice for 1 h. Total protein extract obtained by centrifugation at 21,000 ×g, 10 min at 4 °C. Protein lysates were immunoprecipitated using appropriate antibody crosslinked to agarose for 1 h at 4 °C. The immunoprecipitated protein complexes on the beads were washed three times with lysis buffer and eluted. Elutes were then boiled in the Leamilli sample buffer at 95°C for 10 min and analysed by Western blotting. In all the IP experiments conducted for protein interaction studies, 10% input samples were analysed. When an RNase treatment step was added to the IP protocol, parasite lysate from the same experiment was treated with five Kunitz units of RNase A (Sigma) prior to immunoprecipitation. After incubating for one hour at room temperature, immunoprecipitations were performed as described above. For Lamda protein phosphatase (λ-PPase) experiment IP samples were treated with λ-PPase for 30 min followed by Western blot analysis using anti-phospho-(Ser/Thr) antibody.

2.9. Chromatin immunoprecipitation assay

ChIP was performed using a protocol described previously [36]. Briefly, HFF infected (intracellular) parasites were cross-linked for 10 min with 1% formaldehyde in PBS and quenched with 125 mM glycine for 5 min at room temperature. The HFF monolayers were washed with PBS, scraped into PBS and then collected by centrifugation. The cell pellets were resuspended in PBS and subsequently passed through 23-guage needle to release intracellular parasites from HFF cells. The parasites were then centrifuged at 4°C for 15 min at 700*x*g, resuspended in lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.1% sodium deoxycholate, 1 mM EDTA, protease inhibitors) and the chromatin was sheared by sonication yielding DNA fragments of 500–1,000 bp. The chromatin was purified by centrifugation at 15,000 ×*g* for 10 min at 4°C; 10% of the chromatin was saved as the input sample and the remaining 90% was used for immunoprecipitation. Immunoprecipitations were performed with respective antibodies overnight at 4°C with rotation, washed, and the IP chromatin was eluted with 1% SDS in TE buffer. Both input and IP chromatin were reverse cross-linked by treatment with proteinase K for 2 h at room temperature and purified by phenol: chloroform method. The purified DNA was analysed by real-time PCR using gene-specific primers.

2.10. Kinase assay

ACCEPTED MANUSCRIPT

Kinase assays were performed as described previously [37]. Briefly, 0.5 μg to 1.0 μg of TgSpt5/TgSpt5N/TgSpt5M/TgSpt5C/TgCrk9/TgCrk7/TgCrk9-T1&T2/TgRpb1-CTD in a 30 μl kinase reaction buffer (50 mM Tris-HCl pH 7.5, 10 mm MgCl2, 0.1 mM EDTA, 200 μM ATP and 5 μCi γ-32P ATP) were assayed at 30 °C for 30 min, supplemented with various combinations of ~1.0 μg of TgCycL/TgCycH proteins. The reactions were stopped by the addition of Laemmli sample buffer followed by boiling at 95°C for 10 min. The reactions were electrophoresed on a 10 to 12% SDS PAGE gels and visualized by phosphorimager.

2.11. RT-qPCR analysis

Total RNAs from DMSO or DRB treated parasites were isolated by TRIzol (Invitrogen). RNAs (DNase I treated) were used to generate cDNA using SuperScipt III reverse transcriptase (Invitrogen). RT-qPCR was performed on the 7500 ABI apparatus using cDNA samples with SYBR green PCR Master Mix. The nascent RNA levels were determined as described [38]. The pre-mRNA levels of *PCNA1*, *IMC1*, *MIC2*, *Bip* and *α-Tubulin* genes were determined using primers targeting specific exon-intron or intron-exon junctions, which allowed quantitative analysis of newly synthesized pre-mRNA. Amplification of basal mRNA of the IMC1 which is expected to be unaltered was used as control. The reactions were performed in triplicates using the following cycle conditions: 95 °C, 15 min followed by 40 cycles of 94 °C, 30 s; 55 °C, 40 s and 68 °C, 50 s. Relative transcript levels were calculated by the ΔΔCT (where CT is threshold cycle) method.

2.12. Proliferation of T. gondii

Parasites inside the parasitophorous vacuoles (PVs) were counted as described [39]. HFF cells grown on coverslips were infected with *T. gondii* at a multiplicity of infection of 5. Following 2 h of infection, parasite infected HFF were subjected to DRB treatment for 24 h. The coverslips were then washed with PBS and fixed using 4% paraformaldehyde and scanned under the microscope. One hundred PVs were randomly selected, and parasite replication was monitored by counting the number of tachyzoites per PV.

2.13. Apoptosis assay

Apoptosis analysis of HFF cells treated with DRB for 24 h was carried out by dual labeling of cells with Annexin-fluorescein and propidium iodide (556547, BD Biosciences) followed by flow cytometry measurements.

ACCEPTED MANUSCRIPT

2.14. Yeast complementation

For complementation assays, yeast strains, Spt5 cold sensitive mutant and Spt4 deletion mutant were used [10]. Full-length *Spt5* and *Spt4* genes of *T. gondii* and *Saccharomyces cerevisiae* were amplified by PCR using specific primers (Supplementary Table 1) and cloned into pYES3/CT yeast expression vector, enabling tryptophan selection and C-terminal histidine tag facilitating protein detection. Cold sensitive strains were transformed with plasmid carrying *S. cerevisiae* and *T. gondii Spt5* and selected at the permissive (30°C) or restrictive temperature (15°C) on SD-TRP plates. Yeast *Spt4* mutant strain was similarly transformed with plasmid carrying *T. gondii* or *S. cerevisiae Spt4* gene or empty vector plated on SD-TRP in presence or absence of 6-azauracil (6AU) selection and incubated at 30°C. Western blot analysis was performed as described previously [40], to check for the expression of the complemented genes in respective transformants using anti-His antibody. ScGAPDH expression levels were used as a control.

2.15. Bioinformatics sequence analysis

Domain schematics of Spt5 protein from various organisms was generated using sequence feature information available on the NCBI and ApiDB (ToxoDB, PlasmoDB) sites and My-Domains image creator PROSITE (ExPASy). Phylogenetic analysis (cladogram) of Spt5 protein sequences from *Toxoplasma gondii*, *Homo sapiens*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Cryptosporidium parvum*, *Theileria annulata* and *Plasmodium falciparum* was generated using Mega 7.0.25 by Neighbour-joining method with a bootstrap value of 1000.

Results

3.1. Identification and in- silico analysis of T. gondii Spt5 and Spt4

Spt5 is the only transcription elongation factor found to be conserved in all three domains of life [7]. Spt5 forms a complex with small zinc finger protein Spt4, termed Spt4-Spt5 in yeast or DSIF in humans [13]. Spt4 is found only in eukaryotes and archaea while it is absent in bacteria [8]. Spt5 appears to be essential in all three domains of life, whereas Spt4 has been found to be dispensable in yeast [13]. In protozoan parasites Spt4 and Spt5 are yet to be identified. We first attempted to identify these protein factors in *T. gondii*. BLASTP search of

ACCEPTED MANUSCRIPT

the Toxoplasma database (http://toxodb.org/toxo/) was undertaken to identify *Toxoplasma* orthologs of Spt5 and Spt4 using amino acid sequences of yeast and human Spt5 and Spt4 as query. The search resulted in hits for both Spt5 and Spt4 (TGME49_233000 and TGME49_261220). TGME49_233000 codes for the Spt5 homolog and was termed *TgSpt5*. The Spt4 homolog is encoded by TGME49_261220 and was called *TgSpt4.* Similarly, putative Spt5 from important members of phylum Apicomplexa including *Plasmodium falciparum*, *Cryptosporidium parvum* and *Theileria annulata*, were also identified. Amino acid sequences of Spt5 and Spt4 of various organisms (Apicomplexa parasites, human, yeast and Arabidopsis) obtained from UniProt database were used for further analysis. Sequence identities of Spt5 of the enlisted members of apicomplexan were compared with those of yeast, human and Arabidopsis counterparts using multiple sequence alignment tool: Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). TgSpt5 shares an overall ~24% identity with its human and plant homologs while it is around 20% when compared with budding yeast Spt5 (Supplementary Fig. 1). Among Apicomplexans, TgSpt5 shares approximately 34%, 30% and 27% identity with *Theileria annulata*, *Cryptosporidium parvum* and *Plasmodium falciparum* respectively (Supplementary Fig. 1). Sequence identity is higher when sequence of NGN domain which is important for interaction with Spt4 and RNAPII is compared across organisms (>41% except with yeast where it is ~26%), highlighting its functional importance (Supplementary Fig. 2). Amino acid sequence identity of TgSpt4 to corresponding human protein is ~38% while with that of yeast and Arabidopsis counterparts is 33% each (Supplementary Fig. 3). Among Apicomplexans, TgSpt4 shares nearly 47% identity with *Plasmodium falciparum* and *Theileria annulata* relatives while it is ~45% when compared to *Cryptosporidium parvum* (Supplementary Fig. 3).

Eukaryotic Spt5 is a multi-domain protein consisting of an N-terminal acidic domain, the NGN (NusG N terminal) domain, multiple KOW (Kyprides, Ouzounis, Woese) domains and C-terminal repeats (CTR). The NGN-domain allows interaction with Spt4 and Rpb1, the largest subunit of RNAPII. The KOW domains which vary in number across species, are important for protein-protein and protein-RNA/DNA interactions. These multiple KOW domains are specific to eukaryotic Spt5 and are key to establishment of elongation complex. A recent crystallographic study has delineated the contribution of different domains of Spt5 towards the formation of transcript elongation complex [23]. The CTR with its serine and threonine residues is a target of phosphorylation which mediates recruitment of factors that are involved in co-transcriptional processing and histone modification [13]. We analysed the sequence of Spt5 in the light of its functionally important protein features and domain

ACCEPTED MANUSCRIPT

organization. We found that TgSpt5 bears all the signature domains including N terminal acidic region, NGN domain, and multiple KOW domains except for a C terminal repeat motif (Fig. 1A). Interestingly, the members of Apicomplexa have been found to lack the C terminal repeat motif found in either yeast or metazoans. The domain architecture of Spt5 is conserved across Apicomplexa with variations in terms of amino acid sequence length, number and position of KOW domains and a variable C terminal domain (Fig. 1A). Apicomplexan Spt5 was observed to have lower number of KOW domains as compared to human Spt5 which has 6 of them. The variations may have significant mechanistic implications which requires to be experimentally validated. The overall architectural conservation supports a central and essential role for this protein.

