IDENTIFICATION OF CIS-ACTING ELEMENTS CONTROLLING GENE EXPRESSION IN S. neurona

Rajshekhar Y. Gaji
University of Kentucky, rygaji2@uky.edu

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation
https://uknowledge.uky.edu/gradschool_diss/480

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
ABSTRACT OF DISSERTATION

Rajshekhar Y. Gaji

The Graduate School

University of Kentucky

2006
IDENTIFICATION OF CIS-ACTING ELEMENTS
CONTROLLING GENE EXPRESSION IN S. neurona

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Agriculture
at the University of Kentucky

By
Rajshekhar Y. Gaji
Lexington, Kentucky

Director: Dr. Daniel K. Howe, Associate Professor
Lexington, Kentucky
2006
Copyright © Rajshekhar Gaji 2006
ABSTRACT OF DISSERTATION

IDENTIFICATION OF CIS-ACTING ELEMENTS CONTROLLING GENE EXPRESSION IN S. neurona

*Sarcocystis neurona* is an apicomplexan parasite that is a major cause of equine protozoal myeloencephalitis (EPM). During intracellular development of *S. neurona*, many genes are temporally regulated. To better understand gene regulation, it is important to identify and characterize regulatory elements controlling gene expression in *S. neurona*. To perform this study, it was essential to establish transfection system for this parasite. Hence, the 5’ flanking region of the SnSAG1 gene was isolated from a genomic library and used to construct expression plasmids. In transient assays, the reporter molecules β-galactosidase (β-gal) and yellow fluorescent protein (YFP) were expressed by electroporated *S. neurona*, thereby confirming the feasibility of performing molecular genetic experiments in this organism. Stable transformation of *S. neurona* was achieved using a mutant dihydrofolate reductase thymidylate synthase (DHFR-TS) gene of *T. gondii* that confers resistance to pyrimethamine. This selection system was used to create transgenic *S. neurona* that stably express β-gal and YFP. These transgenic clones were shown to be useful for analyzing growth rate of parasites *in-vitro* and for assessing drug sensitivities. To uncover possible sequence elements involved in promoter activity, the 5’ flanking regions of five *S. neurona* genes were subjected to comparative analysis. This revealed the presence of a 7-base conserved motif GCGTCTC. Using a dual luciferase assay system, the SnSAG1 promoter was subjected to functional analysis. The motif GAGACGC located between -136 and -129 upstream of the transcription start site was found to be essential for SnSAG1 expression. This motif functions in an orientation dependent manner and was shown to play a role in binding nuclear proteins of *S. neurona*.

KEYWORDS: *Sarcocystis neurona*, Transfection, Stable Transformation, Gene Regulation, Promoter

Rajshekhar Y. Gaji
July 13, 2006
IDENTIFICATION OF CIS-ACTING ELEMENTS
CONTROLLING GENE EXPRESSION IN S. neurona

By

Rajshekhar Y. Gaji

Daniel K. Howe
Director of Dissertation

Barry P. Fitzgerald
Director of Graduate Studies

July 13, 2006
RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor’s degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name

Date
DISSERTATION

Rajshekhar Y. Gaji

The Graduate School

University of Kentucky

2006
IDENTIFICATION OF CIS-ACTING ELEMENTS
CONTROLLING GENE EXPRESSION IN S. neurona

DISsertation

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Agriculture
at the University of Kentucky

By
Rajshekhar Y. Gaji
Lexington, Kentucky

Director: Dr. Daniel K Howe, Associate Professor
Lexington, Kentucky

2006
Copyright © Rajshekhar Gaji 2006
ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my major advisor, Dr. Dan Howe for his continuous support, guidance, and encouragement. He is a great mentor who inspires his students to become independent investigators.

I extend my appreciation to the members of my graduate advisory committee, Dr. Ernie Bailey, Dr. Brett Spear and Dr. Frank Cook for their counsel, suggestions and assistance.

I am grateful to Dr. Anthony Sinai, Dr. Teri Lear, Dr. Tom Chambers, Dr. David Horohov, Dr. Charlotte Kaetzel and Dr. Brian Stevenson for providing lab space, reagents, equipment and expertise on specific aspects of my research.

I would also like to thank members of EPM lab Dr. Deqing Zhang, Michelle Yeargan, Rebecca Herman, Jessica Hoane, Gagan Gupta, and Carolyn Crowdus for their assistance in completion of my dissertation research.

I gratefully acknowledge Department of Veterinary Science for funding my graduate studies at University of Kentucky.

Finally I am grateful to my parents, my brother and my wife, Bhagyashree, for their encouragement and support during this study.
TABLE OF CONTENTS

Acknowledgements ............................................................................................................. iii
List of Tables ....................................................................................................................... vii
List of Figures ..................................................................................................................... viii
Chapter One: Introduction ................................................................................................... 1
  I. Sarcocystis neurona ................................................................................................. 1
    1. Taxonomy .............................................................................................................. 1
    2. Life cycle ............................................................................................................. 1
    3. Equine protozoal myeloencephalitis (EPM) ......................................................... 2
    4. Ultrastructure of apicomplexan parasites ............................................................. 2
    5. Cell division ......................................................................................................... 4
    6. Molecular biology ............................................................................................... 4

II. Gene regulation by transcription initiation in eukaryotes ............................................. 5
   1. Core promoter and distal regulatory elements ....................................................... 5
   2. Assembly of transcriptional machinery on core promoter and
      transcription initiation ........................................................................................... 7

III. Cis-acting elements in Apicomplexa ........................................................................... 7

IV. Project Rationale ......................................................................................................... 9

V. Figures .......................................................................................................................... 11

Chapter Two: Molecular genetic transfection of the coccidian parasite Sarcocystis neurona ........................................................ ................................................................. 12
  1. Introduction ......................................................................................................... 12
  2. Materials and Methods ....................................................................................... 13
    2.1 Parasite cultures .............................................................................................. 13
    2.2 Construction and screening of a genomic library ........................................... 13
    2.3 Plasmid constructs ......................................................................................... 14
    2.4 Parasite transfections .................................................................................... 16
    2.5 Microscopic analysis of YFP and β-gal expressing parasites ......................... 16
    2.6 Selection of pyrimethamine resistant clones ................................................. 17
    2.7 Southern blot analysis .................................................................................... 17
2.8 Fluorescence activated cell sorting (FACS) and flow cytometry .....18
2.9 β-gal assays ................................................................................................18
2.10 Growth assays using YFP .................................................................19

3. Results .....................................................................................................19
3.1 Isolation of SnSAG1 5’ flanking region..............................................19
3.2 Transient expression of β-gal and YFP in S. neurona..........................19
3.3 Stable transfection of S. neurona..........................................................20
3.4 Transgenic S. neurona expressing β-gal.................................................21
3.5 Transgenic S. neurona expressing YFP ...............................................22
3.6 Use of transgenic parasites in drug screening......................................22

4. Discussion ...............................................................................................23

5. Figures ......................................................................................................26

Chapter Three: GAGACGC is a cis-acting element critical for SnSAG1 expression in Sarcocystis neurona ........................................................................................................33

1. Introduction .............................................................................................33

2. Materials and Methods ..........................................................................34
2.1 Parasite cultures ....................................................................................34
2.2 Isolation and analysis of S. neurona genomic clones ............................34
2.3 Total RNA isolation and 5’ RACE .........................................................35
2.4 Plasmid constructions ............................................................................35
2.5 Parasite transfections ............................................................................38
2.6 Luciferase assays ...................................................................................38
2.7 Nuclear extract preparation .................................................................39
2.8 Electrophoresis mobility shift assays (EMSA) .....................................39

3. Results .....................................................................................................40
3.1 Bioinformatic analysis of 5’ flanking regions of S. neurona genes and identification of transcription start sites of SnSAG1 ..........40
3.2 Functional analysis of SnSAG1 promoter .............................................41
3.3 The T. gondii SAG1 promoter and S. neurona SAG1 promoters are not efficient for heterologous expression ..........................43
3.4 Motif 1 binds S. neurona proteins .........................................................44
LIST OF TABLES

Table 3.1 Primers used in 5’ RACE & designing plasmid construct......................48
Table 3.2 Primers used in making SnSAG1 5’ deletion constructs..........................49
Table 3.3 Primers used in making SnSAG1 5’ mutant constructs.........................50
LIST OF FIGURES

Fig 1.1 Life-cycle of S. neurona ................................................................. 11
Fig 2.1 Schematic representation of expression plasmids............................. 26
Fig 2.2 S. neurona electroporated with expression constructs......................... 27
Fig 2.3 Southern blot analysis................................................................. 28
Fig 2.4 Analysis of β-gal expression in S. neurona ........................................ 29
Fig 2.5 Analysis of S. neurona expressing YFP ............................................ 30
Fig 2.6 Growth curves of SN3(YFP) .......................................................... 31
Fig 2.6 Growth curves of SN3(LacZ) ......................................................... 32
Fig 3.1 Schematic representation of 5’ flanking regions of S. neurona genes ...... 51
Fig 3.2 Examination of reporter activity by SnSAG1 5’ & 3’ flanking regions ....... 52
Fig 3.3A Schematic representation of SnSAG1 promoter ................................ 53
Fig 3.3B SnSAG1 promoter analysis by 5’ deletion analysis .......................... 54
Fig 3.4 SnSAG1 promoter analysis by 5’ deletion analysis ............................. 54
Fig 3.5 Fine mapping of SnSAG1 promoter ............................................... 55
Fig 3.6 SnSAG1 promoter mapping by mutational analysis .......................... 56
Fig 3.7 Heterologous expression ............................................................ 57
Fig 3.8 Electrophoretic mobility shift assay .............................................. 58
Chapter One: Introduction

I. Sarcocystis neurona

1. Taxonomy

*Sarcocystis neurona* is an obligatory intracellular parasite classified under the phylum Apicomplexa. The phylum includes pathogens of both medical and veterinary importance such as *Toxoplasma gondii, Plasmodium spp., Cryptosporidium parvum, Babesia spp and Eimeria spp*. *Sarcocystis neurona* is identified as a major cause of equine protozoal myeloencephalitis (EPM) in horses [1].

2. Life cycle

*Sarcocystis neurona* makes use of dual host system to complete its life cycle wherein opossums (*Didelphis virginiana*) [2] act as definitive hosts and various small mammals such as raccoons (*Procyon lotor*) [3], nine-banded armadillos (*Dasypus novemcinctus*) [4] and striped skunks (*Mephitis mephitis*) [5] can serve as intermediate hosts (Fig 1.1). Opossums excrete sporocysts in feces after ingesting muscles containing *S. neurona* sarcocysts. When the intermediate host ingests sporocysts, sporozoites are released in the intestine where they invade intestinal epithelial cells. After multiple rounds of asexual propagation, merozoites enter skeletal and cardiac muscles and form tissue cysts, called sarcocysts. When an opossum feeds on the muscles of an intermediate host containing sarcocysts, the parasite undergoes sexual propagation in the intestinal epithelial cells leading to the formation of oocysts. Horses become infected after ingesting feed and water contaminated with sporocysts of *S. neurona*. Horses are considered as aberrant hosts since tissue cyst formation has not been commonly seen in these animals [6]. However, a recent study demonstrated tissue cyst formation in an infected foal [7].
3. Equine protozoal myeloencephalitis

Equine protozoal myeloencephalitis (EPM) is a progressively debilitating neurological disease in horses [8]. *Sarcocystis neurona* is considered as a major causative agent of this disease, while *Neospora hughesi* has also been identified as a cause in a small percentage cases of EPM [9]. In horses, *S. neurona* infection appears to be confined only to the central nervous system. Clinical signs of EPM vary depending upon the site of infection; e.g., involvement of the cerebrum may cause behavioral changes or seizures while lesions in the brainstem and spinal cord may cause gait abnormalities. Extensive damage to grey matter that innervates muscles of the limbs can cause weakness and atrophy of the innervated muscles [10]. One of the helpful clinical symptoms in horses with EPM is asymmetric gate deficit accompanied by focal muscle atrophy [6].