Spt4 contains two functionally important domains zinc-binding motif and NGN binding domain. The zinc-binding motif contains four invariant Cys residues responsible for zinc binding while the NGN-binding domain is important for interaction with Spt5. The comparison of Spt4 sequences revealed the conserved zinc-binding motif, including four invariant Cys residues, and the NGN-binding domain (Supplementary Fig. 2) which is also reflected in TgSpt4 (Fig 1B).

To understand the evolutionary relationship a phylogenetic tree was generated from Spt5 protein sequences from *Toxoplasma gondii*, *Homo sapiens*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Cryptosporidium parvum*, *Theileria annulata* and *Plasmodium falciparum*. It shows TgSpt5 to be evolutionary closer to *Theileria annulata* Spt5 but has branched away from the model eukaryotes yeast, humans as well as plants (Fig. 1C). This suggest a possible divergence from canonical eukaryotic Spt5 in terms of functional and mechanistic aspects.

3.2. TgSpt5 and TgSpt4 express in the tachyzoite stage of the parasite

Transcriptomic data indicated that TgSpt5 and TgSpt4 are expressed in all the developmental stages including highly replicative tachyzoite stage [41]. To ascertain the endogenous protein expression of TgSpt5 and TgSpt4, we utilized specific antibodies against these proteins. In order to raise the antibodies, full length TgSpt5 and TgSpt4 were bacterially expressed and purified using C- terminal histidine tags (Fig. 2A and 2B). Purified recombinant proteins were used to generate specific mouse and rabbit polyclonal antibodies. Antisera obtained against TgSpt5 recognized the recombinant protein as well as the native protein at the expected molecular weight (~127 kDa) (Fig. 2C). TgSpt4 antisera recognized the

ACCEPTED MANUSCRIPT

recombinant protein and confirmed its native expression, both of which were in accordance with the expected molecular weight (~16 kDa) (Fig.2D). Neither antibody showed any cross reactivity with host cell (HFF-Human Foreskin Fibroblast) proteins.

The specific antibodies were further utilized to determine the localization of TgSpt5 and TgSpt4 proteins in the parasite using immunofluorescence analysis (IFA). *T. gondii* inner membrane complex 1 (TgIMC1), a membrane skeletal protein present immediately beneath the plasma membrane serves as a valuable morphological marker. For this study, we bacterially expressed TgIMC1 (Supplementary Fig. 4A) and raised specific polyclonal antibodies against it. TgIMC1 antisera recognized both recombinant and native protein (~75 kDa) (Supplementary Fig. 4B). Both the proteins were predominantly localized in the parasite nucleus (Fig. 2E and F), suggesting a possible role in transcription in *T. gondii*.

3.3. TgSpt5 forms a heterodimeric complex with TgSpt4

Spt5 interacts with small zinc finger protein Spt4 forming a complex which is an important transcription elongation factor from yeast to mammals [13]. To identify and establish the role of Spt5-Spt4 complex in transcription in the parasite, we first attempted to address whether TgSpt5 indeed interacts with TgSpt4. In this regard, we first performed yeast two-hybrid analysis using a *GAL4BD-TgSpt5* fusion construct which was co-transformed with *GAL4AD-TgSpt4* into yeast Y2HGold strain. Interaction between two given proteins was determined by examining growth on medium lacking histidine. TgSpt5 was able to interact with TgSpt4 in the yeast two-hybrid assay as it could restore growth in medium lacking histidine (Fig. 3A). Interaction between p53 with T-antigen and Lamin with T-antigen represented positive and negative control respectively (Fig. 3A bottom panel). Interaction between parasite proteins, cyclin dependent kinase TgCrk7 with TgCycH has been previously demonstrated [35] and served as an additional control for the experiment. Reciprocal yeast two-hybrid experiment was conducted by swapping bait and prey to rule out false positive interaction between TgSpt5 and TgSpt4 (Supplementary Fig. 5). The yeast two-hybrid experiments strongly indicate that TgSpt5 interact with the TgSpt4.

To further validate the yeast two-hybrid interaction, IFA experiments were undertaken using antibodies raised against TgSpt5 and TgSpt4 in replicating tachyzoite parasites. Dual labeling experiments showed that TgSpt5 and TgSpt4 co-localize in the parasite nucleus (Fig. 3B). This observation supported the interaction of TgSpt5 with TgSpt4 and prompted us to examine whether TgSpt4 could be immunoprecipitated using TgSpt5

ACCEPTED MANUSCRIPT

specific antibodies. TgSpt4 successfully co-immunoprecipitated with TgSpt5 confirming the interaction. Reciprocal Co-IP experiments yielded identical results (Fig. 3C). The tested antibodies showed effective individual immunoprecipitation. Together yeast two-hybrid data along with IFA and Co-IP experiments uphold the existence of a TgSpt5-TgSpt4 heterodimeric complex.

3.4. TgSPT5 associates with actively transcribed genes

Eukaryotic Spt5-Spt4 has been found to closely reflect the distribution of RNAPII on transcribed loci [42]. This association with RNAPII starts early in transcription just down-stream of the transcription start site and continues until the site of termination [43, 44]. In fact, Spt5 NGN domain has been shown to directly interact with RNAPII primarily through its Rpb1 domain [13]. This interaction is further stabilized in presence of Spt4, however, no direct interaction of Spt4 with RNAPII has been observed so far [21].
To test if TgSpt5 also interacts with RNAPII through Rpb1 subunit, we conducted Co-IP experiments using TgRpb1 antibodies followed by western blot analysis using TgSpt5 and TgSpt4 specific sera. The immunoprecipitation experiments were undertaken in presence of RNase A to rule out the possibility that the interaction of Spt5 with Rpb1 is indirect, through mutual association with RNA. TgRpb1 could effectively immunoprecipitate TgSpt5 in both presence and absence of RNAse A. However, TgRpb1 antibodies could not immunoprecipitate TgSpt4 under the same experimental conditions (Fig. 4A). In reciprocal experiment, TgSpt5 could also immunoprecipitate TgRpb1 efficiently, both in presence and absence of RNAse A supporting direct interaction between them.
To further verify the role of TgSpt5 in active transcription, we undertook chromatin immunoprecipitation (ChIP) experiments using TgSpt5 antibodies along with phospho-serine 5 (P-Ser5) and phosho-serine 2 (P-Ser2) antibodies which are hallmarks of RNAPII engaged in active transcription with former associated with initiation while latter with productive elongation. We analysed the distribution of P-Ser2 and P-Ser5 with respect to TgSpt5 at the actively transcribed *Toxoplasma* genes namely, *PCNA1*, *Bip* and *IMC1* (Fig. 4B). Tg*BAG1* gene which is not expressed in tachyzoite stage [45] of the parasite is taken as non-transcribed control. Histone H3 antibody was used as positive control for the ChIP experiment. qPCR amplifications from different regions of the selected genes including 5’UTR, gene body and 3’ UTR were used for ChIP analysis (Fig. 4B). We found that TgSpt5 echoes the distribution of P-Ser2 of RNAPII, both being predominantly enriched at the 3’ end

ACCEPTED MANUSCRIPT

of genes as opposed to P-Ser5 of RNAPII which mostly localized towards the 5’ end of the genes (Fig. 4C-E). Both the phosphorylated forms (P-Ser2 and P-Ser5) of RNAPII as well as TgSpt5 were not observed at the TgBAG1 gene loci (Fig. 4F). Positive control Histone H3 was found to be uniformly associated throughout all the selected genes (Fig. 4C-F). The observed enrichment of TgSpt5 along with serine2 phospho form of RNAPII indicate a possible role of TgSpt5 in RNAPII mediated transcription, potentially at the elongation phase of the process. However, `TgSpt5 chromatin distribution is partly distinct from canonical models where its distribution follows a pattern similar to that of total RNAPII which is in more in agreement with its role as a processivity factor. A 5’ peak of Spt5 which is observed, prior to its enrichment towards the 3’ end of the transcribed gene [46] has been found to be missing for the genes tested in this study and requires to be looked at closely for its biological significance.

3.5. TgSpt5 is phosphorylated by TgCrk9

Differential phosphorylation at important serine residues of the CTD of the Rpb1 subunit of RNAPII is considered a regulatory step in the process of mRNA synthesis [47]. Similar targeting of Spt5 CTD has been observed from yeast to mammals and is considered a critical step correlated with productive elongation [48-50]. The phosphorylation event is not only important for ongoing transcript synthesis but co-transcriptional mRNA processing including capping [33]. This phosphorylation is observed to be undertaken by Cdk9 homolog in yeast as well as mammals where it is a part of the PTEFb complex. Interestingly in *T. gondii*, the phosphoproteome data has provided evidence of phosphorylation at serine and threonine residues in the C terminal region of TgSpt5 [34]. Moreover, a Cdk9 homolog called TgCrk9 has recently been reported to phosphorylate TgRpb1 CTD, at crucial Serine-2 residue with a potential role as a key transcriptional kinase [36]. In the light of the cited observations, we proceeded to investigate whether TgCrk9 is the kinase responsible for TgSpt5 phosphorylation. Recombinant proteins including TgSpt5, TgCrk9, TgCycL, TgCrk7, TgCycH and TgRbp1-CTD to be used in the *in vitro* kinase assay were analysed using coomassie stained gel following SDS PAGE (Fig. 5A). We find that TgCrk9 in presence of its cyclin partner TgCycL phosphorylates TgSpt5 *in vitro* (Fig. 5B). In agreement with our previous published data [36], TgCrk9 was able to phosphorylate TgRpb1-CTD (Fig. 5B). Under the same experimental conditions TgCrk7, another key transcriptional kinase was unable to phosphorylate TgSpt5 while it can efficiently phosphorylate TgRpb1-CTD (Fig.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

5B). The TgCrk9-T1&T2 mutant lacking kinase activity was unable to phosphorylate TgSpt5 (Fig. 5B). TgCrk9-T1&T2 used for this study is a kinase deficient mutant where two Threonine residues located in the T loop domain of TgCrk9 kinase have been mutated (T250A & T257A) [36].