4. Ultrastructure of apicomplexan parasites

Apicomplexan parasites are unicellular eukaryotes and possess a distinct set of structures and organelles at the apical end of the cell. *Toxoplasma gondii* is one of the well-studied organisms in this phylum. At the anterior end of the parasite is a special structure called the conoid surrounded by 2 preconoidal rings on top. Two microtubules about 400 nm in length extend from the preconoidal rings and pass through the center of conoid [11]. At the posterior end of the conoid is the polar ring from which 22 microtubules arise and extend two thirds of the length of the parasite body [12]. The plasma membrane forms the outer layer of the cell, and just beneath is the inner membrane complex (IMC), made of flat membranous vesicles [13]. The IMC extends from the anterior to the posterior pole with an interruption at the micropore. This
micropore is hypothesized to play a role in nutrient uptake by endocytosis [14]. The IMC together with the plasma membrane is referred to as the pellicle. The surface of *T. gondii* is covered by an array of glycosyl phosphatidyl inositol (GPI)-anchored proteins designated as surface antigens (SAGs) [15].

The process of invasion of a host cell by apicomplexan parasites is an active process i.e., the parasites make use of their actomyosin motor to gain entry into a host cell [16]. *Toxoplasma gondii* is also equipped with a special set of secretory organelles that are essential for the invasion, intracellular survival, and egress from the host cell. Cigar shaped organelles called micronemes are located at the apical end of the parasite near the conoid. The initial contact with a host cell results in elevation of intracellular Ca\(^{2+}\) that leads to secretion of microneme (MIC) proteins to the parasite surface [17]. The microneme proteins play a role in establishing firm contact between parasite and the host cell, which has proven to be critical for cell invasion [18] [19]. Rhoptries are club shaped organelles located near the apical end, and proteins from this organelle are released as the invasion proceeds. Rhoptry (ROP) proteins are believed to play a role in formation of the parasitophorous vacuole within which the parasite resides during its intracellular life [20]. *Toxoplasma gondii* associates with host cell organelles such as mitochondria and endoplasmic reticulum through ROP proteins [21, 22]. Dense granules are uniformly present through out and proteins (GRA) from this organelle are discharged after the invasion process is almost complete. Dense granule proteins (GRA) modify the parasitophorous vacuole and are believed to help the parasite acquire nutrients from the host cell [23]. A chloroplast-like organelle called the apicoplast is also present in these organisms, the function of which is not clearly understood [24].
_Sarcocystis neurona_ shares most features of the _T. gondii_, although there are some distinct differences. In particular, _S. neurona_ does not have rhoptries in merozoites stage and is not surrounded by a parasitophorous vacuole when present inside a host cell [25] [6].

5. Cell division

After a _S. neurona_ merozoite invades a host cell, it goes through a developmental process termed endopolygeny. This process involves a total of 6 replication cycles that results in the formation of 64 merozoites. However, the first 5 cycles involve only chromosomal duplication and segregation, thus resulting in a 32N cell. Nuclear division and cytokinesis coincide with the completion of the sixth cycle, leading to the production of 64 merozoites [26]. The newly-formed merozoites can either remain in the same host cell or they escape the cell and invade nearby host cells.

6. Molecular biology

Recent studies have begun to provide some insight into the molecular biology of _S. neurona_. About 15000 expressed sequence tags (EST) are now available for _S. neurona_ [27] and an additional 6300 ESTs have been produced for the closely related parasite _S. falcatula_. Comparative analysis of ESTs in apicomplexan parasites revealed that 22 of the top 25 abundantly expressed genes are unique to _S. neurona_ [28]. Of the remaining three, two were similar to surface antigens and the third was similar to ubiquitin like protein [28]. This EST database has facilitated the identification and characterization of _S. neurona_ virulence factors such as surface antigens and secretory proteins. Four major surface antigens have been characterized in _S. neurona_, and these have been designated as SnSAG1, SnSAG2, SnSAG3 and SnSAG4 based on the order of
their discovery [29]. These are GPI-anchored proteins and share homology with surface antigens of *T. gondii*. Two secretory proteins have also been characterized in *S. neurona* and these include SnMIC10 [30] and SnNTP [92].

II. Gene regulation by transcription initiation in eukaryotes

1. *Core promoter and distal regulatory elements*

   In a eukaryotic cell, only a subset of genes is expressed at any given time. This fine regulation of gene expression is achieved through several different mechanisms such as histone modification, DNA methylation, transcription initiation, splicing of mRNA, stability of mRNA, post-translational modification, protein degradation etc. Of these, transcription initiation is a common and an important mode of gene regulation. Transcription of protein coding-genes in higher eukaryotes is carried out by RNA polymerase II, a multisubunit enzyme that catalyzes the synthesis of mRNA from a DNA template.

   The cis-acting elements that play a role in transcription are well defined in metazoans, and these include core promoter elements, enhancers, silencers and insulator sequences. The term core promoter can be defined as the minimal segment of DNA that is sufficient to direct accurate transcription initiation by the RNA polymerase II machinery [31]. The core promoter motifs are located ~35 bp, either upstream or downstream of the transcription start site and include TATA box [32], initiator (Inr) [33], TFIIB recognition element (BRE) [34] and downstream promoter element (DPE) [35] [36]. However, not all genes have the full complement of core promoter elements. A gene may contain all or some or none of these core promoter motifs. For example, analysis of the database of human genes indicated that only 32% of potential promoters
contained the TATA box that is located about 25-30 bases upstream of transcriptional start sites in metazoans [37]. In yeast, however, the TATA box is present at variable locations; i.e., 40-100 bases upstream of the initiator [38]. The consensus sequence for the TATA box is TATAAAG with one or two mismatches [39]. The initiator element is also a core promoter motif and is responsible for transcription initiation from a specific nucleotide. It is found in both TATA box-containing as well as TATA-less promoters.

The consensus sequence for the initiator in mammals is Py Py A (+1) N T/A Py Py with A being the common start nucleotide [40]. However the start nucleotide may not necessarily be adenosine. It can be any nucleotide near the vicinity of the initiator. The BRE is located immediately upstream of the TATA box and binds transcription factor TFIIB in a sequence-specific manner. The consensus sequence for BRE is G/C-G/C-G/A-C-G-C-C, and the 3’C is immediately followed by the 5’ T of the TATA box [34]. The DPE is another core promoter motif that is commonly found in TATA-less promoters [41]. It is located +28 to +32 nucleotides downstream of the transcription start site and has the consensus sequence Py Py A N T/A Py Py in mammals. The DPE functions in association with the Inr to initiate transcription in TATA-less promoters [35], and the spacing between the Inr and DPE is very critical for transcription initiation. Enhancer or silencer elements can be located either near the core promoter, several kb either upstream or downstream from the transcription start site or within introns, and can either activate or repress transcription, respectively. Insulator elements restrict enhancer or silencer elements associated with one gene from inappropriately regulating neighboring genes and also avoid the spread of heterochromatin formation, thus defining the boundary of a gene [42].
2. Assembly of transcriptional machinery on core promoter and transcription initiation

One of the early steps in transcription initiation is the recognition of the TATA box by transcription factor TFIID. TFIID is a multisubunit protein made of TATA binding protein (TBP) and 13 TBP associated factors (TAFs) [31]. TBP is a small 30 kD protein that binds the TATA box in the minor groove in a sequence specific manner [43]. In TATA-less promoters, the initiator together with the DPE can initiate transcription by interacting with TFIID. For example TAF_{II}150 and TAF_{II}250, which are components of TFIID, bind Inr in a sequence specific manner [41]. When TBP binds the TATA box, it forms a saddle around the DNA with its inner face contacting DNA, while the large outer surface is available for interaction with other proteins. The binding of TBP follows bending of DNA by ~80° that provides a larger space for close interaction of transcription factors and RNA polymerase compared to a linear fragment of DNA [44]. This event is followed by sequential binding of general transcription factors (GTFs) – TFIIA, TFIIB, RNA polymerase II –TFIIF complex, TFIIE, TFIIH. This complex is referred to as the preinitiation complex (PIC) [31]. The transcription factor TFIIH has many catalytic units associated with it that includes ATPase, helicase and kinase activity. Using its kinase activity, TFIIH phosphorylates the C terminal domain (CTD) of RNA polymerase II [45]. This process is an essential step for promoter clearance. Most of the TFII factors are released from the PIC and RNA polymerase II begins the process of elongation along the DNA template.

III. Cis-acting elements in Apicomplexa

*T. gondii* is one of the well studied apicomplexan parasite [46]. Sequence analysis of the 5’ flanking region of a major surface antigen gene, *SAG1* indicated the presence of
six tandemly-repeated 27-base motifs about 70 bases upstream of the first transcription start site [47]. Nested 5’ deletion analysis showed that a minimum of two 27-bp repeats are essential for promoter activity above the basal level. The SAG1 promoter when placed in a reverse orientation provided promoter activity similar to the wild type construct thus indicating that it functions in an orientation independent manner [47]. Interestingly, a central heptamer motif A/TGAGACG present in the 27-bp repeats of the SAG1 promoter was also found in promoters of genes coding for dense granule proteins (GRA1, GRA2, GRA5 and GRA6) and tubulin (Tub) [47-51]. The motif was found in both orientations (WGAGACG or CGTCTCW), in different copy number, and at variable locations in the 5’ flanking region of these promoters. Deletion experiments showed that regions containing only one motif (the most proximal to transcription start site) are sufficient for basal promoter activity of GRA2, and mutation of this motif abolished promoter activity [52]. However, some genes of T. gondii such as NTPase and DHFR-TS also contain this motif in the 5’ flanking region, but mapping studies indicated that this motif did not play a role in transcription initiation [53, 54].

BAG1 is a stage specific small heat shock protein in bradyzoites of T. gondii [55]. Transient transfection experiments indicated that a 324-bp 5’ flanking region is sufficient to confer stage-specific expression [56]. T. gondii has two genes of enolase (ENO1 and ENO2) that are stage specific; i.e., ENO1 is expressed in bradyzoites and ENO2 is expressed in the tachyzoite stage [57, 58]. The promoter regions of both enolase genes did not show any homology with each other, and neither contains the WGAGACG motif. A region –1245 to –625 acts as a repressor of ENO1 in tachyzoites [59]. The ENO2 promoter region contains 2 repression regions (–1929 to –1067 and –456 to –222) that
block expression in bradyzoites. Electrophoresis mobility shift assays demonstrated the presence of DNA-binding proteins in tachyzoite and bradyzoite nuclear lysates that bound to stress response elements (STRE), heat shock-like elements (HSE) and other cis-regulatory elements in the upstream regulatory regions of ENO1 and ENO2 [59].

In *Plasmodium* spp, the identification of cis-acting elements and the corresponding trans-acting factors has been a difficult task, as intergenic regions are AT rich and regions resembling classical TATA motifs are abundant upstream of most genes [60]. Deletion analysis of the glycophorin binding protein 130 (*GBP130*) promoter in *P. falciparum* indicated that a 37-bp region located 507 bases upstream of the transcription start site is essential for efficient promoter activity, and mutational analysis showed that the 5-bp element GTATT present in this sequence is a critical cis-acting element [61]. CDP-diacylglycerol synthase (CDS) is an enzyme involved in phospholipid metabolism in *P. falciparum*. Promoter analysis of CDS indicated that a 24-bp region, 1605 bases upstream of transcription start site is required for efficient transcriptional activity [62]. Bioinformatic and functional analysis of a *P. falciparum* heat shock protein (hsp) gene family indicated that a palindromic G box (A/GNGGGGC/A) is involved in transcription [63]. A recent study demonstrated that a consensus TATA box is present 81 bases upstream of the transcription start site and 186 bases upstream of the transcription start site in the promoter regions of knob-associated histidine rich-protein (*kahrp*) & *GBP-130* respectively in *P. falciparum* [64]. The study also showed that native *P. falciparum* TATA binding protein (*PfTBP*) interacts with the TATA boxes in these gene promoters.

IV. Project rationale
Intracellular development of *S. neuronae* is a complex process during which many genes are temporally regulated [30]. For example SnMIC10, a microneme protein is not expressed until terminal stages of endopolygeny [30]. It is likely that there are regulatory sequences upstream of these genes that control expression of proteins at different stages of development. To understand gene regulation it is important to identify and characterize regulatory elements controlling gene expression in *S. neuronae*. As a first step towards this objective molecular genetic transfection system will be established to study this parasite. Using transient transfection system regulatory elements controlling gene expression in *S. neuronae* will be identified.
Figure 1.1. Life-cycle of *S. neurona*. Opossums act as definitive host of *S. neurona* and excrete sporocysts in feces. Small mammals such as armadillos, raccoons and skunks serve as intermediate hosts. When the intermediate hosts ingest sporocysts of *S. neurona*, tissue-cysts referred to as sarcocysts are formed in the muscle. Horses become infected after ingesting feed and water contaminated with sporocysts of *S. neurona*. 
Chapter Two: Molecular genetic transfection of the coccidian parasite *Sarcocystis neurona*

Introduction

*Sarcocystis neurona* is an obligate intracellular parasite and the major causative agent of equine protozoal myeloencephalitis (EPM) [6]. *S. neurona* is classified in the phylum Apicomplexa, which includes pathogens of both medical and veterinary importance such as *Plasmodium spp.*, *Toxoplasma gondii*, *Cryptosporidium spp.*, *Neospora spp.*. *S. neurona* utilizes the opossum as its definitive host [2], and various small mammals can serve as intermediate hosts [4, 5, 65, 66]. Horses are considered aberrant hosts since tissue cyst formation has not been observed in these animals. But a recent study indicates that horses can be natural intermediate hosts [7]. In infected horses, *S. neurona* can parasitize all regions of the central nervous system, and clinical signs depend on the site affected.

Recent studies have begun to provide some insight into the molecular biology of *S. neurona*. Over 15,000 expressed sequence tags (EST) are now available for *S. neurona* [27]. This sequence database has facilitated identification and characterization of several parasite virulence factors such as surface antigens [29] and secretory proteins like SnMIC10 [30] and SnNTP1 [92]. Although *S. neurona* exhibits many features of the Apicomplexa, it possesses several distinct qualities such as a lack of rhoptries in the merozoite stage, the absence of a parasitophorous vacuole, and cell division by endopolygeny [6]. During endopolygeny, the parasite undergoes 6 replication cycles that results in formation 64 daughter merozoites. This prolonged developmental process has
made *S. neurona* a tractable model for cell development and division by apicomplexan organisms [26].

Molecular genetic capabilities are invaluable for investigating the biology of organisms, including parasites. Genetic transfection of an apicomplexan parasite was first accomplished in *T. gondii* [67]. Subsequently, transfection has been achieved in *Plasmodium* [68] [69], *Neospora caninum* [70] and *Eimeria tenella* [71]. The development of reverse genetic tools in these unicellular pathogens has aided in a variety of studies examining various biological processes such as gene regulation [47], host cell invasion [72] and pathogenesis [73]. The objective of this study was to develop molecular genetic tools for *S. neurona*, thereby allowing more in-depth examination of the biology of this parasite. Here we describe transient transfection and a stable transformation system for *S. neurona*.

Materials and methods

2.1 Parasite cultures

*S. neurona* strain SN3 was propagated as merozoites by serial passage in bovine turbinate (BT) cells that were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum, 2 mM sodium pyruvate, Pen/Strep Fungizone (Bio Whittaker, Inc). Freshly lysed extracellular parasites were passed once through 20 G, 23 G and 25 G needles, and purified from the cell debris with 3.0 µm membrane filters, as described previously [30].

2.2 Construction and screening of a genomic library

*S. neurona* genomic DNA was isolated from a freshly-lysed culture of merozoites purified as described above. Merozoites were suspended in preheated lysis buffer
containing 100 mM EDTA (pH 8.0), 10 mM Tris HCl (pH 8.0), 1% Sodium lauryl sarcosyl and 2mg/ml of Proteinase-K and incubated overnight at 50°C. Genomic DNA was extracted three times with phenol:chloroform:isoamyl alcohol, twice with chloroform:isoamyl alcohol, precipitated with ethanol and resuspended in 10 mM Tris (pH 8.0), 1 mM EDTA, 1 mg/ml RNase A. Cosmid library construction was accomplished according to protocols from the manufacturer (Stratagene). Briefly, the chromosomal DNA was partially digested with Sau3AI and ligated into the BamHI site of Supercos-1. The ligation reaction was packaged into phage heads using Gigapack® III XL packaging extract (Stratagene) and transduced into XL1-Blue MR strain of E.coli. To obtain a SnSAG1 genomic clone, colony lifts of the cosmid library were screened with a digoxigenin labeled SnSAG1 DNA probe using DIG High Prime Labeling and Detection kit II (Roche Molecular Biochemicals). The 5’ flanking region of the SnSAG1 genomic clone was sequenced using ABI Prism™ Big Dye Terminator Cycle sequencing reaction mix (PE applied biosystems). Reactions were purified with AutoSeq™ G-50 columns and analyzed on ABI 310 genetic analyzer (PE Applied Biosystems). Sequences were manually edited and analyzed using Vector NTI Advance™ 9.0 (Informax™, Invitrogen).

2.3 Plasmid constructs

Plasmid vectors for expression of yellow fluorescent protein (YFP), β-galactosidase (β-gal) and dihydrofolate reductase thymidylate synthase (DHFR-TS) are shown in Fig 1. The TubYFP-YFP/sagCAT plasmid [74] was digested with HindIII and BglII to remove the 4.1 kb fragment containing the CAT expression cassette and tubulin promoter of T.gondii. A 788 bp fragment of the SnSAG1 5’ flanking region was
amplified using a forward primer that created a HindIII site (primer, 5’-
ACTGAAGCTTACAGTGAGGCAGATATGGC) and a reverse primer that added a
BglII site (primer, 5’- ACTGACAGATCT GCGAGAATGACAGATGGACG). The
resulting PCR product was digested with HindIII and BglII and cloned into the
HindIII/BglII-digested TubYFP-YFP/sagCAT to make the plasmid pSnSAG1/YFP-YFP.

The GRAI promoter in the previously described vector pGRA1/LacZ [70] was
excised with HindIII and NsiI, and replaced with the 5’ flanking region of SnSAG1
amplified using primers that contained HindIII and NsiI sites (forward primer, 5-
ACTGAAGCT TAAACAGTGAGGCAGATATGGC; reverse primer, 5’-
ACTGACCATATGCTGCGAGAAT GACAGATGGACG) to make pSnSAG1/LacZ.

The DHFR-TS ORF was amplified using primers that created an NsiI site at the
start codon and a PacI site at the stop codon (forward primer-5’
ACTGACATGCATAAACC GGTGTGTCTGG; reverse primer-
5’ACTGA CTTAATTAACTAGACGAGCCATCTCCATCTG). The amplification product
was digested with NsiI and PacI, and cloned in place of the LacZ ORF of pSnSAG1/LacZ
to make pSnSAG1/DHFR-TS.

The SnSAG1 5’ flanking region, DHFR-TS ORF and TgSAG1 3’ UTR was
amplified from pSnSAG1/DHFR-TS with primers that created NotI and BamHI sites (forward primer, 5’- ACTGACGCGCCGCAACACGTGAGGCAGATATGGC; reverse primer, 5’-ACTGACGGATCCTGTCGACGGCGATCAACATC) and cloned into NotI, BamHI-digested pSnSAG1/YFP-YFP to make pSnSAG1/YFP-YFP-pSnSAG1/DHFR-
TS.
To generate pSnSAG1/LacZ-pSnSAG1/DHFR-TS, the SnSAG1 promoter, Lac-Z ORF and TgSAG1 3’ UTR was excised from pSnSAG1/LacZ with HindIII and BamHI and inserted into HindIII and BamHI digested pSnSAG1/YFP-YFP-pSnSAG1/DHFR-TS replacing the YFP expression cassette. All PCR reactions were performed using Klen Taq polymerase (Sigma) to reduce mutations introduced by the polymerase.

2.4 Parasite transfections

For transfection, freshly-lysed *S. neurona* merozoites were washed with HHE and resuspended in Cytomix buffer supplemented with 2mM ATP and 5mM glutathione [67]. About 40 µg of plasmid DNA, purified by QIAfilter™ Maxi kit (Qiagen), was mixed with 2x10^7 parasites in a 2-mm gap cuvette (BTX) and electroporated at 2 kV, 129Ω using a BTX ECM 600 electroporation system. Transfected parasites were allowed to recover for 10 min and transferred to a confluent monolayer of BT cells.

2.5 Microscopic analysis of YFP and β-gal expressing merozoites

Monolayers of BT cells were prepared in 4 well Lab-Tek chamber II slides (Nalgene Nunc International) coated with poly-L-lysine or on 24 well plates (Falcon®). Approximately 1x10^5 transfected parasites were added to each chamber or well and the slides were incubated at 37°C in 5% CO₂ for 48 hours. Infected monolayers were washed thrice with PBS and fixed with 2.5% formalin, 0.25% gluteraldehyde in PBS for 15 min at 4°C. For YFP expressing parasites, the slides were rinsed with PBS and mounted in Vectashield mounting medium with DAPI (Vector laboratories Inc.) and examined with a Zeiss axioskope equipped for phase contrast and epifluorescence microscopy. To visualize β-gal expressing parasites, the cells were permeabilized with 0.2% TX-100/PBS for 30 min and blocked with 10% NGS/PBS for 30 min. Slides were
then incubated with 1:500 diluted MAb against *E. coli* β-gal (Developmental Studies Hybridoma Bank) for 1 hour at room temperature. Slides were rinsed and incubated with FITC-conjugated anti-mouse antibody (1:200). The slides were then mounted and examined as mentioned previously.

2.6 Selection of pyrimethamine resistant clones

About 1x10^7 transfected parasites were added to each well in a 6 well plate containing BT cell monolayers. After 24 hours, the medium was changed to RPMI/3% FBS containing 10 µM pyrimethamine. The cultures were maintained till the monolayer was 90% lysed. Parasites were passed through needles and inoculated into a T25 flask for a second round of selection with 10 µM pyrimethamine. After disruption of the monolayer, parasites were cloned by limiting dilution as described previously [75]. Briefly, freshly lysed parasites were passed through needles, filtered, counted and diluted to 4x10^2/ml in RPMI/10% FBS. A 10 µl aliquot of the suspension (4 merozoites) was added to each well of a 96 well plate. Wells containing a single plaque were identified after 7-10 days, after which the medium was changed every 5 days. Once the parasites lysed the monolayer, the content of the well was transferred to a single well of a 24 well plate for expansion of the parasite culture. After lysis of the monolayer in the 24-well plate, the merozoites were transferred to a T25 flask.