Spt5 has been demonstrated to be phosphorylated at serine/threonine residues harboured in the C-terminal repeat region in both yeast and mammalian systems. In absence of a canonical C-terminal repeat region in TgSpt5, it is important to ascertain the location of the potential phosphorylation sites and whether or not they are located at the C-terminal domain. In order to map the region harbouring the potential phosphorylation sites, we purified three different domains of TgSpt5: N terminal (Spt5N: 32-230aa), middle region (Spt5M: 231-830aa) and C-terminal domain (Spt5C: 831-1169aa) proteins. We observe that only Spt5C protein is getting phosphorylated while the Spt5N and Spt5M proteins fail to do so in presence of activated TgCrk9 kinase (Fig. 5C). In agreement with our experimental observation, Toxoplasma phosphoproteome data [34] reveal that the only phosphorylation marks in TgSpt5 are present at the C terminus (T961, S1001 & T1019). While these sites are not present in a repeat sequence context nonetheless may serve to regulate Spt5 function like the CTR regions of yeast and metazoans.

In vitro kinase assay supported TgCrk9 as the kinase responsible for TgSpt5 phosphorylation. In order to investigate whether TgSpt5 undergoes phosphorylation *in vivo* we immunoprecipitated TgSpt5 and immunoblot analysis was done with antibody specific for phospho serine-threonine residue (Fig. 5D). The band detected by the phospho specific antibody was diminished upon λ protein phosphatase treatment establishing its specificity (Fig. 5D). In previous study, TgCrk9 kinase activity was demonstrated to be selectively and effectively inhibited in presence of DRB and flavopiridol [36]. The inhibition of TgCrk9 was observed at considerably lower concentration for DRB as compared to flavopiridol. Therefore, in this study we have used DRB for inhibiting kinase activity of TgCrk9. To test the effect chemical inhibition of TgCrk9 on TgSpt5 phosphorylation, TgSpt5 was immunoprecipitated from parasites treated with two different concentrations of DRB (10 and 15µM) or with DMSO. A dose dependent decrease in phosphorylation was observed with increasing concentrations of Cdk9 selective inhibitor DRB while it remained unchanged upon treatment with DMSO (Fig. 5E). Together the results suggest Crk9 as the kinase responsible for Spt5 phosphorylation in *T. gondii*.

ACCEPTED MANUSCRIPT

3.6. Abolition of TgSpt5 phosphorylation abrogates RNAPII transcription and represses parasite progression

Promising phosphorylation data prompted us to investigate the possible functional implications of TgSpt5 phosphorylation by TgCrk9 for the parasite. ChIP analysis revealed that treatment of *T. gondii* with TgCrk9 selective inhibitor DRB noticeably alters distribution of TgSpt5 on transcribed gene loci as compared to the untreated control parasite. A marked reduction in TgSpt5 enrichment at 3' end was observed in presence of DRB as opposed to untreated parasites on selected genes including *PCNA1, Bip and IMC1* (Fig. 6A-C). The distribution is again reflective of the P-Ser2 distribution across the genes tested which also shows diminished association at 3' end upon inhibitor treatment as compared to untreated control (Fig. 6A-C). The P-Ser5 distribution with predominant enrichment at 5' end of genes remains unaltered while total RNAPII shows increased association at 5' end following inhibitor treatment (Fig. 6A-C). The ratio of P-Ser5/RNAPII, P-Ser2/RNAPII and Spt5/RNAPII were also calculated and the change in chromatin distribution of P-Ser5, P-Ser2 and Spt5 with respect to RNAPII across transcribed genes were plotted which clearly recapitulated the previously observed distribution at all the selected genes (Fig. 6A-C and Supplementary Fig. 6).

In order to examine the effect of TgCrk9 inhibition affecting Spt5 phosphorylation on RNAPII mediated transcription, effect of DRB treatment of parasite on pre mRNA level was evaluated. Parasites were treated with 15 μM of DRB followed by RNA extraction and pre-mRNA analysis using RT-qPCR at highly expressed genes including *PCNA1, IMC1*, *MIC2*, *Bip* and *α-Tubulin* (Fig. 6D). Pre-mRNA levels were determined using primers targeting exon-intron or intron-exon junctions of the candidate genes by RT-qPCR analysis. This allowed quantitative analysis of newly synthesized pre-mRNA. The values were normalized to total IMC1 mRNA (exonic region, basal level) which is expected to be unaltered. Reduction in pre-mRNA levels was registered for all the genes tested upon TgCrk9 inhibition (Fig. 6E). Effect of DRB treatment on overall parasite progression was also examined. Effect of DRB treatment on *T. gondii* proliferation in HFF cells was assessed at 15 μM concentration. DRB treatment limits parasite progression significantly as compared to untreated parasites (Fig. 6F). To rule out the possibility that the parasite repressive effect observed is due to inhibitor induced apoptosis affecting host cell viability and not due to abrogation of TgCrk9 kinase activity, annexinV staining of host cells were performed upon DRB treatment. Less than 8% host cells were positive for annexin V suggesting that the anti-

ACCEPTED MANUSCRIPT

parasitic effect of DRB is most likely due to selective targeting of TgCrk9 (Fig. 6G). The data is in agreement with previous studies [36] and highlights TgCrk9 as a key transcriptional kinase.

3.7. TgSpt5 and TgSpt4 functionally complement S. cerevisiae counterparts

Our observations so far underscore TgSpt5's contribution to RNAPII mediated transcription with its potential significance for the elongation phase. To establish that Spt5 activity is central to the process of transcription by way of being functionally conserved across organism despite phylogenetic distance, we used a cross-species complementation approach. Spt5 gene is essential across organisms including *S. cerevisiae.* A cold sensitive Spt5 strain in *S. cerevisiae* has been previously utilized to characterize Spt5 function in yeast [10]. We utilized the same strain which revealed the role of yeast Spt5 in transcription elongation to test whether TgSpt5 is functionally preserved. Additionally, *S. cerevisiae* Spt4 deletion mutant was also used in a parallel experiment to examine the functional importance of TgSpt4 through heterologous complementation analysis. ScSpt5 cold sensitive strain was transformed with TgSpt5, ScSpt5 along with empty vector (pYES3/CT) and growth of transformed strain was assayed on SD-Trp plate at permissive (30℃) and non-permissive (15℃) temperatures. Similarly, ScSpt4 deletion strain was transformed with Tg*Spt4*, Sc*Spt4* and empty vector (pYES3/CT) and the growth of respective transformed strain was tested on SD-Trp plates in presence or absence of 6AU at 30℃. Strains lacking Spt4 show sensitivity towards 6AU which reduces nucleotide levels *in vivo* [51] resulting in enhanced pausing and transcriptional arrest of RNAPII [52].

Tg*Spt5* which shares 20% identity with Sc*Spt5* could successfully complement the yeast cold sensitive strain similar to its yeast counterpart at non permissive temperature (Fig. 7A). Tg*Spt4* could also functionally rescue the mutant phenotype of Sc*Spt4* deletion mutant similar to yeast Spt4 in presence of 6AU at 30℃ (Fig. 7B). The empty vector control in each case failed to show any growth. To ascertain TgSpt5 and TgSpt4 expression in respective transformed strains, Western blot analysis using anti-His antibodies was undertaken. Both *Toxoplasma* Spt5 and Spt4 were found to be expressing in the transformed strains (Fig. 7C). Taken together; yeast complementation results establish that TgSpt5 and TgSpt4 have important role in transcription in the parasite.

ACCEPTED MANUSCRIPT

Discussion

Studies have concluded that transcription of genes in parasite is regulated as per the requirement of each developmental stage [2, 53] enabled by a robust gene regulatory system. Also there are reasons to believe that transcription as a process may present attractive parasite intervention possibilities considering compounds which modulate transcription have been proposed to be used for treatment of fungal and viral infections and are selectively toxic to cancerous cells [54, 55]. Therefore, it is imperative to understand and fully characterize this fundamental process of transcription. Here we have identified and characterized an important transcription elongation factor Spt5 which essentially aids RNAPII function in conjunction with protein factor Spt4 in *T. gondii*.

Both TgSpt5 and TgSpt4 express in the tachyzoite stage of the parasite. TgSpt5 co-immunoprecipitate with TgSpt4 and co-localize in the parasite nucleus. The heterodimeric protein factor associates with transcribing polymerase through interaction of TgSpt5 with TgRpb1, the largest subunit of RNAPII. Chromatin immunoprecipitation analysis which demonstrates TgSpt5 distribution on actively transcribed genes where it co-localizes with elongating RNAPII marked by phosphorylation at the serine-2 residue of C-terminal heptapeptide repeats, predominantly enriched towards the 3' end of the genes, indicates a potential role in transcription elongation. This distribution was distinguishable from the distribution of serine-5 phosphorylated form of RNAPII which was found mostly towards the 5' end of the genes tested. We further demonstrate that TgSpt5 is phosphorylated most likely by the Cdk9 homolog of the parasite, TgCrk9. Inhibition of this kinase resulted in diminished enrichment of TgSpt5 towards the 3' end of gene mirroring the distribution of elongating serine-2 phosphorylated RNAPII while the serine-5 phosphorylated form of RNAPII remains unaffected. The distribution of TgSpt5 as observed by ChIP analyses is partly distinct from the canonical pattern where it mostly follows the distribution of total RNAPII [31, 42, 46] and is more in agreement with its role as a processivity factor which is loaded before the elongation commences. The association of Spt5 with RNAPII have been observed to take place right after initiation at the 5' end of the genes [11, 56]. A Spt5 peak canonically found near the transcription start site at the 5' terminus, prior to its enrichment at the 3' end of the transcribed gene was observed to be absent in *T. gondii*. Interestingly, similar chromatin distribution of Spt5 has also been observed in Arabidopsis [57]. Incidentally, Arabidopsis lacks NELF proteins, however, any correlation of the observed distribution with lack of NELF proteins requires to be established experimentally. The

ACCEPTED MANUSCRIPT

observed chromatin distribution of TgSpt5 and its biological significance merits thorough investigation including a genome-wide Spt5 chromatin association analysis.

This study also highlights the role of TgCrk9 as a key transcriptional kinase with both TgSpt5 and TgRpb1 CTD as its primary substrates. Selective chemical inhibition of TgCrk9 showed a restraining effect on parasite progression with minimal effect on host cell viability as examined by apoptosis assay. While more experimental evidence with respect to effect of chemical inhibition on host cell viability requires to be assessed nonetheless targeting TgCrk9 appears to be a promising strategy for parasite control considering its significant impact on the essential process of transcription.

Differential phosphorylation of Rpb1 CTD heptapeptide repeats is one of the key transcription regulatory events. The distribution of the phosphorylated forms of RNAPII on transcribed genes is well documented in metazoans and yeast [5, 58 and 59]. The same has been recently demonstrated in *Toxoplasma gondii* [35, 36] where the kinases accomplish this even in presence of uncanonical cyclin partners (TgCrk9-CyclinL) underscoring the functional importance of the event. Phosphorylation of TgSpt5 may have similar regulatory role in modulation of mRNA transcription. Moreover, enrichment of Spt5 along with serine -2 phosphorylated form of RNAPII at the 3’ end of the genes may have implications for mRNA maturation particularly splicing and 3’ end processing as demonstrated previously [60, 61].

Metazoan RNAPII accumulate at promoters of genes representing a population of transcriptionally engaged polymerase called paused polymerase which is competent to resume transcription. This pausing which is considered a mechanism of fine tuning expression of actively transcribed genes is generally achieved by the action of Spt5 comprising the DSIF complex along with multiprotein complex NELF which together exert this seemingly negative effect on transcription elongation. As noted previously in the text homologs of the constituents of NELF have been found to be absent from *Toxoplasma* genome suggesting either an uncanonical mechanism of transcriptional regulation of genes underway in this parasite or likely involvement of novel, yet uncharacterized protein factors. Dissociation of NELF from DSIF is considered the main step towards productive elongation [42, 62]. Therefore, absence of NELF has posed an important question regarding the whole process of transcription in the parasite including the existence of promoter proximal pausing which is a rate limiting step in transcript elongation. It will also be interesting to draw parallels with budding yeast where absence of pause inducing NELF proteins [63, 64] have been correlated with lack of canonical RNAPII pausing as observed in *Drosophila* to man

ACCEPTED MANUSCRIPT

[65]. On the contrary in fission yeast pause like distribution of RNAPII have been observed and appears to be a Spt4 dependent phenomenon [66]. Promoter proximal pausing still remains to be fully understood mechanistically and biochemically. A recent structural study provides detailed insights into the RNAPII pausing in presence of DSIF and NELF underscoring the importance of both the protein complexes [67]. Also the consequence of absence of NELF upon chromatin distribution of DSIF or Spt5 across different regions of a transcribed genes needs to be fully assessed as noted with respect to the seemingly distinct distribution pattern of TgSpt5 at the 5' end of the genes.

In metazoans, transition of paused RNAPII into its elongating form is facilitated by the kinase activity of P-TEFb [12, 30, 68]. The restraining effect of NELF on RNAPII via DSIF is observed to be relieved upon phosphorylation of Spt5 and RNAPII by P-TEFb [18]. Here we have demonstrated that both TgSpt5 and TgRpb1 CTD are targets of phosphorylation by P-TEFb counterpart in *Toxoplasma*, however, in absence of NELF how these phosphorylation events influence transcription requires further investigation. However, it may be speculated that phosphorylation of TgRpb1 CTD and TgSpt5 may augment productive mRNA synthesis by creating platform for recruitment of co-transcriptional RNA processing and chromatin modifying factors as observed for other systems [26, 69-71]. These activities which are directed by CTDs of Spt5 and Rpb1 triggered by phosphorylation, regulates the association of various protein factors [72].

It is of importance to note here that in metazoans mRNA capping enzyme has also been observed to offset the inhibitory effect of NELF and DSIF on elongating polymerase [73]. In fact, interaction between capping machinery and DSIF has also been reported [74]. In *T. gondii* two such putative genes which have been found to be present (TGME49_305320, TGME49_285520) may play a crucial role in the process of productive transcript elongation in addition to its canonical duties as capping enzyme. A possible role of TgSpt5 in mRNA maturation, particularly splicing may also be evaluated based on the overall decremental trend of pre mRNA transcript levels upon abolition of phosphorylation resulting in deregulation of Spt5 function. This is consistent with previous observation whereby mutation in Spt5 has led to faulty splicing [75-77]. The observations may also be attributed to a combined effect of abrogation of TgCrk9 kinase activity influencing both TgSpt5 and RNAPII CTD. A downregulation of pre mRNA levels that we observe also indicate that TgSpt5 may influence transcription initiation in line with a previous report [77]. Therefore, a versatile role of TgSpt5 in transcription and RNA processing may be envisioned.

ACCEPTED MANUSCRIPT

Spt5 is an essential gene across the three domains of life and same may be anticipated in *T. gondii* which would require conditional ablation strategy for studying the gene. A genome wide transcript profiling of the mutant parasites may be valuable. This study also poses questions with respect to significance of TgSpt4 whose functional relevance as a part of Spt5-Spt4 complex as well as its individual role requires thorough investigation. In conclusion, the study has systematically characterized a key transcription factor, TgSpt5, which functions to assist RNAPII possibly at the elongation phase of transcription. The study has evaluated its role and regulation with respect to RNAPII driven transcription. Its important role in transcription is regulated by phosphorylation mediated through TgCrk9 kinase, which has emerged as a key transcriptional kinase and a potential candidate for targeting in *T. gondii.* Absence of key proteins which canonically functions in association with Spt5 suggests a possible divergent molecular mechanism of transcription in the parasite. Involvement of novel protein factors conferring unique mechanistic properties to the process of transcription may not be ruled out. TgSpt5 may have yet uncharacterized unique functions in the parasite which calls for in-depth investigation.

Acknowledgements

This work is funded by the Dept. of Science and Technology, INSPIRE grant (04/2016/000488) to PM. ASD was supported by DST, INSPIRE grant (LFA13-LSBM-57). We thank Prof. Fred Winston, Harvard Medical School, USA for yeast mutant strains; Dr. Ranjan Sen, CDFD, Hyderabad, India for help with the kinase assays; Ms. Prajna Parimita Kar for her valuable help in bioinformatics analysis of protein sequences. PK acknowledges UGC, India for fellowship.

Author contributions

Conceived and designed the experiments; PM, ASD. Performed the experiments; PM, ASD, RG. Analysed data; ASD, PM. Contributed reagents/materials; PM, ASD, RG, PK. Wrote the paper; PM, ASD.

Competing interests

The authors declare no competing financial interests.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

[1] J.P. Dubey, Toxoplasmosis in Animals and Humans, CRC Press, 2010.

[2] J.R. Radke, M.S. Behnke, A.J. Mackey, J.B. Radke, D.S. Roos, M.W. White, The transcriptome of Toxoplasma gondii, BMC Biol, 3 (2005) 26.

[3] G. Orphanides, D. Reinberg, A unified theory of gene expression, Cell, 108 (2002) 439-451.

[4] R. Perales, D. Bentley, "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions, Mol Cell, 36 (2009) 178-191.

[5] S. Buratowski, Progression through the RNA polymerase II CTD cycle, Mol Cell, 36 (2009) 541-546.

[6] M. Hajheidari, C. Koncz, D. Eick, Emerging roles for RNA polymerase II CTD in Arabidopsis, Trends Plant Sci, 18 (2013) 633-643.

[7] J.K. Harris, S.T. Kelley, G.B. Spiegelman, N.R. Pace, The genetic core of the universal ancestor, Genome Res, 13 (2003) 407-412.

[8] C.P. Ponting, Novel domains and orthologues of eukaryotic transcription elongation factors, Nucleic Acids Res, 30 (2002) 3643-3652.

[9] F. Winston, D.T. Chaleff, B. Valent, G.R. Fink, Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae, Genetics, 107 (1984) 179-197.

[10] G.A. Hartzog, T. Wada, H. Handa, F. Winston, Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae, Genes Dev, 12 (1998) 357-369.

[11] F.W. Martinez-Rucobo, P. Cramer, Structural basis of transcription elongation, Biochim Biophys Acta, 1829 (2013) 9-19.

[12] T. Wada, T. Takagi, Y. Yamaguchi, A. Ferdous, T. Imai, S. Hirose, S. Sugimoto, K. Yano, G.A. Hartzog, F. Winston, S. Buratowski, H. Handa, DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs, Genes Dev, 12 (1998) 343-356.

[13] G.A. Hartzog, J. Fu, The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation, Biochim Biophys Acta, 1829 (2013) 105-115.

[14] Y. Yamaguchi, H. Shibata, H. Handa, Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond, Biochim Biophys Acta, 1829 (2013) 98-104.

[15] M.S. Swanson, E.A. Malone, F. Winston, SPT5, an essential gene important for normal transcription in Saccharomyces cerevisiae, encodes an acidic nuclear protein with a carboxy-terminal repeat, Mol Cell Biol, 11 (1991) 3009-3019.

ACCEPTED MANUSCRIPT

[16] Y. Pei, S. Shuman, Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5, J Biol Chem, 277 (2002) 19639-19648.

[17] B.H. Jennings, S. Shah, Y. Yamaguchi, M. Seki, R.G. Phillips, H. Handa, D. Ish-Horowicz, Locus-specific requirements for Spt5 in transcriptional activation and repression in Drosophila, Curr Biol, 14 (2004) 1680-1684.

[18] T. Yamada, Y. Yamaguchi, N. Inukai, S. Okamoto, T. Mura, H. Handa, P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation, Mol Cell, 21 (2006) 227-237.

[19] E.A. Malone, J.S. Fassler, F. Winston, Molecular and genetic characterization of SPT4, a gene important for transcription initiation in Saccharomyces cerevisiae, Mol Gen Genet, 237 (1993) 449-459.

[20] M. Guo, F. Xu, J. Yamada, T. Egelhofer, Y. Gao, G.A. Hartzog, M. Teng, L. Niu, Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation, Structure, 16 (2008) 1649-1658.

[21] A. Hirtreiter, G.E. Damsma, A.C. Cheung, D. Klose, D. Grohmann, E. Vojnic, A.C. Martin, P. Cramer, F. Werner, Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif, Nucleic Acids Res, 38 (2010) 4040-4051.

[22] B. Ding, D. LeJeune, S. Li, The C-terminal repeat domain of Spt5 plays an important role in suppression of Rad26-independent transcription coupled repair, J Biol Chem, 285 (2010) 5317-5326.

[23] H. Ehara, T. Yokoyama, H. Shigematsu, S. Yokoyama, M. Shirouzu, S.I. Sekine, Structure of the complete elongation complex of RNA polymerase II with basal factors, Science, 357 (2017) 921-924.