2.7 Southern blot analysis

Genomic DNA was isolated from transgenic clones as described in section 2.2. Approximately 5 µg of DNA was digested with restriction enzymes, electrophoresed on agarose gel, and transferred by Southern blot [76] to a nylon membrane (Hybond™-N, Amersham biosciences). The membrane was hybridized with digoxigenin-labeled DNA
probe generated from either a portion of the ColE1 origin of replication and ampicillin resistance gene) of the pBluescript SK+ or from the SnSAG1 5’ flanking region.

2.8 Fluorescence Activated Cell Sorting (FACS) and flow cytometry

Freshly lysed parasites were passed through needles, filtered and resuspended in PBS. The parasites were sorted at room temperature under aseptic conditions using a FACSCalibur® instrument (Becton Dickinson, San Jose, CA). Approximately 1x10^4 fluorescent parasites were collected using the single cell sort mode. Merozoites were pelleted by centrifugation and cloned by limiting dilution. The intensity of fluorescent clones was analyzed by further FACS analysis.

2.9 β-gal assays

Pyrimethamine resistant S. neurona clones obtained after transfection with pSnSAG1/LacZ-pSnSAG1/DHFR-TS were pelleted by centrifugation and resuspended in lysis buffer (100 mM HEPES, pH 8.0, 1 mM MgSO_4_, 1% Triton X-100, 5 Mm dithiothreitol). Enzyme activity was analyzed according to procedure described previously [77] using 2 mM chlorophenol red-β-galactopyranoside (CPRG) (Boehringer-Mannheim) as substrate. Reactions were performed in 96 well plates for 30 min at 37°C and then read at A_570 in a microplate reader (Molecular Devices Corporation). T. gondii (RH-2F) and N. caninum (Nc-LacZ) [75] stably expressing β-gal were used as positive controls in the assays. Growth assays of the SN3(LacZ) clone were performed as described previously [78]. Briefly, freshly lysed parasites purified from cell debris were added onto 96 well plate (3x10^5 per well) containing confluent monolayer of BT cells and varying concentrations of drugs. Culture plates were incubated at 37°C in 5% CO_2. At
different time points (days) post infection, infected host cells were lysed and enzyme activity was determined.

2.9 Growth assays using YFP

Growth assays of YFP expressing parasites were performed according to procedures described previously [74]. Briefly, BT cells were seeded onto black, 384 well tissue culture treated plates with optical bottoms (Falcon/Becton-Dickinson). Freshly lysed SN3(YFP).A1 parasites purified from cell debris were resuspended in culture medium without phenol red, and serially diluted in the presence of the drug. Plates were then incubated at $37^\circ$ C, 5% CO$_2$, and read daily in a BMG fluostar fluorescent reader (Hoffenberg, Germany).

Results

3.1 Isolation of 5’ flanking region of SnSAG1

In order to obtain the presumptive promoter region for the SnSAG1 gene, a S. neurona genomic library was screened with a DNA probe for the SnSAG1 ORF [29]. PCR, Southern blot and sequencing analyses confirmed the presence of the SnSAG1 locus in a single hybridizing cosmid clone. About 800 bases of sequence upstream of the ATG start codon was obtained, which presumably contains the promoter elements necessary for gene transcription. Sequence information for the SnSAG1 5’ flank has been deposited in GenBank (DQ286460).

3.2 Transient expression of β-gal and YFP in S. neurona

To demonstrate the feasibility of transgenic expression in S. neurona, 788 bp of the SnSAG1 5’ flanking region (676 bp non-coding region and 112 bp 5’ UTR) was amplified by PCR and directionally cloned upstream of the reporter genes LacZ and YFP
Although β-gal was not detected in *S. neurona* populations transfected with the pSnSAG1/LacZ plasmid, examination by immunofluorescence microscopy at 72 hours post-transfection revealed expression of β-gal in a small proportion of the parasites (Fig 2.2A). Similarly, electroporation of merozoites with the pSnSAG1/YFP-YFP plasmid yielded a comparable sub-population of parasites that exhibited YFP expression, as revealed by fluorescence microscopy (Fig 2.2B). Uniform fluorescence of the merozoites or the developing schizonts was observed for both reporter molecules, thus suggesting cytoplasmic localization of the proteins. These results demonstrated the feasibility of transient expression of foreign genes in *S. neurona* merozoites using the *SnSAG1* gene 5’ flanking region to control transgene expression. Enumerations of fluorescent organisms from two separate electroporations revealed approximately 11% and 7% of the parasite population expressing the transgene (data not shown), although other electroporations appeared to yield a lower proportion of transfected parasites. Therefore, the absence of detectable β-gal activity in the populations of parasites transfected with pSnSAG1/LacZ was attributed to the poor efficiency of transfection.

### 3.3 Stable transfection of *S. neurona*

To establish stable transgenic clones of *S. neurona*, a *T. gondii* DHFR-TS allele that confers resistance to pyrimethamine [79] was cloned into *S. neurona* transfection plasmid under transcriptional control of the *SnSAG1* promoter (Fig 2.1). The resulting pSnSAG1/DHFR-TS plasmid was electroporated into extracellular merozoites and transfected parasites were inoculated onto BT cell monolayers. Development of parasites in the presence of 10 µM pyrimethamine was apparent after 1 week, and the monolayer was 90% lysed after 4 weeks of drug selection. Parasite growth was not observed in
mock transfected parasites. To confirm that plasmid DNA had integrated into chromosomes, genomic DNA from eight pyrimethamine resistant (Pyr$^R$) clones was analyzed by Southern blot using a probe specific for the pBluescript vector. The hybridized fragments migrated slower than the linearized plasmid (Fig 2.3A), indicating that the plasmid was integrated into the parasite genome. Five clones appeared to be siblings, while three clones were unique. Two hybridizing fragments were observed in clone F2 indicating two integration events.

Genomic DNA from Pyr$^R$ recombinant clones was also digested with NdeI, which does not cleave pSnSAG1/DHFR-TS. Hybridization with the SnSAG1 gene 5’ flanking region revealed a restriction fragment from the endogenous SnSAG1 locus in both wild-type SN3 and the Pyr$^R$ clones (Fig 2.3B). A second, slower migrating fragment from the pSnSAG1/DHFR-TS transgene was also observed in Pyr$^R$ clones. The transgene in each Pyr$^R$ clone did not co-migrate with the pSnSAG1/DHFR-TS plasmid and exhibited hybridization intensity that was equal to the endogenous SAG1 fragment. These data are consistent with single-copy integration of the plasmid into the genome of the recombinant clones. Additionally, the plasmid had not integrated by homologous recombination at the SnSAG1 locus since migration of the endogenous fragment was not altered in the Pyr$^R$ clones.

3.4 Transgenic S. neurona expressing β-gal

A stable cell line of SN3 expressing β-gal was obtained by transfection of extracellular merozoites with plasmid pSnSAG1/LacZ-pSnSAG1/DHFR-TS that contained both the DHFR-TS and Lac-Z expression cassettes (Fig 2.1). Analysis of β-gal expression in the SN3(LacZ) clone indicated that enzymatic activity could be detected
with as few as $2 \times 10^3$ transgenic parasites (Fig 2.4). β-gal activity in the *S. neurona* clone was approximately 50% lower than the Nc(LacZ) *N. caninum* clone and the RH-2F *T. gondii* clone.

### 3.5 Transgenic *S. neurona* expressing YFP

To establish transgenic parasites that stably express YFP, the plasmids pSnSAG1/YFP-YFP (200 µg) and pSnSAG1/DHFR-TS (20 µg) were cotransfected into extracellular merozoites, and stable transformants were selected in the presence of 10 µM pyrimethamine. After 4 weeks of drug selection, a very small proportion (<0.5%) of the drug resistant parasites were fluorescent. FACS was employed to enrich for YFP-expressing parasites (Fig 2.5A), and the population of fluorescent merozoites was expanded and subsequently cloned by limiting dilution. Southern blot analysis of five YFP-expressing clones using a DNA probe specific for the pBluescript vector indicated that the clones were all siblings and contained a single copy of each transfected plasmid (Fig 2.5B). Fluorescent cytometry of one clone, SN3(YFP).A1 indicated 1000-fold greater fluorescence in the recombinant parasites relative to wild-type SN3 (Fig 2.5C).

### 3.6 Use of transgenic parasites in growth assays

Recombinant parasites expressing reporter molecules have proven useful for a variety of experimental purposes. To demonstrate the utility of using the transgenic *S. neurona* clones for assessing in-vitro growth of parasites, the *S. neurona* clones SN3(YFP1).A1 and SN3(LacZ) were grown in the presence of diclazuril, toltrazuril or pyrimethamine. A progressive increase in relative fluorescence (Fig 2.6) or β-gal activity (Fig 2.7) was observed over 8-10 days when parasites were grown in the absence of drug treatment. With increasing concentrations of drugs, the relative fluorescence or β-gal
activity increased at a slower rate than non-treated cultures, consistent with reduced
growth of the parasites. In the presence of pyrimethamine, parasites grew at
concentrations higher than the lethal dose of 2.5 µM [80], as expected since they express
a mutant DHFR-TS allele that confers resistance to this folate inhibitor. However, a high
level of pyrimethamine (40 µM) was capable of inhibiting parasite growth. Importantly,
the data generated with the SN3(YFP).A1 clone corresponded well with the data obtained
with the SN3(LacZ) clone, thereby confirming that either reporter molecule can be used
to reliably monitor parasite growth.

Discussion

Development of DNA transfection capabilities provides a powerful approach for
genetic dissection of an organism. In an effort to enhance investigation of the
apicomplexan S. neurona, we have established methods for expressing foreign genes in
this parasite. Reporter molecules YFP and β-gal were efficiently expressed in S. neurona
transfected with plasmid constructs containing SnSAG1 promoter region to drive
transgene transcription. As SnSAG1 is highly represented in the S. neurona EST database
[27] and appears to be constitutively expressed during intracellular development [29], it
was expected to provide an excellent promoter for both transient expression and stable
transformation of this parasite. Indeed, we successfully generated recombinant S.
neurona that are pyrimethamine resistant and stably express YFP and β-gal.

A number of positive selection markers have been developed for the related
apicomplexan T. gondii. These include chloramphenicol acetyltransferase (cat) [81],
mutant dihydrofolate reductase [79], hypoxanthine xanthine guanine phosphoribosyl
transferase (HXGPRT) [82], phleomycin (ble) [83] and tryptophan (trp) synthase [84].
Of these, mutant dihydrofolate reductase gene seems to be a highly efficient selectable marker, but its use has been discouraged in *T. gondii* since it renders parasites resistant to a primary chemotherapeutic agent for the treatment of toxoplasmosis [85]. Given that *S. neurona* is not a human pathogen, the mutant DHFR-TS can be used as a selectable marker without posing a hazard to investigators working with pyrimethamine resistant parasites.

Parasites that stably express the reporter molecules β-gal or YFP have been shown previously to be useful in monitoring *in-vitro* growth and primary screening of new chemotherapeutic agents [78, 86]. As both assays are performed in microtiter plates, multiple drugs can be screened simultaneously for anti-coccidial activity. Reporter molecules like green fluorescent protein (GFP) have also proven useful as fusion proteins to identify signal motifs that direct proteins to specific organelles [87]. Similarly use of a promoterless GFP allowed for identification of novel proteins targeted to various subcellular compartments [88].