[24] D.A. Schneider, S.L. French, Y.N. Osheim, A.O. Bailey, L. Vu, J. Dodd, J.R. Yates, A.L. Beyer, M. Nomura, RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing, Proc Natl Acad Sci U S A, 103 (2006) 12707-12712.

[25] S.J. Anderson, M.L. Sikes, Y. Zhang, S.L. French, S. Salgia, A.L. Beyer, M. Nomura, D.A. Schneider, The transcription elongation factor Spt5 influences transcription by RNA polymerase I positively and negatively, J Biol Chem, 286 (2011) 18816-18824.

[26] Y. Liu, L. Warfield, C. Zhang, J. Luo, J. Allen, W.H. Lang, J. Ranish, K.M. Shokat, S. Hahn, Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex, Mol Cell Biol, 29 (2009) 4852-4863.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[27] Y. Pei, S. Shuman, Characterization of the Schizosaccharomyces pombe Cdk9/Pch1 protein kinase: Spt5 phosphorylation, autophosphorylation, and mutational analysis, J Biol Chem, 278 (2003) 43346-43356.
[28] D. Ivanov, Y.T. Kwak, J. Guo, R.B. Gaynor, Domains in the SPT5 protein that modulate its transcriptional regulatory properties, Mol Cell Biol, 20 (2000) 2970-2983.
[29] M.C. Keogh, V. Podolny, S. Buratowski, Bur1 kinase is required for efficient transcription elongation by RNA polymerase II, Mol Cell Biol, 23 (2003) 7005-7018.
[30] N.F. Marshall, D.H. Price, Purification of P-TEFb, a transcription factor required for the transition into productive elongation, J Biol Chem, 270 (1995) 12335-12338.
[31] A. Shetty, S.P. Kallgren, C. Demel, K.C. Maier, D. Spatt, B.H. Alver, P. Cramer, P.J. Park, F. Winston, Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation, Mol Cell, 66 (2017) 77-88 e75.
[32] H.J. Kim, S.H. Jeong, J.H. Heo, S.J. Jeong, S.T. Kim, H.D. Youn, J.W. Han, H.W. Lee, E.J. Cho, mRNA capping enzyme activity is coupled to an early transcription elongation, Mol Cell Biol, 24 (2004) 6184-6193.
[33] D.L. Lindstrom, S.L. Squazzo, N. Muster, T.A. Burckin, K.C. Wachter, C.A. Emigh, J.A. McCleery, J.R. Yates, 3rd, G.A. Hartzog, Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins, Mol Cell Biol, 23 (2003) 1368-1378.
[34] M. Treeck, J.L. Sanders, J.E. Elias, J.C. Boothroyd, The phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries, Cell Host Microbe, 10 (2011) 410-419.
[35] A.S. Deshmukh, P. Mitra, M. Maruthi, Cdk7 mediates RPB1-driven mRNA synthesis in Toxoplasma gondii, Sci Rep, 6 (2016) 35288.
[36] A.S. Deshmukh, P. Mitra, A. Kolagani, R. Gurupwar, Cdk-related kinase 9 regulates RNA polymerase II mediated transcription in Toxoplasma gondii, Biochim Biophys Acta Gene Regul Mech, 1861 (2018) 572-585.
[37] A.S. Deshmukh, M. Agarwal, P. Mehra, A. Gupta, N. Gupta, C.D. Doerig, S.K. Dhar, Regulation of Plasmodium falciparum Origin Recognition Complex subunit 1 (PfORC1) function through phosphorylation mediated by CDK-like kinase PK5, Mol Microbiol, 98 (2015) 17-33.
[38] T.W. Kelso, K. Baumgart, J. Eickhoff, T. Albert, C. Antrecht, S. Lemcke, B. Klebl, M. Meisterernst, Cyclin-dependent kinase 7 controls mRNA synthesis by affecting stability of

ACCEPTED MANUSCRIPT

preinitiation complexes, leading to altered gene expression, cell cycle progression, and survival of tumor cells, Mol Cell Biol, 34 (2014) 3675-3688.

[39] W. Zhou, J.H. Quan, Y.H. Lee, D.W. Shin, G.H. Cha, Toxoplasma gondii Proliferation Require Down-Regulation of Host Nox4 Expression via Activation of PI3 Kinase/Akt Signaling Pathway, PLoS One, 8 (2013) e66306.

[40] P. Mitra, K. Banu, A.S. Deshmukh, N. Subbarao, S.K. Dhar, Functional dissection of proliferating-cell nuclear antigens (1 and 2) in human malarial parasite Plasmodium falciparum: possible involvement in DNA replication and DNA damage response, Biochem J, 470 (2015) 115-129.

[41] H.M. Fritz, K.R. Buchholz, X. Chen, B. Durbin-Johnson, D.M. Rocke, P.A. Conrad, J.C. Boothroyd, Transcriptomic analysis of toxoplasma development reveals many novel functions and structures specific to sporozoites and oocysts, PLoS One, 7 (2012) e29998.

[42] E.D. Andrulis, E. Guzman, P. Doring, J. Werner, J.T. Lis, High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation, Genes Dev, 14 (2000) 2635-2649.

[43] D.F. Tardiff, K.C. Abruzzi, M. Rosbash, Protein characterization of Saccharomyces cerevisiae RNA polymerase II after in vivo cross-linking, Proc Natl Acad Sci U S A, 104 (2007) 19948-19953.

[44] K. Glover-Cutter, S. Kim, J. Espinosa, D.L. Bentley, RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes, Nat Struct Mol Biol, 15 (2008) 71-78.

[45] W. Bohne, A. Wirsing, U. Gross, Bradyzoite-specific gene expression in Toxoplasma gondii requires minimal genomic elements, Mol Biochem Parasitol, 85 (1997) 89-98.

[46] A. Mayer, M. Lidschreiber, M. Siebert, K. Leike, J. Soding, P. Cramer, Uniform transitions of the general RNA polymerase II transcription complex, Nature structural & molecular biology, 17 (2010) 1272-1278.

[47] M. Heidemann, C. Hintermair, K. Voss, D. Eick, Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription, Biochim Biophys Acta, 1829 (2013) 55-62.

[48] T. Lenasi, B.M. Peterlin, M. Barboric, Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb), J Biol Chem, 286 (2011) 22758-22768.

[49] T. Fujita, I. Piuz, W. Schlegel, The transcription elongation factors NELF, DSIF and P-TEFb control constitutive transcription in a gene-specific manner, FEBS Lett, 583 (2009) 2893-2898.

ACCEPTED MANUSCRIPT

[50] G.T. Booth, P.K. Parua, M. Sanso, R.P. Fisher, J.T. Lis, Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast, Nat Commun, 9 (2018) 543.
[51] F. Exinger, F. Lacroute, 6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae, Current genetics, 22 (1992) 9-11.
[52] S.M. Uptain, C.M. Kane, M.J. Chamberlin, Basic mechanisms of transcript elongation and its regulation, Annual review of biochemistry, 66 (1997) 117-172.
[53] M. Gissot, K. Kim, D. Schaap, J.W. Ajioka, New eukaryotic systematics: a phylogenetic perspective of developmental gene expression in the Apicomplexa, Int J Parasitol, 39 (2009) 145-151.
[54] J.C. Byrd, T.S. Lin, J.T. Dalton, D. Wu, M.A. Phelps, B. Fischer, M. Moran, K.A. Blum, B. Rovin, M. Brooker-McEldowney, S. Broering, L.J. Schaaf, A.J. Johnson, D.M. Lucas, N.A. Heerema, G. Lozanski, D.C. Young, J.R. Suarez, A.D. Colevas, M.R. Grever, Flavopiridol administered using a pharmacologically derived schedule is associated with marked clinical efficacy in refractory, genetically high-risk chronic lymphocytic leukemia, Blood, 109 (2007) 399-404.
[55] J.E. Yeh, P.A. Toniolo, D.A. Frank, Targeting transcription factors: promising new strategies for cancer therapy, Curr Opin Oncol, 25 (2013) 652-658.
[56] D. Grohmann, F. Werner, Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: structure, function and evolution of archaeal RNA polymerase, Research in microbiology, 162 (2011) 10-18.
[57] J. Durr, I.B. Lolas, B.B. Sorensen, V. Schubert, A. Houben, M. Melzer, R. Deutzmann, M. Grasser, K.D. Grasser, The transcript elongation factor SPT4/SPT5 is involved in auxin-related gene expression in Arabidopsis, Nucleic acids research, 42 (2014) 4332-4347.
[58] T. O'Brien, S. Hardin, A. Greenleaf, J.T. Lis, Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation, Nature, 370 (1994) 75-77.
[59] P. Komarnitsky, E.J. Cho, S. Buratowski, Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription, Genes Dev, 14 (2000) 2452-2460.
[60] S.H. Ahn, M. Kim, S. Buratowski, Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing, Molecular cell, 13 (2004) 67-76.

ACCEPTED MANUSCRIPT

[61] L. Davidson, L. Muniz, S. West, 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells, Genes & development, 28 (2014) 342-356.
[62] C.H. Wu, C. Lee, R. Fan, M.J. Smith, Y. Yamaguchi, H. Handa, D.S. Gilmour, Molecular characterization of Drosophila NELF, Nucleic Acids Res, 33 (2005) 1269-1279.
[63] M. Keaveney, K. Struhl, Activator-mediated recruitment of the RNA polymerase II machinery is the predominant mechanism for transcriptional activation in yeast, Mol Cell, 1 (1998) 917-924.
[64] R.D. Alexander, S.A. Innocente, J.D. Barrass, J.D. Beggs, Splicing-dependent RNA polymerase pausing in yeast, Mol Cell, 40 (2010) 582-593.
[65] K. Adelman, J.T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans, Nat Rev Genet, 13 (2012) 720-731.
[66] G.T. Booth, I.X. Wang, V.G. Cheung, J.T. Lis, Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast, Genome Res, 26 (2016) 799-811.
[67] S.M. Vos, L. Farnung, M. Boehning, C. Wigge, A. Linden, H. Urlaub, P. Cramer, Structure of activated transcription complex Pol II-DSIF-PAF-SPT6, Nature, 560 (2018) 607-612.
[68] N.F. Marshall, D.H. Price, Control of formation of two distinct classes of RNA polymerase II elongation complexes, Mol Cell Biol, 12 (1992) 2078-2090.
[69] B.M. Peterlin, D.H. Price, Controlling the elongation phase of transcription with P-TEFb, Mol Cell, 23 (2006) 297-305.
[70] N. Czudnochowski, C.A. Bosken, M. Geyer, Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition, Nat Commun, 3 (2012) 842.
[71] K. Zhou, W.H. Kuo, J. Fillingham, J.F. Greenblatt, Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5, Proc Natl Acad Sci U S A, 106 (2009) 6956-6961.
[72] M. Smolle, J.L. Workman, Transcription-associated histone modifications and cryptic transcription, Biochim Biophys Acta, 1829 (2013) 84-97.
[73] S.S. Mandal, C. Chu, T. Wada, H. Handa, A.J. Shatkin, D. Reinberg, Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II, Proc Natl Acad Sci U S A, 101 (2004) 7572-7577.
[74] M.J. Moore, N.J. Proudfoot, Pre-mRNA processing reaches back to transcription and ahead to translation, Cell, 136 (2009) 688-700.

ACCEPTED MANUSCRIPT

[75] D.L. Lindstrom, S.L. Squazzo, N. Muster, T.A. Burckin, K.C. Wachter, C.A. Emigh, J.A. McCleery, J.R. Yates, 3rd, G.A. Hartzog, Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins, Molecular and cellular biology, 23 (2003) 1368-1378.
[76] G. Diamant, L. Amir-Zilberstein, Y. Yamaguchi, H. Handa, R. Dikstein, DSIF restricts NF-kappaB signaling by coordinating elongation with mRNA processing of negative feedback genes, Cell reports, 2 (2012) 722-731.
[77] G. Diamant, T. Eisenbaum, D. Leshkowitz, R. Dikstein, Analysis of Subcellular RNA Fractions Revealed a Transcription-Independent Effect of Tumor Necrosis Factor Alpha on Splicing, Mediated by Spt5, Molecular and cellular biology, 36 (2016) 1342-1353.

Figure legends

Fig. 1. Domain organization and phylogenetic analysis of TgSpt5. (A) Schematic diagrams of Spt5 from different organisms show conserved N- terminal acidic region, NGN domain, KOW (Kyrpidis, Ouzounis, Woese) domains and C-terminal repeats (CTR). Protein sequences were retrieved from UniProtKB database. The respective UniProt IDs of the proteins analysed have been provided in parenthesis. Tg-*Toxoplasma gondii* (S8GAM4), Pf-*Plasmodium falciparum* (W4IV75), Cp-*Cryptosporidium parvum* (Q7YZ47), Ta-*Theileria annulata* (Q4UAX3), Sc-*Saccharomyces cerevisiae* (P27692), At-*Arabidopsis thaliana* (O80770), Hs-*Homo sapiens* (O00267). (B) Schematic diagram of Toxoplasma Spt4 protein (A0A125YJM8) shows zinc binding motif and NGN domain. (C) Phylogenetic tree generated from the protein sequence data shows TgSpt5 to be evolutionary closer to *Theileria annulata* Spt5 but has branched away from the model eukaryotes yeast, humans as well as plants. Bootstraps values are mentioned at corresponding nodes of the phylogenetic tree.

Fig. 2. Expression and localization of TgSpt5 and TgSpt4**.** (A and B) Coomassie stained gel showing uninduced, IPTG induced and purified recombinant histidine tagged TgSpt5 and TgSpt4 proteins. (C and D) Expression of TgSpt5 and TgSpt4 in tachyzoite stage of parasite was analyzed by Western blotting with anti-TgSpt5 and anti-TgSpt4 polyclonal antibodies. The polyclonal antibodies recognized the recombinant as well as native proteins (TgSpt5 ~127 kDa and TgSpt4 ~16 kDa). Antibodies did not show cross-reactivity with the host HFF

cell proteins. (E and F) IFA images in tachyzoite stage show predominant nuclear localization of TgSpt5 and TgSpt4. Scale bar 5 μm.

Fig. 3. TgSpt5 interacts with TgSpt4. (A) Yeast two-hybrid interaction between *GAL4BD-TgSpt5* and *GAL4AD-TgSpt4*; and control constructs *GAL4BD-TgCrk7* and *GAL4AD-TgCycH* was tested. To test the interaction, *GAL4BD* and *GAL4AD* constructs were cotransformed and yeasts were spotted in serial dilution from 1 to 10^{-3} on nutritionally selective SD/-His/-Leu/-Trp or SD/-Ade/-His/-Leu/-Trp plates. (B) Immunoflorescence analysis using anti-TgSpt5 and anti-TgSpt4 immune sera showed colocalization of the proteins in the parasite nucleus. Nucleus was stained using DAPI. Scale bar 5 μm. (C) IP using anti-TgSpt5 antibody followed by immunoblotting (IB) with anti-TgSpt4 antibody and *vice-versa* were performed to confirm interaction between TgSpt5 and TgSpt4. TgSpt5 could effectively co immunoprecipitate TgSpt4. Identical results were obtained for the reciprocal experiment. Efficient immunoprecipitation was observed for individual TgSpt5 and TgSpt4 antibodies.

Fig. 4. TgSpt5 distribution on RNAPII transcribed genes. (A) IP using anti-TgRpb1 antibody followed by immunoblotting (IB) with anti- TgSpt5 antibody and *vice-versa* were performed to confirm interaction between TgSpt5 and TgRpb1 subunit of RNAPII. IP was conducted using parasite lysate treated with or without RNase A. TgRpb1 could effectively co-immunoprecipitate TgSpt5 in both presence and absence of RNase A, however, it failed to immunoprecipitate TgSpt4 under the same experimental conditions. Identical results were obtained for the reciprocal experiment. Efficient individual immunoprecipitation was observed for TgRpb1and TgSpt5 antibodies. (B) Schematics of *PCNA1, Bip, IMC1* and *BAG1* genes with targeted qPCR amplicons for ChIP analysis is indicated underneath. (C-F) Distribution of P-Ser5, P-Ser2, TgSpt5 and TgH3 on *PCNA1, Bip, IMC1* and *BAG1* genes was determined by ChIP of *T. gondii* parasites. Values were normalized against individual inputs. Error bars, s.e.m. (n=3 independent experiments).

Fig. 5. TgSpt5 is phosphorylated by TgCrk9. (A) Coomassie stained gel showing purified recombinant proteins: TgSpt5, TgCrk9, TgCycL, TgCrk7, TgCycH and TgRbp1-CTD used in the *in vitro* kinase assay. (B) TgCrk9 in presence of TgCycL could phosphorylate TgSpt5 as well as TgRpb1-CTD while TgCrk7 in presence of TgCycH could phosphorylate TgRpb1-CTD however it failed to phosphorylate TgSpt5. TgCrk9 mutant deficient in kinase activity

was also unable to phosphorylate TgSpt5 as well as TgRpb1-CTD. Lower panel shows Coomassie stained gel showing various protein reaction lanes as loading control. (C) Kinase assays were performed using recombinant TgSpt5N, TgSptM, TgSpt5C in the presence of TgCrk9/TgCycL. Activated TgCrk9 kinase could phosphorylate TgSpt5C but not the TgSpt5M and TgSpt5N. Coomassie stained gel showing loading of variants of TgSpt5 proteins in kinase reactions (Lower panel). (D) TgSpt5 phosphorylation was demonstrated using IP with TgSpt5 antibody followed by lambda protein phosphatase (λ-PPase) treatment and Western blot analysis using anti-phospho-(Ser/Thr) antibody. (E) IP of TgRpb1 followed by Western blot analysis for TgSpt5 phosphorylation in *T. gondii* parasites treated with DMSO or 10 and 15 μM DRB shows progressive reduction of TgSpt5 phosphorylation with increasing concentration of DRB. Similar level of TgRpb1 (lower panel D, E) indicates equal amount of parasite proteins used in IP.

Fig. 6. Abolition of TgSpt5 phosphorylation abrogates mRNA transcription. (A-C) Distribution of RNAPII, P-Ser5, P-Ser2 and TgSpt5 on *PCNA1, Bip,* and *IMC1* genes was determined by ChIP of *T. gondii* parasites treated with 15 μM DRB for 2h. For each position of genes, the ratio between P-Ser2 or Spt5 and total RNAPII is shown (A-C). (D) Schematic of *PCNA1*, *IMC1*, *MIC2*, *Bip* and *α-Tubulin* genes with bar underneath indicating the relative position of the region analysed by RT-qPCR. (E) RT-qPCR analysis of pre-mRNA of *PCNA1*, *IMC1*, *MIC2*, *Bip* and *α-Tubulin* genes following 2 h of 15 μM DRB treatment. The pre-mRNA levels of genes were determined using primers targeting specific exon-intron or intron-exon junctions, which allowed quantitative analysis of newly synthesized pre-mRNA. The values were normalized to total IMC1 mRNA (exonic region, basal level) which is expected to be unaltered. (F) Intracellular multiplication of *T. gondii* parasites in HFF cells following DRB treatment. The number of parasites in the parasitophorous vacuoles were counted and converted into percentage. (G) Measurement of apoptosis in HFF cells was undertaken following DRB treatment for 24 h using flow cytometry of annexin V-propidium iodide double stained cells. Mean and SEM values were calculated from three biological replicates. Data sets marked with asterisks are significantly different from control as assessed by t-test: *$P < 0.05$, **$P < 0.01$ or ***$P < 0.001$.

Fig. 7. Tg*Spt5* and Tg*Spt4* functionally complements *S. cerevisiae* counterparts. (A) Yeast strain with chromosomal copy mutation in *Spt5* gene displaying cold sensitive phenotype was transformed with plasmid encoding Tg*Spt5* and Sc*Spt5* along with empty vector

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

(pYES3/CT). The growth was assayed at 30°C and 15°C.(B) Yeast strain with chromosomal copy deletion of *Spt4* gene showing sensitivity to 6AU was transformed with Tg*Spt4*, Sc*Spt4* and empty vector and growth was assayed in presence and absence of 6AU at 30°C. (C) *T. gondii* genes Tg*Spt5* and Tg*Spt4* expressed in the respective transformants as shown by Western blotting using anti-His antibody. *S. cerevisiae* GAPDH was used as loading control (lower panel).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

- Spt (Suppressor of Ty) 5 functions is central to RNAPII mediated transcription in *Toxoplasma gondii*
- TgSpt5 acts together with TgSpt4 forming a hetero-dimeric complex
- TgSpt5 associates with elongating RNAPII on actively transcribed genes predominantly enriching at 3' ends
- TgSpt5 is likely regulated through phosphorylation by TgCrk9 kinase.