The development of molecular genetic capabilities for *S. neurona* has opened many new avenues of investigation into the biology of this parasite. The genetic tools described here in have already proven valuable for examining development and segregation of the apicoplast during endopolygeny [26]. A variety of useful genetic manipulations are now feasible, such as gene knockouts, complementation by introduction of heterologous genes, and use of reporter molecules to monitor gene expression. Prior studies utilizing transient transfection have implicated putative regulatory elements that control gene expression in *T. gondii* and *plasmodium falciparum* [47, 52-54, 63]. In similar fashion, cis-acting elements required for gene transcription in
*S. neurona* can be identified using transient transfection system (Chapter Three). These studies should provide a better understanding of the mechanisms that temporally control gene expression during intracellular development.
Figure 2.1. Schematic representation of plasmids used for expression of YFP, β-gal, and mutant DHFR-TS in *S. neurona*. The 5’ flanking region of the SnSAG1 gene was utilized to drive transcription of the transgenes.
Figure 2.2. *S. neurona* electroporated with reporter constructs transiently expressed β-gal (A) and YFP (B). Parasites were transfected with either pSnSAG1/LacZ or pSnSAG1/YFP-YFP and inoculated onto monolayer of BT cells. YFP expressing parasites were viewed with a Zeiss axioscope equipped for epiflourescence microscopy. β-gal expressing parasites were revealed by immunoflourescence microscopy using a monoclonal antibody against *E.coli* β-gal. Bars = 5µm.
Fig. 2.3. Southern blot analysis of pyrimethamine resistant (Pyr$^R$) clones of *S. neurona*.

(A) Genomic DNA digested with *Hind*III, which cleaves at one site in the pSnSAG1/DHFR-TS plasmid construct (see Fig 1), was hybridized with a probe specific for the pBluescript vector. Clones D8, E2, G10, C8, and F2 appeared to be siblings, clone D10 has 2 copies of the transfected plasmid, and F6 & D4 are unique clones. (B) Genomic DNA was digested with NdeI, which does not cleave pSnSAG1/DHFR-TS, and hybridized with the *SnSAG1* gene 5′ flanking region. The endogenous *SnSAG1* fragment (arrow) was observed in wild type SN3 and the recombinant clones. The transgene fragments (*) did not comigrate with the pSnSAG1/DHFR-TS plasmid (supercoiled and nicked forms; arrowheads) and were only present in the recombinant clones. MW = molecular weight marker.
Figure 2.4. Analysis of β-gal expression in a transgenic clone of *S. neuronae*. Greater enzymatic activity was detected with increasing parasite numbers. The *T. gondii* RH-2F and *N. caninum* NC (LacZ) clones that stably express β-gal were used as positive controls. β-gal activity was not detected in wild-type SN3 merozoites.
Figure 2.5. Isolation of *S. neurona* clones expressing YFP. (A) Merozoites expressing YFP were selected by gating highly fluorescent cells (R1), which represented about 0.45% of the drug resistant population. (B) Southern blot analysis of YFP-expressing clones of *S. neurona* obtained by cotransfection of plasmids pSnSAG1/YFP-YFP and pSnSAG1/DHFR-TS demonstrated that all clones appeared to be siblings and have a single copy of each plasmid. Genomic DNA was digested with HindIII, which cuts once in each plasmid, and a probe specific for pBluescript was used for hybridization. No HindIII sites are present in the probe sequence. (C) SN3 wild-type and stable YFP-expressing clone SN3 (YFP1).A1 were analyzed by flow cytometry. Transgenic *S. neurona* were about 1000 times more fluorescent than the wild-type parasites.
Figure 2.6. Growth curves of SN3 (YFP) A1 merozoites in the presence of diclazuril, toltrazuril and pyrimethamine. Parasites were loaded into 384-well plates containing confluent monolayers of BT cells, the plates were incubated for 10 days in the presence of each drug at various concentrations, and relative fluorescence was read daily with a fluorescence plate reader. Average values from three identical wells were used in plotting the growth curve.
Figure 2.7. Growth curves of SN3 (LacZ) in the presence of diclazuril, toltrazuril and pyrimethamine. Parasites were added (3x10^5 per well) into 96-well plates containing confluent monolayers of BT cells and varying concentrations of drugs. At different time points (days) post infection, infected host cells were lysed in the wells and assayed for β-gal activity. Average values from three identical wells were used in plotting the growth curve.
Chapter Three: GAGACGC is a cis-acting element critical for SnSAG1 expression in *Sarcocystis neurona*

Introduction

*Sarcocystis neurona* is an apicomplexan parasite identified as the major cause of equine protozoal myeloencephalitis (EPM) in horses [6]. Opossums act as the definitive host for *S. neurona* [2], and various small mammals such as raccoons, armadillos and skunks [3] [4] [5] can serve as intermediate hosts. The parasite shares many features of apicomplexan organisms such as an obligatory intracellular life & gliding motility and *S. neurona* EST project [27] has revealed numerous genes that are homologous to *Toxoplasma gondii* surface [29] and secretory [30] [92] antigens. However, there are some distinct differences such as absence of rhoptries, parasitophorous vacuole in merozoite/schizont stages in *S. neurona* [6].

Apicomplexan parasites utilize a variety of mechanisms for cell division. For example *T. gondii* divides by endodyogeny while *S. neurona* divides by endopolygeny. During *S. neurona* endopolygeny, it is apparent that there is temporal control of gene expression. Specifically, both SnMIC10 [30] and SnNTP1 [92] are not expressed until terminal stages of endopolygeny. However, the mechanism of gene regulation during intracellular development of *S. neurona* is unknown. It is likely that there are regulatory elements upstream of these genes that control gene expression at different stages of development.

In higher eukaryotes core promoter motifs and distal regulatory elements are well defined [42]. The core promoter motifs are located ~35 bases either upstream or downstream of the transcription start site and include TATA box [32], initiator [33],
TFFIIB recognition element (BRE) [34], and downstream promoter element (DPE) [35]. However, not all genes have the full complement of core promoter elements. A very few cis-acting sequence elements have been identified in apicomplexan parasites. The heptanucleotide motif WGAGACG has been identified in the promoters of SAG1 and GRA genes of Toxoplasma [47, 52] and a palindromic G-box was found in heat shock protein (hsp) gene family of Plasmodium spp [63]. A consensus TATA box has been identified in knob-associated histidine rich-protein (kahrp) & glycoporphin binding protein 130 (GBP-130) gene promoters of P. falciparum [64].

To gain a better understanding of gene regulation during intracellular development of S. neurona, we have initiated studies to define regulatory sequences in this organism. As a first step towards this objective, the 5’ flanking region of SnSAG1 was subjected to comparative and functional analysis using a recently developed transfection system for S. neurona (Chapter Two). Herein, we describe the identification of a conserved 7-base motif that is essential for gene expression.

Materials and Methods

2.1 Parasite cultures

S. neurona strain SN3 was propagated as merozoites by serial passage in bovine turbinate (BT) cells. BT cells were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum, 2 mM sodium pyruvate, and Pen/Strep Fungizone (Bio Whittaker, Inc). Freshly-lysed extracellular parasites were passed once through 20 G, 23 G and 25 G needles, and purified from the cell debris with 3.0 µm membrane filters, as described previously [29].

2.2 Isolation and analysis of S. neurona genomic DNA clones
To obtain genomic DNA clones of the SnSAG2, SnSAG3, SnNTP and Sarco_21 loci, colony lifts of a S. neurona genomic cosmid library (Chapter Two) were hybridized with digoxigenin (DIG)-labeled DNA probes, and hybridized colonies were detected using DIG High Prime Labeling and Detection kit II (Roche Molecular Biochemicals). The 5’ flanking region of the each genomic clone was sequenced using ABI Prism™ Big Dye Terminator Cycle sequencing reaction mix (PE applied biosystems). Sequencing reactions were purified with AutoSeq™ G-50 columns and analyzed on an ABI 310 genetic analyzer (PE Applied Biosystems). Sequences were manually edited and analyzed using Vector NTI Advance™ 9.0 (Informax™, Invitrogen).

2.3 Total RNA isolation and 5’ RACE

To identify transcription start sites of SnSAG1, 5’ RACE was performed using First Choice® RLM-RACE (Ambion) according to the manufacturer’s instructions. Briefly, total RNA was purified from S. neurona merozoites using TRIZOL (Invitrogen). RNA was treated with calf intestinal alkaline phosphatase (CIP), and the reaction was stopped by phenol:chloroform extraction. The RNA was treated with Tobacco Acid Pyro-phosphatase (TAP), followed by ligation of a 45-base RNA adapter using T4 RNA ligase. A random-primed reverse-transcription was then performed using M-MLV reverse transcriptase, followed by a nested PCR. The outer 5’ RLM-RACE PCR was performed using the 5’ RACE outer primer and a SnSAG1 outer primer (Table 3.1). The inner 5’ RLM-RACE PCR was performed using primers 5’ RACE inner primer and a SnSAG1 inner primer (Table 3.1). The amplification products were analyzed by agarose gel electrophoresis and nucleotide sequencing, as described above.

2.4 Plasmid Constructions
Sequences of primers used in this study are provided in Tables 3.1-3.3. The renilla luciferase (RLUC) ORF was amplified from pRL-CMV (Promega) using the primers RLU.HindIII.F and RLU.BamHI.R that created HindIII and BamHI sites, respectively (Table 3.1). The amplification product was cloned into HindIII and BamHI-digested pBluescript SK+ to make pRLUC. The *SnSAG2* 3’ flanking sequence was amplified from a *SnSAG2* cosmid clone using primers (*SnSAG2* 3’.BamHI.F and *SnSAG2* 3’.NotI.R) that created BamHI and NotI sites (Table 3.1), and the product was cloned into BamHI and NotI-digested pRLUC to make pRLUC/SnSAG2. The *SnSAG1* 5’ flanking region was amplified using primers (*SnSAG1* 5’.XhoI.F and *SnSAG1* 5’.HindIII.R) that created XhoI and HindIII sites (Table 3.1), and the amplification product was ligated into XhoI and HindIII-digested pRLUC/SnSAG2 to make pSnSAG1/RLUC/SnSAG2.

The firefly luciferase (FLUC) ORF of pGL2-Basic (Promega) was amplified with primers (FLU.HindIII.F and FLU.BamHI.R) that created HindIII and BamHI sites (Table 3.1). The amplification product was ligated into HindIII and BamHI-digested pBluescript SK+ to make pFLUC. The *SnSAG1* 3’ flanking region was amplified from a *SnSAG1* genomic clone with primers (*SnSAG1* 3’.BamHI.F and *SnSAG1* 3’.NotI.R) that created BamHI and NotI sites (Table 3.1). The product was cloned into BamHI and NotI-digested pFLUC to make pFLUC/SnSAG1.

Amplification of the *SnSAG1* promoter fragments was performed with forward primers that added a XhoI site and a single reverse primer that added a HindIII site (*SnSAG1* 5’.HindIII.R; Table 3.2). The amplification fragments were cloned upstream of the FLUC ORF in pFLUC/SnSAG1 to create *SnSAG1* promoters that varied in length.
These deletion constructs were named based on the length of the promoter from the first transcription start site.