ACCEPTED MANUSCRIPT

A

1 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300

TgSpt5
PfSpt5
CpSpt5
TaSpt5
ScSpt5
AtSpt5
HsSpt5

Acidic region
NGN domain
KOW domain
C-terminal repeats

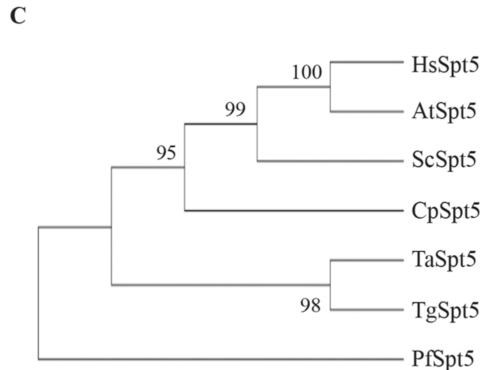


Figure 1

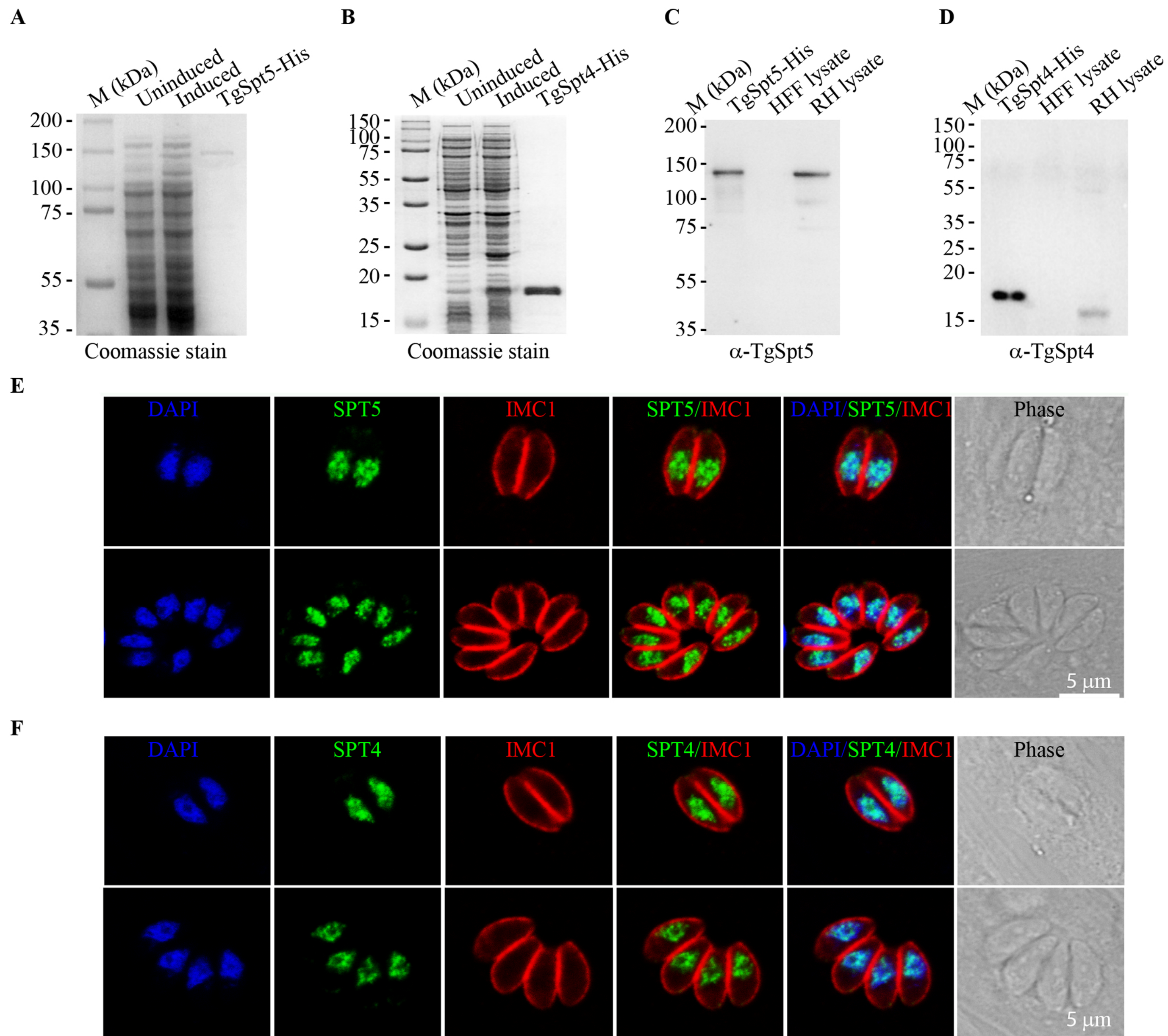
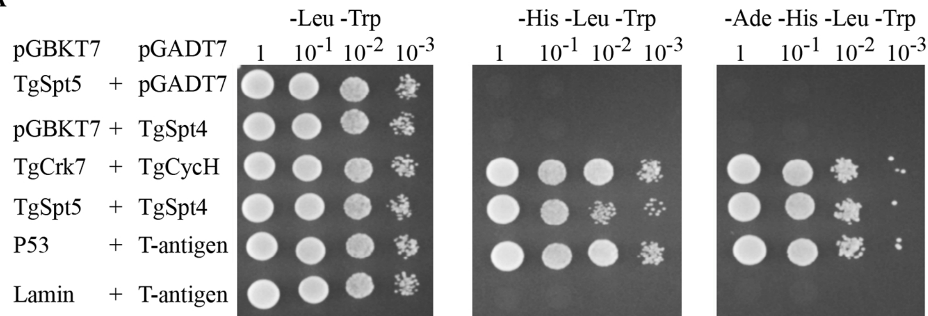
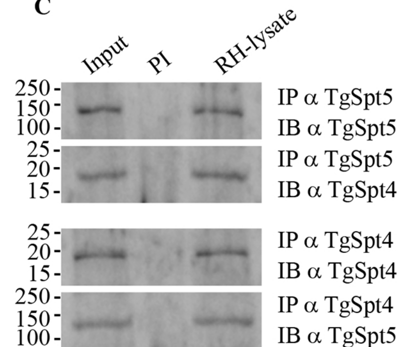