Modified promoter fragments were created by overlap extension PCR [89] using forward (F2), reverse (R2), forward mutation (FM) and reverse mutation (RM) primers (Table 3.3). PCR 1 was performed using primers FM and R2, whereas PCR 2 was done using primers RM and F2. The products of PCR 1 & PCR 2 were mixed and used as template in PCR 3, which was performed using primers F2 and R2 that contained XhoI and HindIII sites, respectively. The product of PCR 3 was digested with XhoI and HindIII and cloned into XhoI and HindIII-digested pFLUC/SnSAG1, thus creating plasmids pSnSAG1(-179\textsuperscript{M1})/FLUC/SnSAG1, pSnSAG1(-136\textsuperscript{M2})/FLUC/SnSAG1, pSnSAG1(-136\textsuperscript{M3})/FLUC/SnSAG1 and pSnSAG1(-179\textsuperscript{M1,FLIP})/FLUC/SnSAG1.

To replace the region between Motif 1 and 2, a PstI site was introduced one base upstream of Motif 2 in pSnSAG1(-136)/FLUC/SnSAG1 to make pSnSAG1(-136.PstI)/FLUC/SnSAG1. The amplification product obtained with primers Supercos.XhoI.F & Supercos.PstI.R and pSupercos 1 (Stratagene) as template was digested with XhoI & PstI and subcloned into XhoI & PstI-digested pSnSAG1(-136.PstI)/FLUC/SnSAG1 to make pSnSAG1(-136\textsuperscript{R})/FLUC/SnSAG1 construct. The SnSAG1 promoter fragment was amplified with primers (-223.HindIII.F & -223.XhoI.R) that created HindIII & XhoI sites (Table 3.1) and cloned into HindIII & XhoI-digested pFLUC/SnSAG1 to make the pSnSAG1(-223\textsuperscript{FLIP})/FLUC/SnSAG1 construct.

Amplification of the TgSAG1 promoter was performed with primers (TgSAG1.XhoI.F and TgSAG1.HindIII.R) that created XhoI & HindIII sites (Table 3.1). The amplification product was cloned into XhoI & HindIII-digested pFLUC/SnSAG1 &
pRLUC/SnSAG2 to make pTgSAG1/FLUC/SnSAG1 & pTgSAG1/RLUC/SnSAG2 respectively.  

All PCR reactions were performed using Klen Taq polymerase (Sigma) to reduce mutations introduced by the polymerase. All plasmid constructs were confirmed by sequencing.

2.5 Parasite Transfections

Freshly-lysed *S. neurona* merozoites were washed with Hanks buffer, HEPES, EGTA (HHE) and resuspended in Cytomix buffer supplemented with 2mM ATP and 5mM glutathione [67]. A 40 µg aliquot of the experimental FLUC luciferase plasmid and 10 µg of the RLUC control plasmid (pSnSAG1/RLUC/SnSAG2) were mixed with 1x10^7 parasites in a 2-mm gap cuvette (BTX), and the parasites were electroporated at 2 kV, 129Ω using a BTX ECM 600 electroporation system. Transfected parasites were allowed to recover for 10 min at RT and transferred to a confluent monolayer of BT cells in 6-well plates.

2.6 Luciferase assays

Seventy-two hours post-transfection, parasite cultures were analyzed for luciferase activity using a Dual-luciferase® reporter assay system (Promega). Briefly, medium was removed from the culture plates, the monolayers were washed once with PBS, and 200 µl of passive lysis buffer was added to the wells. Lysed cells were collected by scraping, transferred to microcentrifuge tubes, and subjected to two freeze-thaw cycles. 20 µl of cell lysate was analyzed for firefly luciferase activity in a TD-20/20 luminometer (Turner Designs). Stop & Glow® reagent was added to quench FLUC luminescence and initiate RLUC activity, which was then quantified in the luminometer.
Firefly luciferase activity was normalized to the level of RLUC activity to control for variable transfection efficiencies. The firefly luciferase activity observed in parasites transfected with pSnSAG1(-701)/FLUC/SnSAG1 was set at 100%.

2.7 Nuclear extract preparation

Nuclear extracts were prepared from 2x10^8 *S. neurona* merozoites using NE-PER® (Pierce) according to the manufacturer’s instructions. Briefly, freshly-lysed merozoites were harvested and resuspended in 200 µl of ice-cold cytoplasmic extraction reagent (CER)-I. The parasite lysate was incubated on ice for 10 min followed by addition of 11 µl of ice-cold CER-II. The tube was incubated on ice for one minute, vortexed, and centrifuged at 16000 x g for 5 min. The supernatant fraction (cytoplasmic extract) was discarded and the insoluble pellet containing merozoite nuclei was resuspended in 100 µl of ice-cold nuclei extraction reagent (NER). The resuspension mixture was incubated on ice for 40 min, with vortexing every 10 min. The tube was centrifuged at 16000 g for 10 min, and the supernatant (nuclear extract) fraction was transferred to clean tube and stored in 20 µl aliquots at -80°C.

2.8 Electrophoresis mobility shift assay (EMSA)

For the EMSAs, three different probes were made. Probe 1 was a double stranded oligo containing wild-type copy of motif 1 (AGGAAACGAGACGCCAGCA). Probe 2 was a double stranded oligo containing mutant copy of motif 1 (AGGAAACTCAGTTAGCCAGCA). Probe 3 was a PCR product containing all 3 motifs of *SnSAG1* promoter obtained using primers (forward, 5’ AATGCTAGCTGCTCCGGGTGC; reverse, 5’ ACAGATGGACAGCGTGCTCTG). The probes were end labeled with [γ-32P]ATP (ICN, Costa Mesa, CA) as described previously
Binding reactions were performed for 30 minutes at room temperature in 10 µl volumes containing approximately 2 µg S. neurona nuclear extract and 1 µl (~50,000 cpm) radiolabeled oligonucleotide in binding buffer [4% glycerol, 1 mM MgCl₂, 0.5 mM EDTA (pH 8.0), 0.5 mM DTT, 50 mM NaCl, 10 mM Tris (pH 7.5) and 50 µg poly(dI-dC)]. Reaction products were separated in 4% nondenaturing polyacrylamide gels and analyzed with a Storm Phosphorimager System (Molecular Dynamics, Sunnyvale, CA).

Results

3.1 Bioinformatic analysis of 5’ flanking regions of S. neurona genes and identification of transcription start sites of SnSAG1

The sequencing of the 5’ flanking region of SnSAG1 has been described previously (Chapter Two). To obtain sequence for the promoter region of various S. neurona genes, a S. neurona genomic library was screened with probes specific for the SnSAG2, SnSAG3, SnNTP and Sarco_21 ORFs. A single genomic clone for each locus was isolated, and approximately 0.6 kb of sequence upstream of the ATG start codon was obtained. Comparative analysis of the 5’ flanking sequences using MEME and Vector NTI® software revealed a 7-base motif GCGTCTC that was present in either orientation and in variable copy numbers (Fig 1). Additionally, the location of these motifs was highly variable across the different promoters. The analysis did not reveal the standard eukaryotic promoter elements TATA-box, DPE, CAAT-box or TFIIB recognition element (BRE).

To determine the transcription start site of the SnSAG1 gene, 5’ RACE was performed. This analysis indicated the presence of two transcription start sites (Fig 3.1). The first transcription start site is 88 bases upstream of ATG start codon, while the
second transcription start site is 122 bases upstream of the start codon. The transcription start sites for the four other *S. neurona* genes are based on EST data and presumptive.

3.2 Functional analysis of the SnSAG1 promoter

To define a *S. neurona* promoter, the *SnSAG1* 5’ flanking region was subjected to functional analysis using a dual luciferase reporter system. As an initial test to establish the importance of both the *SnSAG1* 5’ and the 3’ flanking regions for gene transcription, parasites were transfected with plasmid constructs that lacked the 5’ flanking region, the 3’flanking region or both. High levels of FLUC activity were observed when the full-length parental construct pSnSAG1(-701)/FLUC/SnSAG1 was electroporated. In contrast, FLUC reporter activity was not detected with the constructs pFLUC and pFLUC/SnSAG1, while the pSnSAG1/FLUC yielded about 40% luciferase activity relative to the full length construct (Fig 3.2). These data indicated that both the 5’ and the 3’ flanking regions of *SnSAG1* are needed for efficient expression of FLUC reporter gene.

To identify cis-acting elements that play a role in SnSAG1 gene expression (see Fig 3.3A), a series of *SnSAG1* 5’ deletion constructs were transfected into parasites and analyzed for FLUC reporter activity. The FLUC activity detected from parasites transfected with the full length parent construct pSnSAG1(-701)/FLUC/SnSAG1 was set at 100%. Transfection with the deletion constructs from -627 to -179 gave levels of reporter activity that was comparable to the full length (-701) promoter (Fig 3.3B). In contrast, FLUC activity was not detected from parasites transfected with the -101 and -23 constructs. These data indicated that the 78 bp region between -179 and -101 contains sequences that are critical for FLUC expression.
To more precisely define the sequences necessary for gene expression, the 78-bp region between -179 and -101 was subjected to further deletion analysis (Fig 3.4). Luciferase activity was detected in parasites transfected with the construct that contained 149 bases upstream of the transcription start site (-149), while transfection with the plasmid construct that included 123 bases upstream of the transcription start site (-123) did not yield reporter activity. These results indicated that the 26-bp region between -149 and -123 contains sequences required for FLUC expression. Importantly, one of three GCGTCTC motifs is present in this 26-base region of SnSAG1 promoter (Fig 3.3A).

To determine if the motif 1 (GAGACGC) located between -149 and -123 is an important sequence element for gene expression, constructs were made that delete or alter the motif, and reporter activity produced in parasites transfected with these plasmids was analyzed (Fig 3.5). Parasites transfected with the -129 construct showed no reporter activity, while parasites transfected with plasmid that included 7-base motif GAGACGC (-136) yielded full promoter activity (Fig 3.5). This indicated that motif 1 is a sequence element that is critical for reporter molecule expression. To further confirm that motif 1 is critical for expression, a construct that contained a mutant copy of motif 1 (-179\textsuperscript{M1}) was made and promoter activity analyzed. Luciferase activity was not detected when the -179\textsuperscript{M1} construct was transfected into parasites (Fig 3.5), thus confirming that the motif 1 present between -129 and -136 is a critical sequence element that is required for FLUC expression.

Two additional motifs, motif 2 and motif 3 are present in the \textit{SnSAG1} 5’ flanking region (Fig 3.3A). To know if these two additional motifs play a role in gene transcription, plasmids containing mutant copies of the motifs were constructed. As
shown in Fig. 3.5, mutation of motif 2 (-136\textsuperscript{M2}) reduced luciferase activity by 50% compared to the full-length promoter. In contrast, mutation of motif 3 (-136\textsuperscript{M3}) did not affect luciferase activity. These data imply that motif 2 is involved in promoter activity but is not essential, while motif 3 may not be important for gene expression.

To know if there are additional regulatory elements between motif 1 and motif 2, the \textit{SnSAG1} promoter region between these motifs was replaced with a heterologous DNA fragment of identical length. When the resulting construct (-136\textsuperscript{R}) was transfected into parasites, no FLUC activity was detected implying that there are sequence elements between motif 1 and motif 2 that are critical for gene expression.

The GCGTCTC motif is present in forward and reverse orientations in \textit{S. neurona} gene promoters. To know if motif 1 can function in both orientations, a construct was designed in which motif 1 was placed in the reverse orientation GCGTCTC (-223\textsuperscript{M1.\text{FLIP}}). FLUC activity was not detected in parasites transfected with this construct indicating that orientation of motif 1 is critical for its function (Fig 3.6). In the related apicomplexan parasite SAG1 promoter is able to function in reverse orientation [47]. To know if \textit{SnSAG1} promoter is also a divergent promoter the fragment from -223 to +85 was placed in reverse orientation. Transfection of the resulting construct (-223\textsuperscript{FLIP}) into parasites yielded about 25% FLUC activity relative to full-length construct (Fig 6). These results suggest that the \textit{SnSAG1} promoter can function in reverse orientation, but at a reduced capacity.