Figure 2





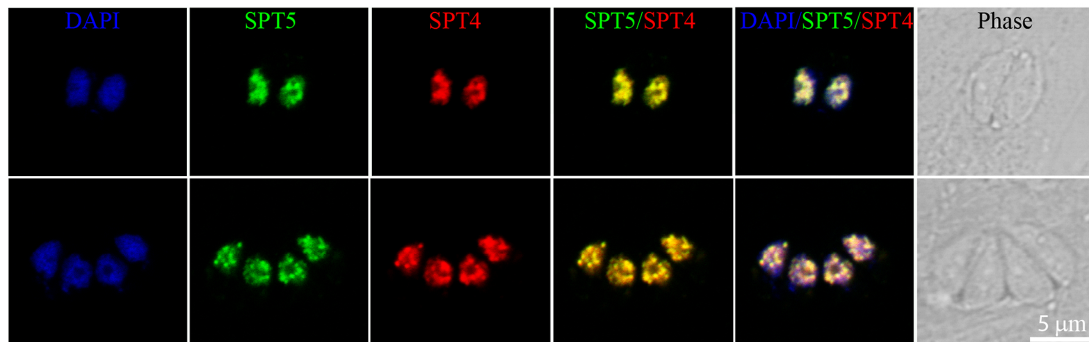


Figure 3

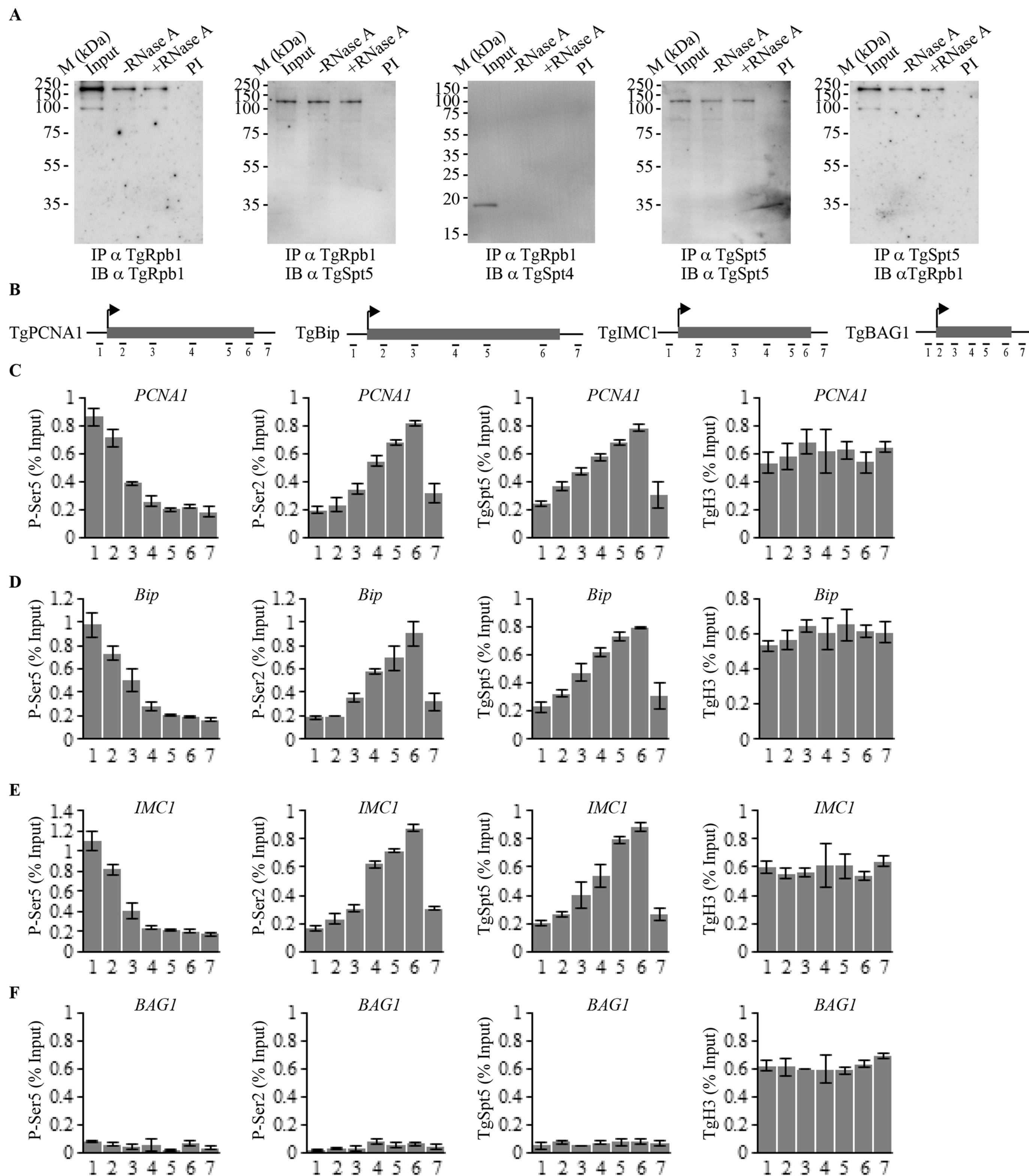
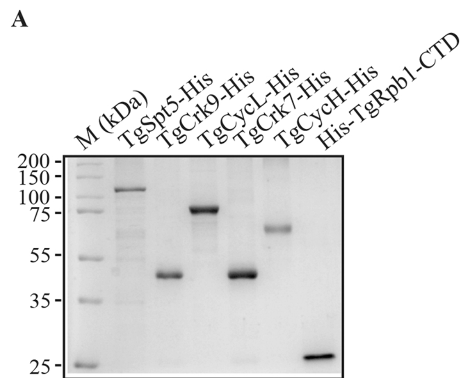
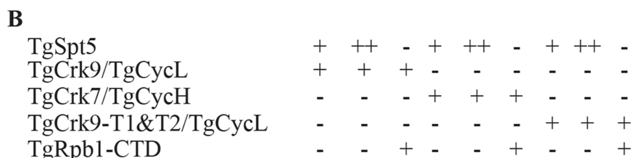
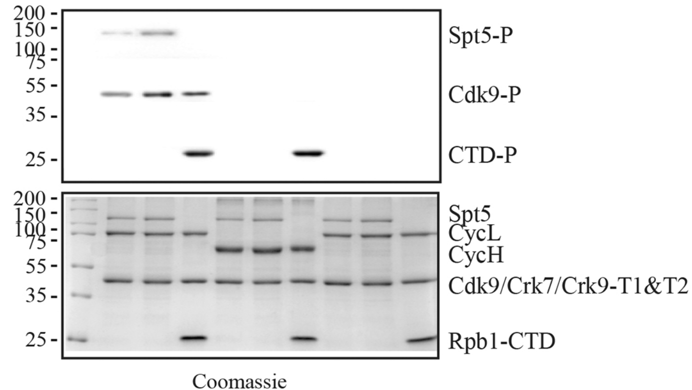
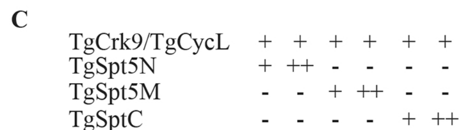


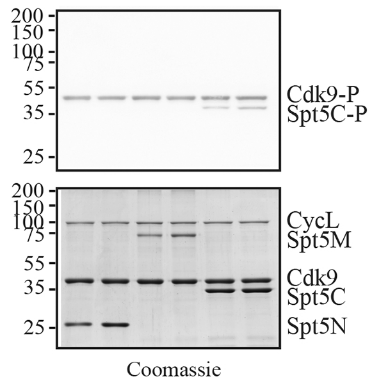
Figure 4

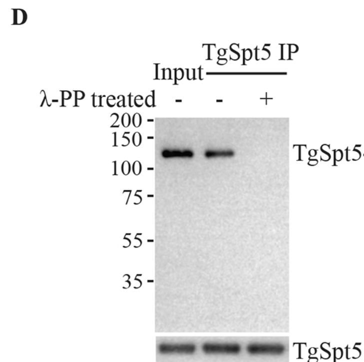












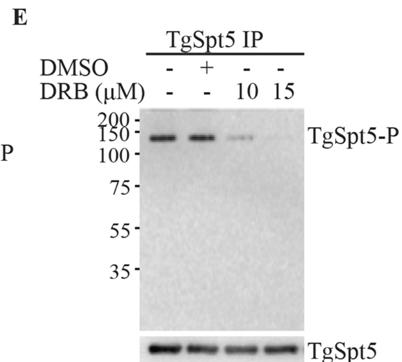


Figure 5

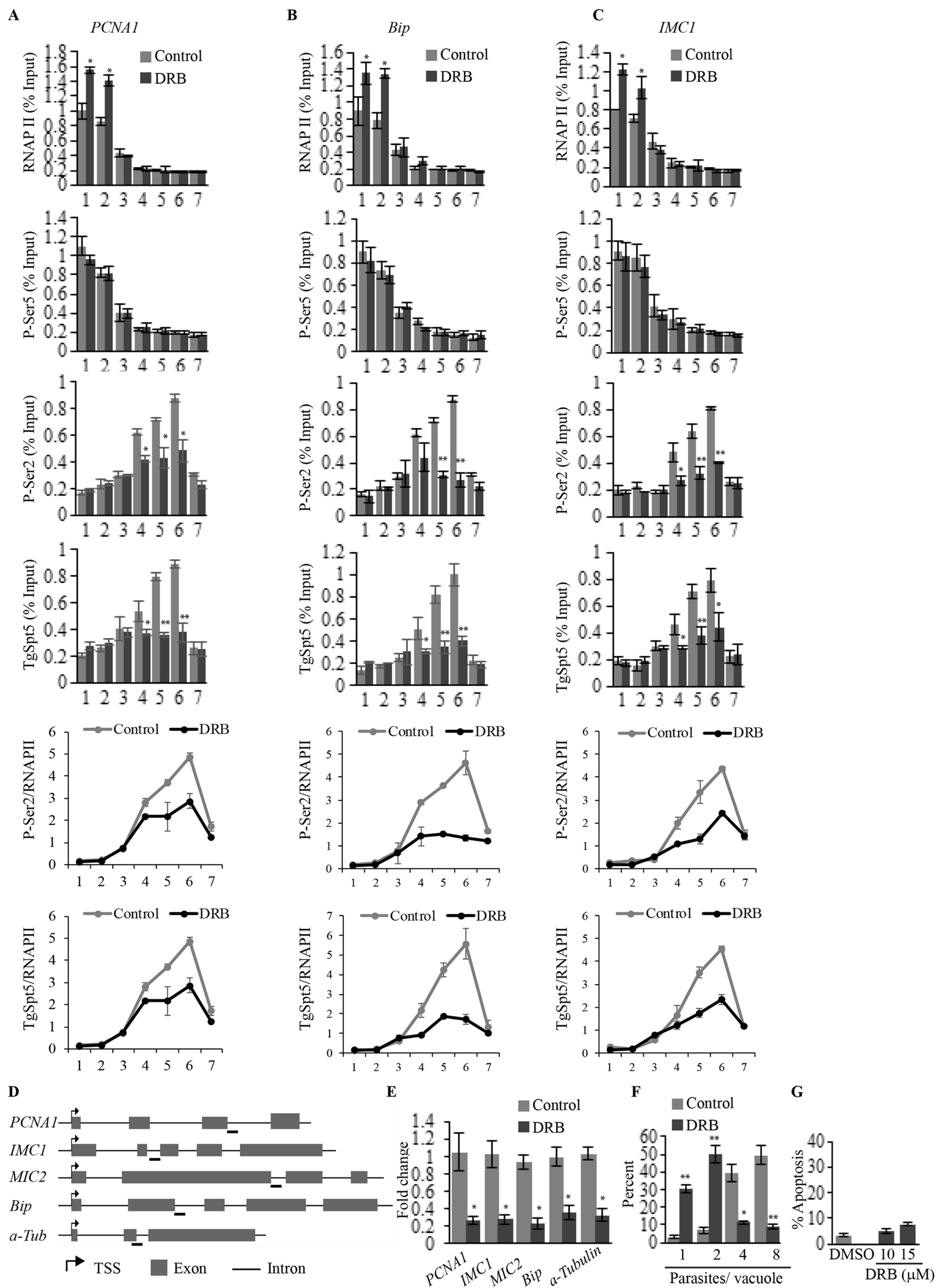


Figure 6

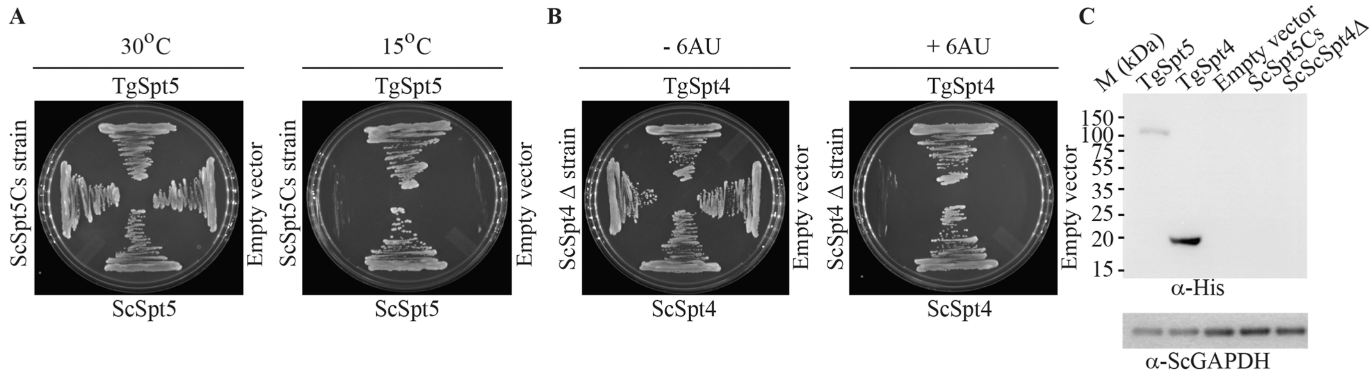


Figure 7