3.3 The \textit{T. gondii} TgSAG1 and \textit{S. neurona} SnSAG1 promoters are not efficient for heterologous expression
Motif 1, which appears to be critical for SnSAG1 expression, is very similar to the cis-acting element WGAGACG that has been identified in some *T. gondii* gene promoters [47, 52]. To determine whether the *T. gondii* TgSAG1 promoter would be recognized by the *S. neurona* transcriptional machinery, *S. neurona* merozoites were transfected with pTgSAG1/FLUC/SnSAG1. This construct yielded only 2% FLUC activity relative to the *SnSAG1* full length promoter construct (Fig 3.7A). Similarly, *T. gondii* tachyzoites transfected with pSnSAG1/FLUC/SnSAG1 construct produced only 4% FLUC activity compared to the pTgSAG1/FLUC/SnSAG1 construct (Fig 3.7B). These data suggest that the promoters in *T. gondii* and *S. neurona* will not function efficiently in the heterologous organism, despite the similarity of the sequence motifs that have been identified as important for gene expression in these related species.

3.4 Motif 1 binds *S. neurona* proteins

The functional analysis of *SnSAG1* promoter indicated that motif 1 is critical; motif 2 is essential but not necessary and motif 3 does not play a role in gene transcription. To know if motif 1 is a protein binding motif, EMSAs were performed with a *S. neurona* nuclear extract and DNA probes containing the motif. An electrophoretic shift of a 21-bp probe that included the wild-type motif 1 and 14 adjacent nucleotides was observed after incubation with the nuclear extract (Fig 3.8). A 100-fold excess of the non-labeled probe prevented protein binding to the labeled probe (Fig 3.8). A probe containing a mutant copy of motif 1 did not efficiently bind proteins in the nuclear extract, as revealed by a significant reduction in the amount of shifted probe (Fig 3.8). A 254 bp fragment of the SnSAG1 promoter that included all 3 motifs was similarly capable of binding proteins in the nuclear extract as shown by a
Large complex of shifted probe (Fig 3.8). A 100-fold excess of non-labeled probe abolished binding of proteins to the labeled fragment. However, when a 100-fold excess of the 21-bp probe was used as competitor, the shifted complex was unaffected, suggesting that there are additional protein-binding sequence elements in the 254-base fragment.

Discussion

*S. neurona* undergoes a complex method of cell division termed endopolygeny, during which there is temporal control of gene expression. The biology of this parasite with regard to gene regulation has not been defined, primarily because of unavailability of molecular genetic tools. The recent development of a transfection system for *S. neurona* has enabled the studies described herein aimed at identifying sequence elements that regulate gene expression. In this study, we have identified a 7-base motif required for SnSAG1 expression. This element appears to function in an orientation dependent manner by interacting with transcription factors.

As the deletion or alteration of the motif in the *SnSAG1* promoter completely abolishes promoter activity the GAGACGC motif may be one of the core promoter elements in *S. neurona*. In *Toxoplasma*, an identical motif WGAGACG was found in the promoter regions of *GRA1*, *GRA2*, *GRA5*, *GRA6* and *TUB1* genes [52]. This element is present in either orientations, at variable locations and multiple copy number similar to 7-base motif of *S. neurona* [52]. A single copy of the motif proximal to transcription start site is sufficient for basal promoter activity of *GRA2*. Similar to our findings in this study, not all of these motifs play a role in transcription of *GRA* genes [52]. The *Toxoplasma* SAG1 promoter contains six, tandemly-repeated, 27-base motifs and a
minimum of two 27-bp repeats are sufficient for promoter activity. Interestingly, the heptanucleotide motif (WGAGACG) is the central core of the 27-base motifs of TgSAG1 promoter [47]. We also identified GAGACGC motif in promoter regions of NcMIC10 and NcNTP1 genes of *N. caninum* (Gaji and Howe, unpublished data). However, this motif has not been described in *Plasmodium* spp, thus suggesting that GAGACGC motif is a conserved cis-acting element only in tissue dwelling coccidia.

In *Plasmodium*, the available information suggests that gene regulatory elements may be different from other parasites in the phylum. A palindromic G-box (A/GNGGGGC/A) is involved in transcription of the *hsp* gene family in *Plasmodium* [63]. While no classical eukaryotic promoter elements are present *hsp* family gene promoters [63], a consensus TATA-box is present in *kahrp* & *GBP-130* gene promoters that interacts with the native *P. falciparum* TATA binding protein [64].

It is unlikely that the 7-base core promoter motif is involved in developmental regulation of gene expression in *S. neurona* since the motif is present in SnSAG1, a constitutively expressed gene [29] and also in SnNTP1 [92], a temporally regulated gene during endopolygeny. Sequence elements that act as repressors govern stage specific expression of enolase genes in *Toxoplasma* [59]. *T. gondii* has two genes of enolase (*ENO1* and *ENO2*) that are stage specific; *ENO1* is expressed in bradyzoites and *ENO2* is expressed in tachyzoites [57, 58]. The region –1245 to –625 acts as a repressor of *ENO1* in tachyzoites and 2 repression regions in the *ENO2* promoter region, (–1929 to –1067 and –456 to –222), block expression in bradyzoites [59]. Recent studies have also suggested that epigenetics or chromatin remodeling is an important mode of gene regulation in apicomplexan parasites [91].
In summary, this is the first report describing identification of a discrete core promoter element in *Sarcocystis* species. Further analysis of *SnSAG1* promoter will help in identifying additional regulatory motifs and thus provide an insight about the mechanism of transcriptional regulation in *S. neurona*.

The knowledge of cis-acting elements and transcription factors in apicomplexan parasites is still rudimentary. Bioinformatic analysis using programs designed for higher eukaryotes may not be an efficient way of identifying regulatory sequence elements in apicomplexans due to sequence divergence [46]. But comparative genomics within Apicomplexa combined with functional analysis could be a better approach.
Table 3.1. Primers used in 5’ RACE & designing plasmid constructs

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SnSAG1 outer primer</td>
<td>TGCATAGACCTGCATCTGAC</td>
</tr>
<tr>
<td>SnSAG1 inner primer</td>
<td>TAACAGTCTTTCCATGCCTGAC</td>
</tr>
<tr>
<td>RLU.HindIII.F</td>
<td>GATCAAGCTTAGCCACCATGGGCTTCGAAAG</td>
</tr>
<tr>
<td>RLU.BamHI.R</td>
<td>GATCGGATCCCCGCTCTAGAATTATTTGTC</td>
</tr>
<tr>
<td>SnSAG2 3’.BamHI.F</td>
<td>ACTGGGATCCGCGTGAAGATGAAATTGTC</td>
</tr>
<tr>
<td>SnSAG2 3’.Notl.R</td>
<td>ACTGGGCGGCCGCTATGACCGTCCCGCTCGTGTAA</td>
</tr>
<tr>
<td>SnSAG1 5’.Xhol.F</td>
<td>ACTGCTCGAGAAACAGTGAGGCAGATATG</td>
</tr>
<tr>
<td>SnSAG1 5’.HindIII.R</td>
<td>ACTGACAAGCTTGCGAGAATGACAGATGGACAG</td>
</tr>
<tr>
<td>FLU.HindIII.F</td>
<td>GATCAAGCTTTGAAATGGAAGACGCAAAAAAC</td>
</tr>
<tr>
<td>FLU.BamHI.R</td>
<td>GATCGGATCCGCTCATCGTGAATTACGTTAC</td>
</tr>
<tr>
<td>SnSAG1 3’.BamHI.F</td>
<td>ACTGGGATCCCTAATCCTGCCGTTAGGTC</td>
</tr>
<tr>
<td>SnSAG1 3’.Notl.R</td>
<td>ACTGGCGGCCGCAATGGCCGATGCTAC</td>
</tr>
<tr>
<td>Supercos.Xhol.F</td>
<td>ACTGCTCGAGGAGCGAGCGAGCGTGGATGG</td>
</tr>
<tr>
<td>Supercos.PstI.R</td>
<td>ACTGCTCGAGGAGCGAGCGAGCGTGGATGG</td>
</tr>
<tr>
<td>TgSAG1.Xhol.F</td>
<td>ACTGCTCGAGCTTTACCATGCGTCTAG</td>
</tr>
<tr>
<td>TgSAG1.HindIII.R</td>
<td>ACTGAAGCTTTCTGCAAAAAACCAGAAG</td>
</tr>
<tr>
<td>-223.HindIII.F</td>
<td>ACTGAAAGCTTCGAATCAGATGACAG</td>
</tr>
<tr>
<td>-223.Xhol.R</td>
<td>ACTGCTCGAGGAGCGAGATGACAG</td>
</tr>
</tbody>
</table>
Table 3.2. Primers used in making *S. nigrum* 5’ deletion constructs

<table>
<thead>
<tr>
<th>Name of forward primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>-701</td>
<td>ACTGCTCGAGGAAACAGTGAGGCGAGATATG</td>
</tr>
<tr>
<td>-627</td>
<td>ACTGCTCGAGCTGTTAGTCTAAATTG</td>
</tr>
<tr>
<td>-478</td>
<td>ACTGCTCGAGCGGTGTTATCTCCACCAGGG</td>
</tr>
<tr>
<td>-328</td>
<td>ACTGCTCGAGATACTGCTAGTAAACTGGTTC</td>
</tr>
<tr>
<td>-179</td>
<td>ACTGCTCGAGAATGCTAGCTGTTCCACCAGGG</td>
</tr>
<tr>
<td>-101</td>
<td>ACTGCTCGAGCGGCGGTATAGGCGGTAC</td>
</tr>
<tr>
<td>-23</td>
<td>ACTGCTCGAGCGGAGCCTTTAATTACTTCCACCAGGG</td>
</tr>
<tr>
<td>-149</td>
<td>ACTGCTCGAGAATGCGAGGAAACGACGC</td>
</tr>
<tr>
<td>-123</td>
<td>ACTGCTCGAGAAGCCCTTGTTTCGATAC</td>
</tr>
<tr>
<td>-136</td>
<td>ACTGCTCGAGGAGCCAGCCAGAAGCCAGGC</td>
</tr>
<tr>
<td>-129</td>
<td>ACTGCTCGAGGAGCCCTTTCACAGGCGG</td>
</tr>
<tr>
<td>-223</td>
<td>ACTGCTCGAGCGAGAATCTAGATATGATGACAG</td>
</tr>
</tbody>
</table>
Table 3.3. Primers used in making *SnSAG1* 5’ mutant constructs by overlap extension PCR

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motif 1. F2</td>
<td>ACTGCTCGAGAATGCTAGCTGTCGCCGGGTGC</td>
</tr>
<tr>
<td>Motif 1. R2</td>
<td>ACTGACAAGCTTGCGAGAATGACAGATGGACAG</td>
</tr>
<tr>
<td>Motif 1. RM</td>
<td>GAAATGCAGGAAACTTTTTTTGCCAGCAAGCCG</td>
</tr>
<tr>
<td>Motif 1. FM</td>
<td>CGGCTTGCTGGCAAAAAAAGTTTCTGCATTTC</td>
</tr>
<tr>
<td>Motif 2. F2</td>
<td>ACTGCTCGAGGAGAGACGCGCCCCAGCACAGCCG</td>
</tr>
<tr>
<td>Motif 2. R2</td>
<td>ACTGACAAGCTTGCGAGAATGACAGATGGACAG</td>
</tr>
<tr>
<td>Motif 2. RM</td>
<td>GCACACATGCTTTTTTTAAATGTCGTATC</td>
</tr>
<tr>
<td>Motif 2. FM</td>
<td>GATACGCACATTAAAAAAAGCATGTGTGCA</td>
</tr>
<tr>
<td>Motif 3. F2</td>
<td>ACTGCTCGAGGAGAGACGCGCCCCAGCACAGCCG</td>
</tr>
<tr>
<td>Motif 3. R2</td>
<td>ACTGACAAGCTTGCGAGAATGACAGATGGACAG</td>
</tr>
<tr>
<td>Motif 3. RM</td>
<td>CGTCACAAATTGAAAAAAGACCTTTAATA</td>
</tr>
<tr>
<td>Motif 3. FM</td>
<td>GATACGCACATTAAAAAAAGCATGTGTGCA</td>
</tr>
<tr>
<td>Motif 1. FLIP. F2</td>
<td>ACTGCTCGAGAATGCTAGCTGTCGCCGGGTGC</td>
</tr>
<tr>
<td>Motif 1. FLIP. R2</td>
<td>ACTGACAAGCTTGCGAGAATGACAGATGGACAG</td>
</tr>
<tr>
<td>Motif 1. FLIP. RM</td>
<td>GAAATGCAGGAAACGCTCTCGCGCAAGCAAGCCG</td>
</tr>
<tr>
<td>Motif 1. FLIP. FM</td>
<td>CGGCTTGCTGGCGAGAAGCCGTTCCTGCATTTC</td>
</tr>
</tbody>
</table>
Figure 3.1. Schematic representation of 5’ flanking regions of *S. neurona* genes. Sequence analysis revealed GCGTCTC motifs in either orientation, at variable locations and differing copy number. Direct orientation of the motif (GCGTCTC) is shown by hatched box and indirect orientation (GAGACGC) is represented by open box. Transcription start sites are shown by bent arrows. The transcription start sites of *SnSAG1* were identified by 5’ RACE. The transcription start sites for the four other *S. neurona* genes are putative based on the first nucleotide in *S. neurona* EST database.
Figure 3.2. Examination of reporter activity provided by SnSAG1 5’ and 3’ flanking regions. The analysis indicated that both 5’ and 3’ flanking regions are needed for efficient transcription of the reporter gene. Mean ± SEM were determined from three independent experiments.
Figure 3.3A. Schematic representation of *SnSAG1* full length promoter region used in functional analysis. This included 701 bases of 5’ flanking region and 85 bases of 5’ UTR. There is one indirect copy, motif 1 (M1) and two direct copies, motif 2 (M2) & motif 3 (M3) of GCGTCTC motifs in the 5’ flanking region. The two transcription start sites are shown by bent arrows. The transcription start site 88 bases upstream of ATG start codon was designated as first transcription start site (+1). The numbers below the motifs indicate distance from first transcription start site.

Figure 3.3B. *SnSAG1* promoter mapping by 5’ deletion analysis. Transient luciferase activity of deletion mutants indicated that the cis-acting element/s required for FLUC expression is/are located between 179 and 101 bases upstream of the transcription start site. Mean ± SEM were obtained from three independent experiments.
Figure 3.4. *SnSAG1* promoter mapping by 5’ deletion analysis. Based on results of the previous experiment (Fig 3B), the region between -101 and -179 was subjected to further deletion analysis. The results indicated that 26 bp region between -149 and -123 contains sequence element essential for FLUC expression. Mean ± SEM were obtained from three independent experiments.
Figure 3.5. Fine mapping of *SnSAG1* promoter by 5’ deletion and mutational analysis. The results indicate that GAGACGC motif located between -129 and -136 is essential for FLUC expression. Dot filled boxes in constructs in -179\textsuperscript{M1}, -136\textsuperscript{M2} and -136\textsuperscript{M3} represent mutation of motifs 1, 2 and 3 respectively. Zigzag line in construct -136\textsuperscript{R} represents replacement of *SnSAG1* promoter sequence between motif 1 and motif 2 by heterologous sequence of same length. Mean ± SEM were obtained from three independent experiments.
Figure 3.6. *SnSAG1* promoter mapping by mutational analysis. Orientation is important for functioning of motif 1 and the *SnSAG1* promoter can function at a reduced capacity in reverse orientation. Motif 1 is placed in reverse orientation in construct -179\(^{M1.FLIP}\). In construct -223\(^{FLIP}\) the *SnSAG1* promoter fragment (-223 to +85) is placed in reverse orientation. Open and closed block arrows in constructs -223 & -223\(^{FLIP}\) indicate the orientation of *SnSAG1* promoter fragment. Mean ± SEM were determined from three independent experiments.
Figure 3.7A. *T. gondii SAG1* promoter is not well recognized by *S. neurona* transcription machinery. *S. neurona* merozoites were transfected independently with pSnSAG1/FLU/SnSAG1 and pTgSAG1/FLUC/SnSAG1. pSnSAG1/RLUC/SnSAG2 was used as internal control for transfection efficiency. The reporter activity was determined 72 hours post-transfection and luciferase activity obtained with SnSAG1 promoter was set at 100%. Mean ± SEM were determined from three independent experiments.

Figure 3.7B. *SnSAG1* promoter is not well recognized by *T. gondii* transcription machinery. *T. gondii* (RH-1) tachyzoites were transfected independently with pSnSAG1/FLU/SnSAG1 & pTgSAG1/FLUC/SnSAG1. pTgSAG1/RLUC/SnSAG2 was used as internal control for transfection efficiency. The reporter activity was determined 24 hours post-transfection and luciferase activity obtained with TgSAG1 promoter was set at 100%. Mean ± SEM were determined from three independent experiments.
Figure 3.8. Electrophoresis mobility shift assay indicated that M1 is important for binding to proteins in *S. neurona* nuclear extract. The assay was performed using three labeled probes - probe 1 (Lanes 1,2,3) was 21 bp in length with wild-type copy of M1 (straight line); probe 2 (Lanes 4,5,6) was same as probe 1, but contained mutant copy of M1; probe 3 (Lanes 7,8,9,10) was 254 bp in length that included all three motifs (dotted line). In lane 3, non-labeled probe 1 was the competitor; in lane 6, non-labeled probe 2 was used as the competitor; in lane 9, non-labeled probe 3 was used as the competitor and in lane 10, non-labeled probe 1 was used as the competitor. Arrows indicate the location of the shifted probe.
Chapter Four: Conclusions

*S. neurona* is an obligatory intracellular parasite classified in phylum Apicomplexa and is a major cause of EPM in horses. The biology of this parasite with regard to host cell invasion, pathogenesis, metabolism, gene regulation etc has not been defined partly due to unavailability of molecular genetic tools. The parasite divides by a complex process termed endopolygeny after invading a host cell. During this process many genes are developmentally regulated. It is important to identify regulatory sequences controlling gene expression to understand gene regulation. Hence the primary objective of this study was to identify regulatory elements that play a role in gene expression in *S. neurona*. However, to conduct this study, it was essential to establish transfection system for this parasite. One of the basic components needed was promoters of *S. neurona* genes to drive expression of transgenes. To obtain promoters of *S. neurona* genes, a cosmid based *S. neurona* genomic library was made and some of the highly expressed gene clones were isolated. The 5’ flanking region of these gene clones were sequenced to obtain presumptive promoter elements that were cloned upstream of reporter genes. Transfection of reporter plasmid constructs into *S. neurona* resulted in successful expression of transgenes YFP & β-gal. An important molecular genetic tool is the establishment of stable transformation system for an organism. Mutant DHFR-TS allele of *T. gondii* has proven to be a strong selectable marker in related apicomplexans, *T. gondii* and *Plasmodium* spp. The expression of this gene confers resistance against the drug pyrimethamine, a folate inhibitor. Stable transformation of *S. neurona* was achieved using mutant DHFR-TS. Using this selection system transgenic *S. neurona* stably expressing reporter molecules YFP & β-gal were generated and were shown to be useful for assessing drug sensitivity *in-vitro*. Theses transgenic parasites
can be used for primary screening and identification of potential drug candidates against *S. neurona*.

After successfully establishing transfection system for *S. neurona*, SnSAG1 gene promoter was subjected to comparative and functional analysis. Comparative analysis of 5’ flanking regions of *S. neurona* genes, SnSAG1, SnSAG2, SnSAG3, SnNTP1 and Sarco_21 revealed a 7-base conserved motif GCGTCTC. Dual luciferase system was used for functional analysis of SnSAG1 promoter. Nested 5’ deletion and mutational analysis indicated that GAGACGC motif located between -136 and -129 is an essential cis-acting element for SnSAG1 expression. This is a protein binding motif that functions in an orientation dependent manner. The GAGACG motif of SnSAG1 is identical to WGAGACG element that has been shown to play a role in transcription of many *T. gondii* genes. This motif is also present in *N. caninum* gene promoters suggesting that GAGACGC is a conserved core promoter element in tissue dwelling coccidia.

The results of this study indicate that there are additional regulatory elements between motif 1 and 2 that are essential for transcription. Mutational analysis of the region between these two motifs will allow for identification of these regulatory sequences. The 5’ UTR of SnSAG1 can also be subjected to mutational analysis to see if there are core promoter elements similar to down-stream promoter element (DPE) in higher eukaryotes. The DNA foot-print analysis is also another important tool to identify additional protein binding motifs in the SnSAG1 promoter region. The protein that interacts with motif 1 can be identified by mass spectrometry. The proteins that interact with GAGACGC binding protein can be characterized using yeast two-hybrid system.
The molecular pathogenesis of *S. neurona* is a very poorly understood process. The route by which the *S. neurona* released in the intestine of the horse reach the central nervous system is not known. Using stable transformation system transgenic parasites stably expressing reporter molecule firefly luciferase can be created and the path by which *S. neurona* travels in the body of the horse can be worked out. It is also feasible to perform genetic manipulations such as gene knock-outs to precisely define the function of different proteins and hence virulent proteins involved in molecular pathogenesis of *S. neurona* can be identified. The prolonged intracellular development and the availability of molecular genetic tools & cellular markers make *S. neurona* a useful model to study developmental biology of apicomplexan parasites. Identification of cis-acting elements and transcription factors that interact with these sequences will help in understanding the mechanism of transcriptional and developmental regulation in *S. neurona*. Those transcription factors that play a role in expression of virulent genes in *S. neurona* can be ideal drug targets for treating EPM.
REFERENCES


80. Lindsay, D.S. and J.P. Dubey, Determination of the activity of pyrimethamine, trimethoprim, sulfonamides, and combinations of pyrimethamine and


Vita

Date of Birth:

Sep 6, 1973

Education:

1996 - 1998        M.V.Sc (Animal Biotech), Madras Veterinary College, India
1991 - 1996        B.V.Sc, University of Agriculture Sciences, Dharwar, India

Professional Positions:

1999 - 2001        Veterinary Officer, Government of Karnataka, India

Honors/Scholarships:

1. Paul Mellon Graduate Student Fellowship in Equine Veterinary Science, Fall 2001- present.

2. University of Kentucky, Graduate School Travel Grant, April 2005.


4. Kentucky Opportunity Fellowship, Fall 2002 - Fall 2003, University of Kentucky.


Publications:

Howe, D.K., Gaji, R.Y., Mroz-Brarrett, M., Gubbels M. J., Striepen, B and S. Stamper, 2005. *Sarcocystis neurona* merozoites express a family of immunogenic surface antigens that are orthologues of the *Toxoplasma gondii* surface antigens (SAGs) and SAG-related sequences. *Infection and Immunity*, 73 (1023-33).


Abstracts/Presentations:


