

University of Windsor

Scholarship at UWindsor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

2019

Investigating Natural Antisense Transcripts in *Toxoplasma gondii*

Ambreen Fahim
University of Windsor

Follow this and additional works at: <https://scholar.uwindsor.ca/etd>

Recommended Citation

Fahim, Ambreen, "Investigating Natural Antisense Transcripts in *Toxoplasma gondii*" (2019). *Electronic Theses and Dissertations*. 7701.
<https://scholar.uwindsor.ca/etd/7701>

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

Investigating Natural Antisense Transcripts in *Toxoplasma gondii*

By

Ambreen Fahim

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Chemistry and Biochemistry
in Partial Fulfillment of the Requirements for
the Degree of Master of Science
at the University of Windsor

Windsor, Ontario, Canada

2019

© 2019 Ambreen Fahim

Investigating Natural Antisense Transcripts in *Toxoplasma gondii*

by

Ambreen Fahim

APPROVED BY:

P. Karpowicz
Department of Biological Sciences

O. Vacratsis
Department of Chemistry & Biochemistry

S. Ananvoranich, Advisor
Department of Chemistry & Biochemistry

May 17, 2019

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Natural antisense transcripts (NATs) are non-coding RNAs that can regulate the expression of their counterpart protein-coding transcript. While NATs are widespread in eukaryotic genomes, very little is known about their mechanism. Our study focuses on gaining a better understanding of the function of NATs in *Toxoplasma gondii*, a pathogenic unicellular eukaryote. We recently characterized the gene encoding the first committed enzyme in SUMOylation, named Ubiquitin-like protease 1 (TgUlp1), and showed that the expression of *TgUlp1* is vital to the life cycle of *T. gondii*. Interestingly, the locus of *TgUlp1* also transcribes a NAT species. Using a dual luciferase assay and RT-qPCR, we identified the promoters of *TgUlp1* mRNA and NAT and measured their transcript levels in tachyzoites and bradyzoites. We found that *TgUlp1* mRNA and NAT are differentially regulated at the transcriptional level *via* promoter activity and transcript turnover. Furthermore, the products of *TgUlp1* NAT processed by RNase III retain the ability to lower the expression of reporters carrying *TgUlp1* mRNA sequences, suggesting the involvement of RNA interference pathway. In Dicer-knockout (TgDicer-KO) and Argonaute-knockout (TgAgo-KO) transgenic strains, a higher level of *TgUlp1* NAT, and much lower level of *TgUlp1* mRNA was detected, suggesting that Dicer and Ago may be involved in maintaining *TgUlp1* mRNA. Although we were unable to determine the direct effect of *TgUlp1* NAT on mRNA, we showed that the introduction of *TgUlp1* NAT by electroporation negatively affected the level of *TgUlp1* mRNA. Consequently, regulation by *TgUlp1* NAT would also affect TgUlp1 protein levels and ultimately the SUMOylation pathway in *T. gondii* which plays an important role in host cell invasion and cyst genesis. However, underlying mechanisms remain to be investigated.

DEDICATION

To Ammi, Papa, Ahsan, and Hamza

Words could never express how much your love and support mean to me

ACKNOWLEDGEMENTS

In the name of Allah, Most Gracious, Most Merciful. My religion is my anchor and motivation which helps me in everything I do in my life.

First and foremost, I would like to thank my supervisor, Dr. Sirinart Ananvoranich. You allowed me into your lab four years ago and have guided me throughout the entire time I have known you. I appreciate all the time and effort you have spent with me, the motivation and confidence and all the opportunities that you provided me with. You have encouraged me to become a better student and researcher which I will always remember. I would also like to thank my committee, Dr. Vacratsis and Dr. Karpowicz, for their advice and suggestions. Also thank you to the rest of the Chemistry and Biochemistry Department.

I would like to thank to past and present members in the Ananvoranich lab, Scott Roscoe, Anna Crater-Potter, Farzana Afrin, Mikayla Roberts, Grace Wen, and Yasin Avci. Each of you has made this a great experience and end to my undergraduate and graduate years.

I want to thank my friends, Labiba, Hafsah, Raghad, Sara and Rana. You guys have no idea what my work really is about, but you have been unbelievably supportive and continued to motivate me. Thank you for venting sessions, board games, and the stress-relieving hangouts.

And finally, none of this is possible without my family who encourage, support and love me unconditionally. My parents have done nothing but encourage me to do the best I can in whatever I want to do, and my brothers who always have my back. Love you guys.

TABLE OF CONTENTS

DECLARATION OF ORIGINALITY	III
ABSTRACT.....	IV
DEDICATION.....	V
ACKNOWLEDGEMENTS	VI
LIST OF TABLES	IX
LIST OF FIGURES	X
LIST OF APPENDICES	XI
LIST OF ABBREVIATIONS	XII
CHAPTER 1 – LITERATURE REVIEW	1
1.1 Noncoding RNA.....	1
1.2 Classification of ncRNA	1
1.3 Natural Antisense Transcripts	3
1.4 <i>Cis</i> -NATs	3
1.5 Regulatory functions of <i>Cis</i> -NATs	5
1.5.1 Transcriptional Interference	5
1.5.2 Chromatin Modification	8
1.5.3 RNA Editing.....	8
1.5.4 RNA Masking.....	9
1.6 <i>Trans</i> -NATs	10
1.7 ncRNA and RNA Interference	10
1.8 NATs in Apicomplexa	11
1.9 <i>Toxoplasma gondii</i> , the Model Apicomplexan	12
1.10 Discovery of <i>Toxoplasma gondii</i>	12
1.10.1 Life cycle of <i>T. gondii</i>	13
1.10.2 Asexual Life Cycle	15
1.10.3 Toxoplasmosis and Treatment.....	17
1.10.4 <i>T. gondii</i> Clonal Lineages.....	17
1.11 Post-Translational Regulation by SUMOylation	18
1.11.1 SUMOylation in <i>T. gondii</i>	21
1.11.2 Ubiquitin-like protease 1 in <i>T. gondii</i>	21
1.12 Research Hypotheses and Objectives.....	22
CHAPTER 2 – MATERIALS AND METHODS.....	24
2.1 Cell and Parasite Culture.....	24

2.1.1	Mammalian cell culture	24
2.1.2	<i>Toxoplasma gondii</i> culture	24
2.2	Construction of Plasmids for Promoter Reporter Assays.....	24
2.3	<i>In vitro</i> transcription of <i>TgUlp1</i> NAT	25
2.3.1	Generating Template with T7 Promoter	25
2.3.2	<i>In vitro</i> transcription of NAT	26
2.3.3	Preparation of Small RNA.....	26
2.3.4	PAGE Analysis.....	26
2.4	Electroporation	26
2.5	Dual Luciferase assay.....	27
2.6	RT-PCR Analysis.....	28
2.6.1	RNA Isolation.....	28
2.6.2	Reverse Transcriptase.....	28
2.6.3	RT-qPCR.....	28
2.7	Immunofluorescence Assay	29
CHAPTER 3 – RESULTS.....		30
3.1	Establishing a Dual Luciferase Reporter to Study Promoters.....	30
3.2	Identifying <i>TgUlp1</i> mRNA and NAT Promoters	36
3.3	<i>TgUlp1</i> mRNA and NAT Promoter Activity in Bradyzoites	38
3.4	<i>TgUlp1</i> mRNA and NAT Expression	42
3.5	The Effect of <i>TgUlp1</i> NAT on Gene Expression.....	43
3.5.1	Structural Analysis of <i>TgUlp1</i> NAT.....	43
3.5.2	NAT and its Derivatives on the Expression of <i>RnLuc</i> Transcripts	47
3.6	Effect of <i>TgUlp1</i> NAT on the level of <i>TgUlp1</i> mRNA.....	50
3.7	<i>TgUlp1</i> mRNA and NAT Expression in TgDicer-KO and TgAgo-KO	54
CHAPTER 4 – DISCUSSION.....		55
4.1	Promoters for <i>TgUlp1</i> mRNA and NAT	55
4.2	Expression Of <i>TgUlp1</i> mRNA And NAT During Its Asexual Cycle	58
4.3	<i>TgUlp1</i> NAT and its Regulatory Effect	60
4.4	Dicer and Ago Knockout Strains of <i>Toxoplasma gondii</i>	62
CONCLUSION		65
REFERENCES.....		67
APPENDICES.....		75
VITA AUCTORIS		81

LIST OF TABLES

Table 3.1	RnLuc Activity of Control Plasmids for Promoter Assay.....	34
Table 3.2	Stem Loops Predicted to be Cleaved by Dicer.....	45

LIST OF FIGURES

Figure 1.1	Classification of Noncoding RNAs by Their Origin	2
Figure 1.2	Classification of <i>Cis</i> -NATs Based on Their Orientation Relative to Their Target Gene.....	4
Figure 1.3	Mechanisms of Transcriptional Interference	7
Figure 1.4	Life Cycle of <i>Toxoplasma gondii</i> in Definitive and Intermediate Hosts	14
Figure 1.5	<i>T. gondii</i> Asexual Life Cycle Within A Host Cell	16
Figure 1.6	The SUMOylation pathway.....	20
Figure 1.7	<i>TgUlp1</i> Locus	22
Figure 3.1	Putative Regions Tested for Promoter Activity.....	31
Figure 3.2	Representative Gel Image of PCR Analysis of Putative Promoters ...	32
Figure 3.3	Plasmid Constructs Used in the Promoter Assays.....	33
Figure 3.4	Relative Luciferase Activity of Putative 2.5-kb <i>TgUlp1</i> mRNA promoter Compared to Strong Tubulin Promoter.....	35
Figure 3.5	Promoter Activity in Type I and Type II <i>T. gondii</i> Tachyzoites	37
Figure 3.6	Immunofluorescence Assay of Cyst Formation in Type I and II <i>T. gondii</i>	39
Figure 3.7	Relative activity of RnLuc Under the Control of Putative Promoters in Type I <i>T. gondii</i> Bradyzoites.....	40
Figure 3.8	Change in Sense and Antisense Promoter Activity in Type II Bradyzoites	41
Figure 3.9	Expression of <i>TgUlp1</i> mRNA and NAT in Tachyzoites and Bradyzoites	42
Figure 3.10	<i>TgUlp1</i> NAT Structure Prediction.....	44
Figure 3.11	Predicted Binding Sites on <i>TgUlp1</i> mRNA Used To Detect Self-Regulation By <i>TgUlp1</i> NAT.....	48
Figure 3.12	Effect of <i>TgUlp1</i> NAT on Engineered RnLuc Expression.....	49
Figure 3.13	Experiment Flow for Detecting Effect of <i>in vitro</i> NAT	51
Figure 3.14	<i>TgUlp1</i> mRNA Expression After Mock and <i>in vitro</i> NAT Electroporation	52
Figure 3.15	<i>TgUlp1</i> mRNA and NAT Expression After Mock Electroporation ...	53
Figure 3.16	<i>TgUlp1</i> mRNA and NAT expression in TgDicer-KO and TgAgo-KO compared to its' parental.....	54
Figure 4.1	Identified Promoters of <i>TgUlp1</i> mRNA and NAT	57

LIST OF APPENDICES

APPENDIX A	75
Figure A1 pTUBRnLuc	75
Figure A2 pRnLuc	76
Figure A3 Diagram Showing Primers Used for <i>TgUlp1</i> RT-qPCR.	76
APPENDIX B	77
Table B1 Primers Used to Construct Plasmids For The Promoter Assay	77
Table B2 pPutPromRnLuc Used in the Promoter Assay	77
Table B3 Primers Used to Make Templates For <i>in Vitro</i> RNA	78
Table B4 Primers Used for RT-qPCR.....	78
Table B5 Amount of Plasmid Electroporated Based on Experiment.....	78
Table B6 Initiator Element Identified in <i>TgUlp1</i> Promoters.....	79
Table B7 GAGACG Motif Identified in <i>TgUlp1</i> Promoters.....	80
Table B8 TGCATGC Motif Identified in <i>TgUlp1</i> Promoters.....	80

LIST OF ABBREVIATIONS

-R	RNase III Digested
Ago	Argonaute
AS	Antisense Strand
Bz	Bradyzoites
cDNA	complementary DNA
CDS	Coding Sequence
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
Dol-FTC	FITC-conjugated <i>Dolichos biflorus</i> lectin
ds	Double stranded
DTT	Dithiothreitol
EP	Electroporation
FFLuc	<i>Firefly</i> luciferase
HFF	Human Foreskin Fibroblasts
hr	Hour
Hz	Hoechst
kDa	Kilodalton
KO	Knockout
lncRNA	Long ncRNA
miRNA	microRNA
mRNA	messenger RNA
NAT	Natural antisense transcript
ncRNA	non-coding RNA
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PV	Parasitophorous Vacuole
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNAP	RNA Polymerase II

RnLuc	<i>Renilla</i> luciferase
RT-qPCR	Reverse Transcriptase Quantitative PCR
SAGE	Serial Analysis of Gene Expression
siRNA	short interfering RNA
sRNA	short RNA
ss	single stranded
SS	Sense Strand
SUMO	Small ubiquitin-like modifier
Tg	<i>Toxoplasma gondii</i>
TSS	Transcription start site
TUB	Tubulin
Tz	Tachyzoites
UTR	Untranslated region
Ulp1	Ubiquitin-like protease 1

CHAPTER 1 – LITERATURE REVIEW

1.1 Noncoding RNA

Whole transcriptome analysis has revealed that up to 90% of eukaryotic genomes are transcribed [1]. However, only 1–2% of the transcripts are protein coding. Majority of the remaining ~98% of transcripts are functional non-protein-coding RNA (ncRNA) molecules. ncRNAs play an important role in cellular processes, including gene imprinting, differentiation and development, antiviral response, apoptosis, cell cycle control and more [2, 3].

1.2 Classification of ncRNA

ncRNA can be classified by their length. An arbitrary 200-nucleotide threshold is used to separate short and long ncRNA. To date, the best studied short ncRNA are microRNA (miRNA) and small interference RNA (siRNA), which regulate gene expression through the RNA interference (RNAi) pathway. Long ncRNAs (lncRNAs) are larger than 200 nts and can be over 100,000 nts in length. LncRNAs make up the largest portion of ncRNA and are highly diverse in structure and function [3]. LncRNAs are detected in the genomes of animals, plants, yeast, prokaryotes and even viruses. However, they are poorly conserved when compared to well-studied short ncRNAs like miRNAs or siRNAs [4].

LncRNAs are further classified based on their origin within the genome, including their location and orientation (Figure 1.1). Intergenic lncRNA are transcribed from the location between two genes, regardless of their orientation. Intronic lncRNAs are transcribed entirely from introns of protein-coding genes. Sense lncRNAs are transcribed from the sense strand of protein-coding genes and overlap the exons. Antisense ncRNAs

are transcribed from the antisense strand of protein-coding genes, and also referred to as **natural antisense transcripts (NAT)**.

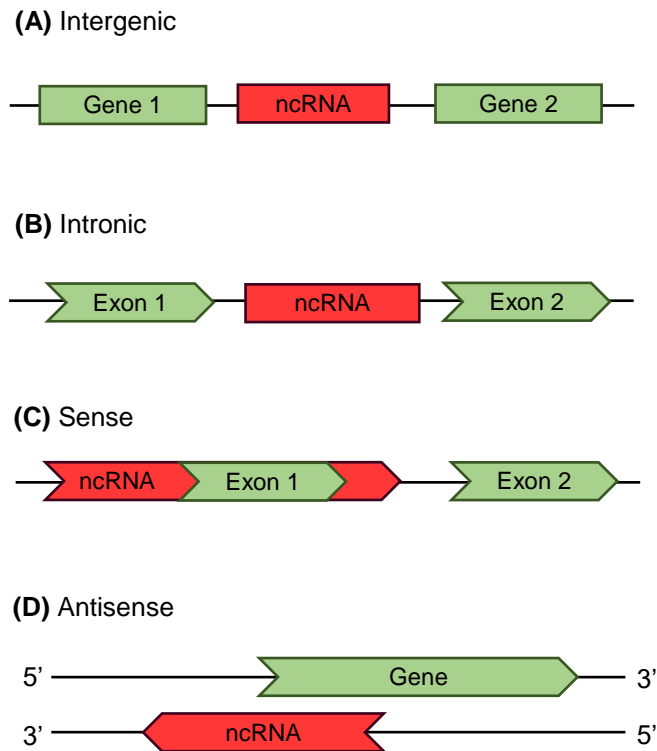


Figure 1.1 Classification of Noncoding RNAs by Their Origin

ncRNAs can be classified into four types according to their position in the genome. Green bars/arrows represent the transcription region and direction of sense transcript and the red bars/arrows represent the transcription region and direction of the ncRNA. Modified from [3].

(A) **Intergenic ncRNAs** are transcribed from the genomic sequences located between two genes, regardless of their orientation.

(B) **Intronic ncRNAs** are derived entirely from the introns of protein encoding genes.

(C) **Sense lncRNAs** are transcribed from the sense strand of protein-coding genes and overlap with exons of the gene.

(D) **Antisense lncRNAs** are transcribed from the antisense strand of protein-coding genes.

1.3 Natural Antisense Transcripts

NATs are widespread in eukaryotes and are recognized as important regulators of gene expression. Up to 72% of transcriptional units in mice and 22 - 40% in human (depending on cell type) are transcribed in both orientations, and comparable numbers have been suggested for other eukaryotes [5, 6]. Similar to mRNA, NATs are capped, polyadenylated, and spliced [3]. Their transcription is also controlled by promoters and enhancers and transcribed by RNA polymerase II. Unlike mRNAs, NATs preferentially accumulate in the nucleus, in addition to other cellular compartments, such as mitochondria [7]. NAT can be broadly grouped into two categories based on the location of their target as *cis*- and *trans*-NATs.

1.4 *Cis*-NATs

Cis- NATs have been extensively studied in many eukaryotes and found to play a variety of regulatory roles. They are transcribed from the same genomic locus as their target but from the opposite DNA strand. Therefore, *cis*-NATs have perfect base-pairing to their target transcript. This RNA-RNA interaction can result in translational inhibition, mRNA degradation, or promote RNA stability [8].

There are three types of *cis*-NATs based on their position relative to the position of their target gene (Figure 1.2). In head-to-head orientation, the sense and antisense transcripts overlap on their 5'-end. In addition, the 5'-UTR of the sense gene may harbour a bidirectional promoter capable of initiating transcription for both sense and antisense transcripts. Tail-to-tail describes transcripts overlapping at the 3'-ends. In full overlap, NAT completely overlaps with the sense transcript. The transcription starts site (TSS) of

NAT occurs within the sense gene, resulting in high level of complementarity between the sense and antisense transcript.

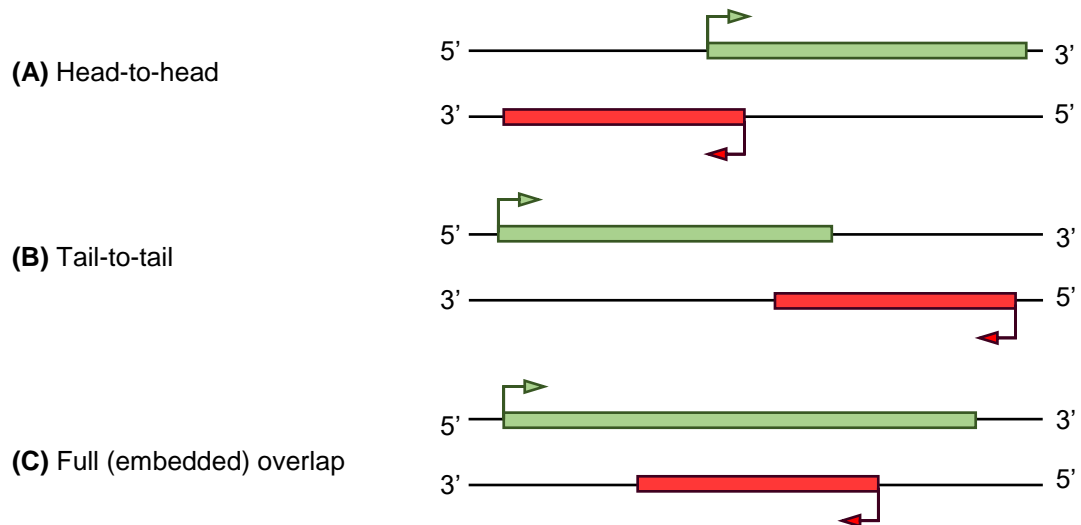


Figure 1.2 Classification of *Cis*-NATs Based on Their Orientation Relative to Their Target Gene

Cis-NATs can be classified into three types according to their positions relative to the sense counterpart. Green bars represent the transcription region of the sense transcript, and the red bars represent the transcription region of the antisense transcript. Arrows represent transcription start site (TSS) and direction of transcription. Modified from [3].

(A) Head-to-head: the sense and antisense transcripts overlap on their 5'-ends.

(B) Tail-to-tail: sense and antisense transcripts overlap on their 3'-ends.

(C) In a full overlap (or embedded overlap), the NAT transcription region is totally included within the sense transcription region.

1.5 Regulatory functions of *Cis*-NATs

Cis-NATs exert their activity through various mechanisms including transcription interference, chromatin modification, RNA editing, and RNA masking.

1.5.1 Transcriptional Interference

Transcriptional interference refers to the suppression of one transcriptional process by the direct influence of a second transcriptional process. The close proximity of the *cis*-NAT transcriptional unit to the target transcriptional unit makes it possible for gene regulation to occur without the direct pairing of the RNA molecules themselves. Four mechanisms of transcriptional interference that affect sense and antisense gene pairs have been hypothesized (Figure 1.3) [9].

(i) Promoter Competition: A sense and antisense gene pair in a head-to-head orientation will share a bidirectional promoter that initiates transcription in both orientations. During the initiation phase, steric hindrance and competition for the binding of RNA Polymerase II (RNAP) and other transcriptional elements can occur for bidirectional promoter. This restricts the RNAP binding to one promoter, enhancing its' activity and leads to the downregulation of the other transcript. Over 7000 sense and antisense transcript pairs were identified in the human genome and found that 76% exhibit a head-to-head formation, sharing a bidirectional promoter, and within this 58% of NATs begin from a 500 bps region upstream of the TSS of the sense gene [10].

(ii) Sitting Duck Mechanism: An RNAP complex that is too slow to transition from open to an elongation complex is considered a 'sitting duck', which can be dislodged by an elongating RNAP from a different promoter downstream.

(iii) Occlusion: Occlusion occurs when an elongating RNAP is passing through a promoter, therefore blocking the binding of another RNAP. However, transcription elongation occurs rapidly in most organisms, meaning that occlusion is very brief, so even extremely strong interfering promoters will not produce much occlusion by elongating RNAP [11].

(iv) RNAP Collision: When both sense and antisense transcription has already initiated, the collision of both RNAP complexes during elongation phase can lead to the premature termination of one of the complexes. As elongating RNAP envelops both strands of DNA, it is likely that both RNAP cannot continue transcription past one another. RNAP collisions have been imaged by atomic force microscopy, showing that one elongating RNAP forces the opposing RNAP to stall and backtrack [11].

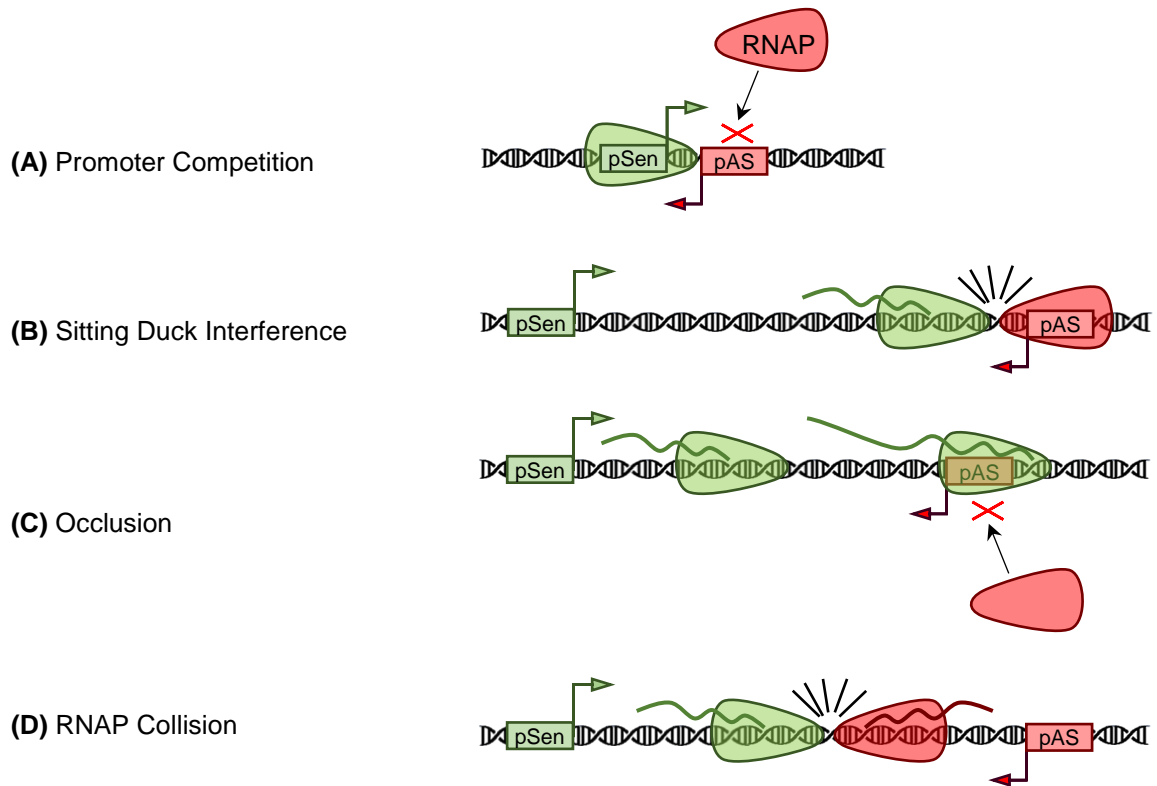


Figure 1.3 Mechanisms of Transcriptional Interference

Promoters of sense genes are represented as green boxes (pSen), and promoters of antisense genes are represented red boxes (pAS). Green and red arrows represent direction of transcription for sense and antisense transcription respectively. Rounded triangles represent RNA Polymerase II (RNAP). Modified from [9].

(A) Promoter Competition: In the initiation phase, promoters of head-to-head NATs compete for the RNAP and regulatory elements.

(B) Sitting Duck Mechanism: RNAP that is too slow to start is dislodged by an elongating RNAP.

(C) Occlusion: Elongating RNAP is passing through a promoter, therefore blocking the binding of another RNAP.

(D) RNAP Collision: In the elongation phase, collision between RNAP can lead to a transcriptional interference.

1.5.2 Chromatin Modification

Cis-NATs can regulate gene expression through epigenetic mechanisms. A classic example is X chromosome inactivation by X-inactive specific transcript (XIST). XIST is expressed from a region of chromosome X called the X inactivation center (XIC). XIST transcriptionally silences one of the pairs of X chromosomes in early developmental process in mammalian females to provide dosage equivalence between males and females. XIC also expresses a *cis*-NAT called Tsix, which silences XIST by directing modification of the chromatin structure [12].

Cis-NATs also influence DNA methylation. DNA methylation is a fundamental epigenetic mechanism that regulates gene transcription. It involves the addition (or removal) of methyl group to cytosines that are typically found in CpG-rich regions of genome, referred to as CpG islands. CpG islands are often found at promoters and first exons regions [13]. At unmethylated CpG islands, transcription of the gene is possible; at methylated CpG islands the transcription of the gene is blocked. *Cis*-NATs have been implicated in both methylation and demethylation of DNA to regulate gene expression. For example, the *cis*-NAT of Hemoglobin, alpha 2 (HBA2) mediates methylation CpG island in the promoter for HBA2, resulting in in gene silencing [14]. Conversely, NAT can also be responsible for DNA demethylation to increase transcription. An example is the NAT for the tumor suppressor TCF21, which facilitates demethylation of the TCF21 promoter, leading to its expression [15].

1.5.3 RNA Editing

RNA editing is an enzymatic reaction catalyzed by adenosine deaminases that act on RNA (ADARs) that converts adenosines (A) into inosines (I). ADARs target dsRNA

such as those formed by sense and antisense transcripts [16]. After A-to-I editing, inosine is interpreted as guanine during splicing or translation. Such modification may control the localization or the stability of the edited transcripts. For example, the pre-mRNA of PRUNE2 and its overlapping NAT named PCA3, creates a dsRNA that forms a complex with ADAR proteins. This leads to A-to-I editing of PRUNE2 transcript, resulting in protein downregulation and increase in tumor cell growth [17].

1.5.4 RNA Masking

RNA masking refers to the formation of a sense and antisense dsRNA which provides protection against post-transcriptional regulation factors that target the gene. This provides physical protection by interfering with splicing and translation machineries. For example, the expression of the transcription repressor Zeb2 is dependent on RNA masking done by its NAT. When the 5'-UTR of Zeb2 mRNA is spliced, it disrupts the sequence for the internal ribosome entry site (IRES) required for translation. Therefore, Zeb2 mRNA is no longer able to bind ribosomes and translate the protein. When the Zeb2 NAT sequence overlaps the 5'-UTR, it prevents the binding of the spliceosome to the 5' splice site. Consequently, the IRES is conserved and Zeb2 protein is translated [18].

Conversely, sense and antisense transcript binding may also lead to degradation of a transcript, such as in the case of MALAT1, a nuclear-retained long noncoding RNA that promotes malignancy. The NAT of MALAT1, named TALAM1, binds the 3'-UTR of MALAT1 and allows for cleavage by RNase P. This leads to 3'-end processing and maturation that is essential for MALAT1 stability and function [19]. RNA masking also prevents miRNA and siRNA from binding as well as protect transcripts from RNases which target single stranded RNA [20, 21].

1.6 *Trans*-NATs

Trans-NATs are transcribed from a different location than their targets and could have complementarity to multiple transcripts with mismatches. Although there are very few studies on long *trans*-NATs, *in silico* analysis have identified *trans*-NATs in abundance in many eukaryotes such as humans, mice, plants, zebrafish, flies, and worms [22].

There is various evidence to suggest that *trans*-NATs perform diverse regulatory functions. For example, the 2.2-kb NAT transcribed from the HOXC gene cluster on chromosome 12, called **H**OXC **a**ntisense **i**ntergenic **R**NA (HOTAIR) is capable of modifying histones to regulate gene expression. HOTAIR directly interacts with **P**olycomb **R**epressive **C**omplex **2** (PRC2) with its 5'-end to induce tri-methylation on histone H3, lysine 27 and interacts with **l**ysine-**s**pecific histone **d**emethylase **1**A (LSD1) on its 3'-end to induce demethylation on histone H3, lysine 4. These combined modifications, performed *in trans* on chromosome 2, leads to a repressive chromatin structure, thus silencing of multiple genes [23]. Long *trans*-NATs have also been implicated in mRNA degradation and translational repression [24].

1.7 ncRNA and RNA Interference

RNA interference is a biological process specific to eukaryotes, in which short ncRNA (~22 nts) called siRNA and miRNA inhibit gene expression. Formation of si/miRNA requires dsRNA precursors to be cleaved by Dicer, an RNase III enzyme. Studies have shown that NAT can serve as a dsRNA precursor needed for si/miRNA formation [25, 26]. dsRNA created by internal hairpins loops in NATs can be digested by Dicer to form NAT-miRNA. NAT can form long dsRNA with its sense transcript which is

digested by Dicer to produce NAT-siRNA. After digestion by Dicer, NAT-si/miRNA is bound to Argonaute (Ago) and integrated into the RNA-induced silencing complex (RISC). NAT-si/miRNA guides Ago to target mRNA molecules for silencing or degradation based on complementary sequence [27]. NAT-siRNA forms perfect base pairs with their target mRNA and can act in *cis* to downregulate the sense gene expression [28]. On the other hand, NAT-miRNA forms imperfect base pairs with target mRNA, allowing it to act in *trans* to regulate multiple genes.

1.8 NATs in Apicomplexa

Majority of the studies on NATs have been focused in humans; very little is known about NATs in Apicomplexa, a phylum of eukaryotic unicellular organisms that consist of numerous pathogenic parasites of human and domestic animals. *Plasmodium falciparum*, which causes malaria, is responsible for millions of deaths each year in the developing world. *Eimeria spp.* and *Cryptosporidium spp.* are important enteric pathogens in humans while *Neospora spp.* and *Theileria spp.* are veterinary pathogens. *Toxoplasma gondii* and *Cryptosporidium parvum* have caused water-borne disease outbreaks [29].

In *Plasmodium*, NATs are very frequent and associated with ~24% of all open reading frames [30]. A high frequency of antisense RNAs was also observed in *T. gondii* [31]. Radke et al., [31] used serial analysis of gene expression (SAGE) to examine the *T. gondii* transcriptome and reported ~21% of antisense transcription. Both *Plasmodium* and *T. gondii* also show an inverse relationship between the frequency of antisense transcripts and the level of sense transcription [31]. To date, there is no information regarding NAT regulation, function, or its implications in cell biology in *T. gondii*.

1.9 *Toxoplasma gondii*, the Model Apicomplexan

In contrast to bacterial pathogens, apicomplexan parasites are eukaryotic and share many metabolic pathways with their hosts. This makes therapeutic target development extremely difficult – a drug that harms an apicomplexan parasite is also likely to harm its host. Research on these parasites is challenging because it is difficult to maintain live parasite cultures in the laboratory and to genetically manipulate these organisms. However, *T. gondii* is both easily cultured in the lab and readily amenable to genetic manipulation. This allows researchers to study many biological or biochemical functions of proteins in *T. gondii* that cannot be done in other apicomplexans. Studies are readily performed in *T. gondii* due to the high efficiency of transient and stable transfection and the availability of many cell markers. While results in *T. gondii* might not always be applicable to other Apicomplexa due to differences between the parasites, *T. gondii* remains an important model system for understanding the biology of apicomplexan parasites [32]

1.10 Discovery of *Toxoplasma gondii*

In 1908, Nicolle and Manceaux were studying the tissue of the rodent *Ctenodactylus gundi* and found a parasite they believed to be *Leishmania*. However, they realized they had discovered a new parasite species and named it *Toxoplasma gondii* due to its crescent morphology; toxo meaning “bow” and plasma meaning “creature” [33]. *T. gondii* is an obligate protozoan parasite belonging to the phylum Apicomplexa. Most apicomplexans are characterized by a unique organelle called an apicomplast, which perform essential metabolic functions for the viability of the parasites [34]. They also have complex life cycles alternating between sexual and asexual cycles in different hosts.

T. gondii can infect any nucleated mammalian or avian cell, including humans, causing toxoplasmosis. An estimated 15 to 85% of the world adult human population is chronically infected with *T. gondii*, depending on geographical location [35]. In infected individuals with healthy immune systems, *T. gondii* infection is asymptomatic. In immunocompromised individuals, the infection can result in serious illness and death.

1.10.1 Life cycle of *T. gondii*

T. gondii goes through a sexual cycle within definitive hosts and an asexual cycle within intermediate hosts. The parasite exists in three forms: sporozoites, tachyzoites, and bradyzoites. Oocysts – containing sporozoites – are only produced by sexual reproduction in the definitive host, felines, and passed in feces which are then ingested by intermediate hosts such as humans (Figure 1.4). The asexual cycle continues in the intermediate host.

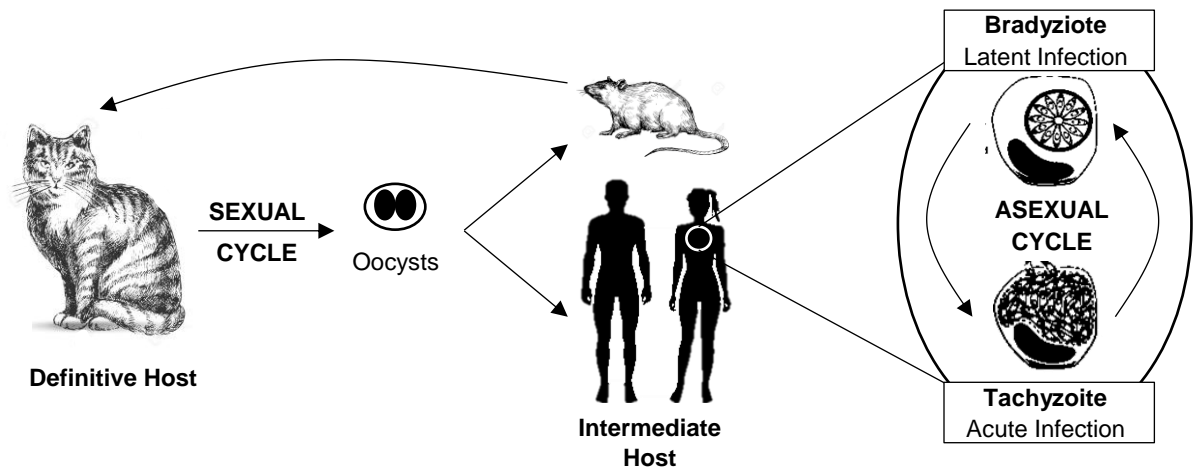


Figure 1.4 Life Cycle of *Toxoplasma gondii* in Definitive and Intermediate Hosts

T. gondii undergoes a sexual cycle in the definitive host to produce oocysts that are excreted through feces into the environment. Oocysts are ingested by intermediate hosts where the parasite undergoes an asexual cycle. Initial infection by tachyzoites is acute. Once the host organism triggers an inflammatory response, tachyzoites are forced to convert to bradyzoites to become latent. When the host organism becomes immunocompromised, bradyzoites convert back to tachyzoites to cause recurrent infection. Modified from [32].

1.10.2 Asexual Life Cycle

Nicolle and Manceaux found *T. gondii* in its tachyzoite forms, also known as its proliferative form. Tachyzoites infect and grow through a process called the lytic cycle (Figure 1.5). First, *T. gondii* infects a host cell by attaching to and penetrating the cell membrane using gliding motility. Once inside the host cell, the parasite is enclosed in the parasitophorous vacuole (PV) in the host cell cytoplasm. The PV is derived from the host cell membrane as well as lipids and proteins from the parasite [36]. Inside the PV, the parasite divides asexually every 6-9 hours through a process called endodyogeny in which two daughter cells are produced within the mother cell, which is then consumed by the offspring prior to their separation [37]. The host cell ruptures when it can no longer support the growth of tachyzoites, and parasites move on to infect the next available cell.

Destruction of host cells by tachyzoites growth triggers the host's immune response. In response to this environmental stress, tachyzoites convert to bradyzoites, also known as the slowly dividing form of the parasite. Bradyzoites are enclosed in tissue cysts that have a glycosylated cyst wall, made up of host and parasite materials that protects the parasite from the host immune system. Tissue cysts are most prevalent in neural and muscular tissues, including the brain, eyes, and skeletal and cardiac muscles [37]. Tissue cysts can persist for the life of the host without triggering an immune response, causing chronic infection. If the host organism becomes immunocompromised, the bradyzoites can differentiate back into the tachyzoite form and can cause damage in the host [37].

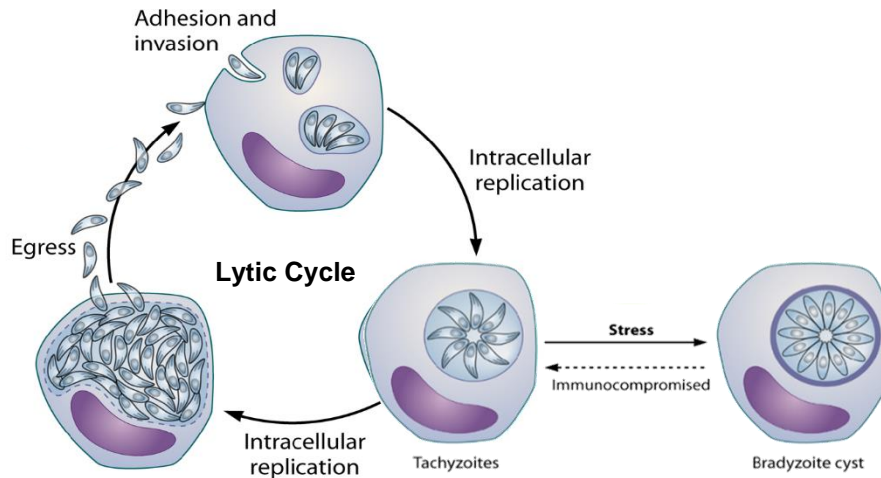


Figure 1.5 *T. gondii* Asexual Life Cycle Within A Host Cell

Asexual cycle in intermediate host cells results in proliferation of *T. gondii* through the lytic cycle to produce tachyzoites. Tachyzoites attach and penetrate through the cell membrane. In the cytoplasm, the tachyzoite is enclosed in a parasitophorous vacuole where it replicates intracellularly by endodyogeny. The host cell ruptures when it can no longer support the growth of tachyzoites within it and parasites move on to infect the next available cell, completing the lytic cycle. Stress can force tachyzoite into bradyzoites, which are enclosed in a tissue cyst. Bradyzoites continue to replicate by endodyogeny but much slower. When the host organism becomes immunocompromised, bradyzoites convert back to tachyzoites. Modified from [38].

1.10.3 Toxoplasmosis and Treatment

Toxoplasmosis can be acquired congenitally and by ingesting infected meat or water contaminated with oocysts, which are shed in the feces of infected cats. Acute toxoplasmosis, caused by tachyzoites in the lytic phase, can manifest with flu-like symptoms including swollen lymph nodes, headaches, fever, fatigue, or muscle aches [39]. Immunocompetent individuals can suppress the acute infection and become latent (asymptomatic). However, once individuals become immunocompromised, they are at risk of recurrent toxoplasmosis which can cause serious illness. This includes damage to the central nervous system causing speech abnormalities, motor deficits, seizures, psychosis, and is potentially fatal if left untreated [40].

Toxoplasmosis is usually asymptomatic; thus, treatment of the infection is limited only to individuals showing signs of acute infection, newly infected pregnant mothers, or preventative treatment for individuals at high risk of recurrent toxoplasmosis [40]. Standard treatment for toxoplasmosis includes a combination of sulfadiazine and pyrimethamine [40]. Despite their effectiveness, these drugs can only target tachyzoites and have no effect on the remaining tissue cysts. They also have serious side effects; pyrimethamine is associated with bone marrow toxicity and sulphadiazine with allergy, making these drugs unsuitable for long term treatment [33].

1.10.4 *T. gondii* Clonal Lineages

Three clonal lineages of *T. gondii* have been identified; type I, type II and type III [41]. Genomic analysis indicates the clonal lineages differ by only 1–2% at the nucleotide level. All three strains can infect their intermediate host orally, bypassing the need for the sexual stage of the parasitic life cycle which limits genetic recombination and the formation

of a wider range of genotypes. Furthermore, infection in intermediate hosts by multiple strains is rare, therefore recombinant genotypes are unlikely to emerge [42].

Although there is no significant difference at the genomic level, studies in mice have shown differences in pathogenesis among the different clonal lineages. Type I strain is highly virulent in comparison to type II and III strains [43]. However, studies in North America and Europe has identified type II strains as the most prevalent cause of human toxoplasmosis in congenital and AIDS patients, while type III is largely confined to animals [44].

1.11 Post-Translational Regulation by SUMOylation

SUMOylation refers to the reversible addition of a small protein named small ubiquitin-like modifier (SUMO) to a protein substrate. SUMO has significant homology to ubiquitin, a highly conserved modifier protein found in all eukaryotic cells. SUMO and ubiquitin share a similar three-dimensional structure, yet less than 20% amino acid sequence similarity [45]. They also share a very similar pathway in conjugating to their protein substrates. However, while ubiquitinated proteins are generally targeted for degradation, SUMOylated proteins are regulated for different functions within a cell. Studies suggest that SUMOylation is involved in many aspects of cell function such as DNA damage repair, maintenance of genome integrity, transcription regulation, and nuclear transport [46]. The SUMO protein is translated as a precursor peptide, around 100 amino acids. It has been evolutionary conserved exclusively in eukaryotes. In higher eukaryotes such as mammals and plants, there can be up to as many as eight SUMO isoforms [47].

SUMOylation is initiated by the C-terminal cleavage of extra amino acids in the precursor SUMO peptide by a protease named Ubiquitin-like protease 1 (Ulp1) (Figure 1.6A). Ulp1 is a cysteine peptidase that is highly specific for the SUMO peptide, as it recognizes the tertiary structure of SUMO rather than an amino acid sequence [48]. Cleavage of the SUMO precursor reveals a di-glycine motif that will ultimately be linked to lysine side chains in target protein substrate. After cleavage, the SUMO protein is activated by the E1 activating enzyme. This involves the formation of adenylated SUMO at the C-terminal carboxyl group of SUMO (Figure 1.6B). The SUMO-AMP bond breaks and is followed by the formation of a thioester bond between C-terminal carboxyl group of SUMO and cysteine residue in E1 activating enzyme (Figure 1.6C) [49]. Next, SUMO is transferred to E2 conjugating enzyme which binds the target protein. SUMO is then conjugated to the target protein by E3 ligase enzyme that recognizes the ψ KxD/E consensus sequence (where ψ is a large hydrophobic residue and x is any residue) [50]. The carboxyl group of the di-glycine motif is ligated to the ψ -amino group of a lysine residue on the substrate (Figure 1.6D). Ulp1 is also responsible for the removal of SUMO from SUMOylated proteins (Figure 1.6E) [48].

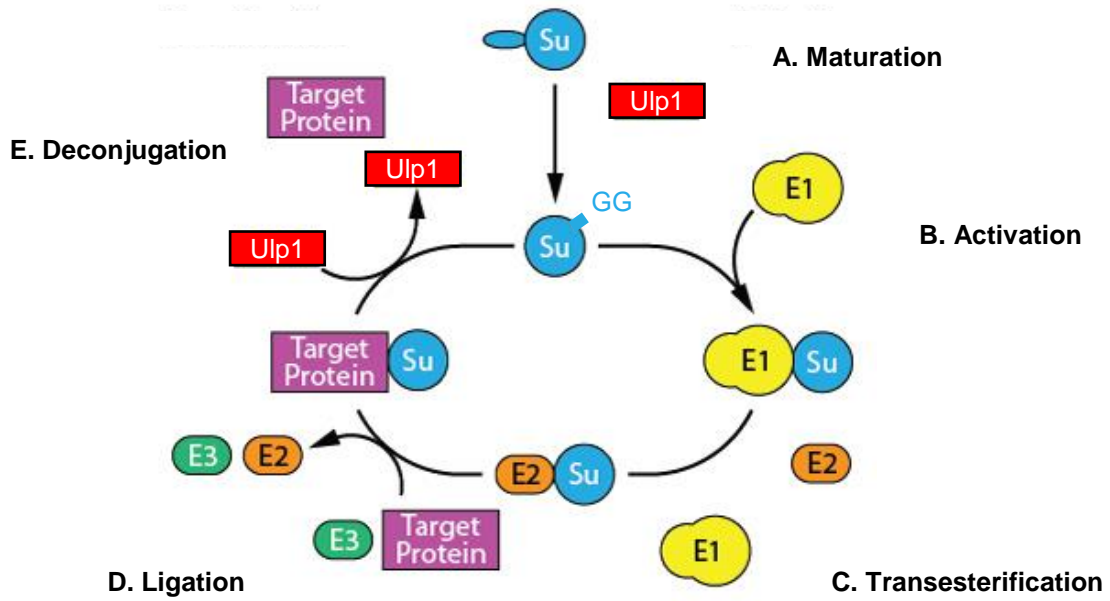


Figure 1.6 The SUMOylation pathway

The SUMOylation pathway is initiated by the cleavage of a SUMO precursor at the C-terminal end to reveal a diglycine motif by Ulp1 (A). The mature SUMO is transferred to an E1 then E2 complex (B, C). E3 ligates SUMO to a protein substrate (D), which recognizes the ψ KxD/E consensus sequence (where ψ is a large hydrophobic residue and x is any residue). Ulp1 also removes SUMO from conjugated protein substrates (E). Modified from [51].

1.11.1 SUMOylation in *T. gondii*

Only one gene is encoded for SUMO in lower eukaryotes, including *T. gondii* and other Apicomplexans [52]. Braun et al., [52] characterized the SUMOylation pathway in *T. gondii*. They determined that the TgSUMO knockout or overexpression by a strong promoter was determined to be lethal. They identified 120 putative SUMOylated proteins. This revealed SUMOylation to be involved in many diverse cellular processes such as metabolism, protein translation and folding, signalling, transport and more. After developing an antibody against recombinant TgSUMO, they used it to localize free SUMO and SUMOylated proteins in the parasite. They observed that TgSUMO was present at the tachyzoite membrane at the point of contact during invasion. Once successfully invaded, TgSUMO was localized to the parasite nucleus. In bradyzoites, they observed TgSUMO at the PV membrane that encapsulates the parasite within the host cell. All of this is indicative of the important role that SUMOylation plays during host cell invasion and survival of the parasite.

1.11.2 Ubiquitin-like protease 1 in *T. gondii*

Crater et al., [53] characterized the *T. gondii* homolog of ubiquitin-like protease and named it TgUlp1. They found that *TgUlp1* is negatively regulated by Tg-miR60, an abundant miRNA species in *T. gondii* type I strain. Misregulation of *TgUlp1* is detrimental to the parasite, indicating its essential role in the parasite's life cycle. Surprisingly, they also discovered a NAT transcribed from the same locus and referred to it as *TgUlp1* NAT. *TgUlp1* NAT was detected by RT-PCR to be transcribed from the intron 6/7 region of *TgUlp1* (Figure 1.7). Due to the complementary sequence between *TgUlp1* mRNA and NAT, it was hypothesized that the expression of *TgUlp1* mRNA is self-regulated by

TgUlp1 NAT. This would be the first example of a self-regulating gene identified in *T. gondii*.

1.12 Research Hypotheses and Objectives

The objective of my study is to gain a better understanding of the mechanism controlling the expression of NAT, using *T. gondii* as a model organism. I aim to identify the elements controlling the expression of *TgUlp1* NAT and elucidate its mechanism of function. The specific aims are as follows:

(I) To identify the promoter controlling the expression of *TgUlp1* mRNA and NAT, I will use a dual luciferase assay. It is hypothesized that the promoter for *TgUlp1* mRNA would lie upstream from exon 1 on the sense strand, and the *TgUlp1* NAT would lie upstream from intron 7 on the antisense strand (Figure 1.7). If putative promoter sequences are capable of initiating transcription, it will successfully drive the expression of RnLuc reporter protein which can be quantified.

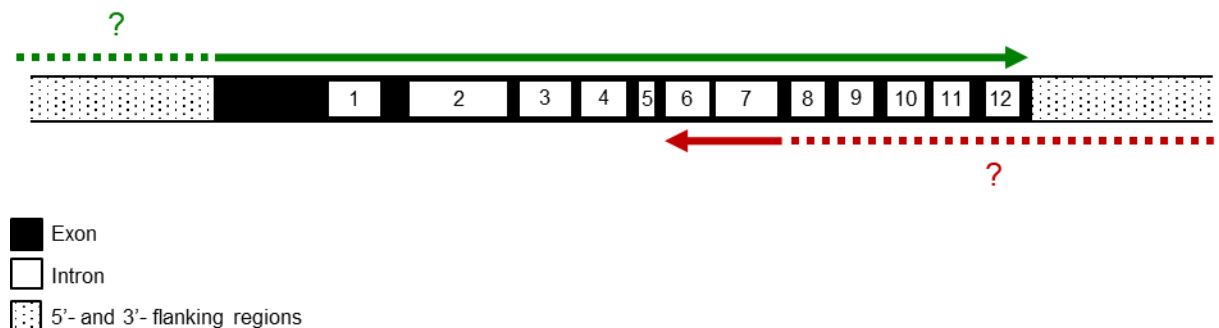


Figure 1.7 *TgUlp1* Locus

TgUlp1 has 13 exons indicated as black boxes and 12 introns indicated as numbered white boxes. Dotted areas are the presumed 5'- and 3'-regions flanking *TgUlp1* CDS. Green arrow represents mRNA transcription direction and TSS. Red arrow represents NAT transcription direction and TSS. Green and red dotted lines indicate the location of putative promoters for mRNA and NAT, respectively.

(II) To characterize the expression pattern of *TgUlp1* mRNA and NAT between tachyzoites and bradyzoites, I will perform RT-qPCR. Many studies have documented genes that are differentially expressed in either tachyzoites or bradyzoites, suggesting that differential gene regulation plays a major role in coordinating stage conversion in *T. gondii*. My study will provide more insight into the role of NATs in gene expression and stage conversion in *T. gondii*.

(III) To determine the regulatory role of *TgUlp1* NAT, I will use a dual luciferase assay. *TgUlp1* mRNA has eleven predicted *Tg-mir60* family binding sites [53]. When three sites were tested with mir60a – a member of *Tg-mir60* family – only one of these sites were downregulated, and the other two were unaffected. These sites might be downregulated by other members of mir60, such as those derived from *TgUlp1* NAT. *TgUlp1* NAT transcription region overlaps from intron 6 to intron 7, making it possible to form dsRNA with the mRNA. *TgUlp1* NAT secondary structure also contains many hairpin structures. Thus, it is hypothesized that both these dsRNA and hairpin structures could serve as precursors of siRNA and miRNA for RNAi.

(IV) To investigate the involvement of key enzymes, Dicer and Ago, of RNAi in the expression of *TgUlp1* transcripts, RT-qPCR will be performed in Dicer knockout (TgDicer-KO) and Ago knockout (TgAgo-KO) strains. The RNase III activity of Dicer and ribonuclease activity of Ago will be required for the processing and function of *TgUlp1* NAT. Therefore, changes in *TgUlp1* mRNA expression in the transgenic strains will implicate RNAi in the function of *TgUlp1* NAT.

CHAPTER 2 – MATERIALS AND METHODS

2.1 Cell and Parasite Culture

2.1.1 Mammalian cell culture

Human foreskin fibroblasts (HFF, ATCC-1041) grown and maintained in Dulbecco's Modified Eagle Medium (DMEM) with high D-glucose and L-glutamine (Invitrogen, #12100046) supplemented with 10% cosmic calf serum (ThermoFisher Scientific, Hylcone, # SH30087.03), and 5 µg/ml antibiotic-antimycotic (Invitrogen, # 15140-122) at 37 °C in 5% CO₂.

2.1.2 *Toxoplasma gondii* culture

Confluent monolayers of HFF were infected with *T. gondii*. The infected monolayer is maintained using Minimum Essential Medium (MEM, Invitrogen, #61100061) supplemented with 1% (v/v) fetal bovine serum (FBS, Sigma) and 5 µg/ml antibiotic-antimycotic (Invitrogen, #15240-062) incubated at 37 °C in 5% CO₂. Type I and Type II strains of *T. gondii* (NIH-AIDS Reagent Program, #2859 and #2858) were maintained in media without selection. TgAgo-KO strain was obtained from Dr. Boothroyd, Stanford University School of Medicine and also maintained in media without selection. TgDicer-KO was created by Farzana Afrin and maintained in selection media containing mycophenolic acid (25 µg/mL) and xanthine (50 µg/mL).

2.2 Construction of Plasmids for Promoter Reporter Assays

Primers used for plasmid construction are listed in Table B1. Putative promoter sequences were amplified by PCR using Type I genomic DNA as template and performed following the manufacturer's protocol (Phire Hot Start II DNA Polymerase, #F122S). The PCR reaction tubes were placed in the 96-well BioRad T100TM Thermal Cycler for 35

cycles. Three overlapping regions upstream from exon 1 were amplified from the sense strand using primers a, b, c and d to give amplicons of 500, 1135 and 2518 bps, respectively. Four overlapping regions downstream from intron 1 were amplified from the antisense strand using primers e, f, g, h, and i to give amplicons of 529, 1167, 2159, and 2670 bps respectively. Amplified fragments were purified by gel extraction using the QiaexII Gel Extraction kit (Qiagen, #20021) following the manufacturer's protocol and used as inserts in ligation reactions. pTUBRnLuc was digested with NheI and HindIII to remove the tubulin (TUB) promoter (2717bps). The linearized pTUBRnLuc plasmid (6071bps) was purified using QiaexII Gel Extraction kit (Qiagen #20021) following the manufacturer's protocol and used as a vector in ligation reactions. The ligation reaction was performed using the Cold Fusion Cloning Kit (System Biosciences, #MC010A-1) or NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, Inc., #E5520S) following the manufacturer's protocol. Resultant plasmids carry putative promoters upstream from RnLuc coding sequence which would drive RnLuc transcription. Plasmids were subjected to restriction endonuclease analyses to confirm their identity before being used in the dual luciferase assay.

2.3 *In vitro* transcription of *TgUlp1* NAT

2.3.1 Generating Template with T7 Promoter

Primers used for plasmids construction are listed in Table B3. PCR was performed using OneTaq® Quick-Load® 2X Master Mix (New England Biolabs, Inc., #M0485S) according to the manufacturer's protocol to generate DNA templates for the synthesis of single- and double-stranded RNA with the sequence encompassing intron 6 and 7 regions of *TgUlp1* (1,167 bps). T7 promoter sequence was incorporated onto the reverse primer

was used to generate single stranded *TgUlp1* NAT (ssNAT). T7 promoter sequence on both primers was used to generate double stranded *TgUlp1* NAT (dsNAT).

2.3.2 *In vitro* transcription of NAT

PCR products were used as a template for T7 transcription. *In vitro* transcription reactions were carried out in 50 μ L reaction mixtures in the presence of T7 RNA polymerase (~5 units), 80 mM HEPES-KOH, pH 7.5, 24 mM NaCl, 2 mM spermidine, 40 mM DTT, 10 mM rNTPs, and ~1 unit pyrophosphatase at 37 °C for 16 hours. Reactions were subsequently extracted with one volume of phenol-chloroform mixture (1:1), and resultant RNAs were precipitated and quantified (Thermo Scientific NanoDraop2000).

2.3.3 Preparation of Small RNA

In vitro transcribed *TgUlp1* NAT was treated with ShortCut RNase III prior to transfection for dual luciferase assay according to manufacturer's protocol (New England Biolabs, Inc., #M0245S). The products were checked by PAGE analysis.

2.3.4 PAGE Analysis

RNase III digestion products of *in vitro* transcribed *TgUlp1* NAT were suspended in formamide 2X buffer and separated on a 12.5% denaturing urea gel. The gel was stained for 30 minutes with SYBR Gold Nucleic Acid Gel Stain (ThermoFisher Scientific, #S11494) in TBE buffer and imaged using BioRad Molecular Imager FX.

2.4 Electroporation

Freshly lysed parasites were harvested and used for transfection via electroporation using a BTX ECM 630 (1500 volts, 25 Ω , and 25 μ F). For each electroporation, 2 μ g of FFLuc plasmid and varying amounts of RnLuc plasmids were mixed with harvested parasites in 400 μ L of electroporation mixture (120 mM KCl, 0.15 mM CaCl₂, 10 mM

$\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ (pH 7.6) 2 mM EDTA, 5 mM MgCl_2 , 2 mM ATP, 5 mM glutathione). Table B5 lists the type and amount of plasmid electroporated depending on experiment. Following electroporation, the parasites were allowed to infect confluent HFF monolayers and grown under different conditions depending on the experiment.

2.5 Dual Luciferase assay

Parasites were grown under testing conditions (24 hours under neutral pH, 5% CO_2 , or up to 72 hours under alkali conditions, atmospheric CO_2 conditions). The parasites were harvested and lysed with 100 μL of 1x Passive Lysis Buffer (Promega, #E1531) and incubated for 15 minutes at room temperature. Lysates were cleared of debris by centrifugation (12,000 x g for 1 minute) and used in the dual luciferase assay, which was carried out in a 2-step fashion. To measure FFLuc activity, 20 μL of lysate was added to a freshly made reaction mixture (100 μL) containing 200 μM D-luciferin, 20 mM Tricine, 10 mM MgSO_4 , 5 mM DTT, 250 μM ATP and 250 μM Coenzyme A. The mixture was incubated for 5 seconds at room temperature and luminescence was measured with a 20/20n Luminometer (Turner Biosystems). For RnLuc assay, 20 μL of lysate was added to a freshly made reaction mixture (100 μL) containing 0.1 μM Coelenterazine, 100 mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ (pH 7.6) 500 mM NaCl, 1 mM EDTA, and 0.02% BSA. The mixture was incubated for 5 seconds at room temperature and luminescence measured with a 20/20n Luminometer (Turner Biosystems). RnLuc activity were normalized to FFLuc activity for a direct comparison across independent experiments. Three independent experiments were performed for every dual luciferase assay.

2.6 RT-PCR Analysis

2.6.1 RNA Isolation

Total RNA was isolated from freshly lysed parasites using TriReagent (Molecular Research Center, #TR 118) and treated with RQ1 RNase free DNase (Promega, #M6101) following manufacturer's protocol. Treated RNA was extracted with one volume of phenol-chloroform mixture (1:1) three times, and resultant RNA was precipitated and quantified (Thermo Scientific NanoDraop2000). Integrity of the RNA was determined by gel electrophoresis on 1% agarose gel. Samples were stored at -20°C for future use.

2.6.2 Reverse Transcriptase

First strand cDNA synthesis was performed using ~ 500 ng isolated total RNA. The reaction was performed using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) (New England Biolabs, Inc., #M0253S) following the manufacturer's protocol. A negative control reaction in the absence of M-MLV was performed at the same time. Samples were stored at -20°C for future use.

2.6.3 RT-qPCR

The cDNA reaction mixtures were diluted 1/1000 dilution for tachyzoite samples and 1/10 dilutions for bradyzoite samples and analyzed by using Fast EvaGreen® qPCR Master Mix (2x) (Cat #31003) and OneStepPlus thermocycler (Applied Biosystems) following the manufacturer's protocol. Primers used for RT-qPCR are listed in Table B4. Relative *TgUlp1* mRNA and NAT expression levels in comparison to GAPDH expression levels were measured and calculated using $2^{-\Delta\Delta\text{Ct}}$ protocol. RT-qPCR were analyzed using both biological and technical triplicates.

2.7 Immunofluorescence Assay

Cells were fixed in 3% paraformaldehyde in PBS for 10 minutes, permeabilized with 0.25% Triton X-100 in PBS for 15 minutes, and non-specific sites were blocked with 5% equine serum in PBS for 1 hour at room temperature. Slides were incubated for 1 hour with FITC-conjugated *Dolichos biflorus* lectin (1:300, L9142, Sigma) to stain cyst walls formed under bradyzoite culturing conditions. Staining of the nuclei was carried out by incubation in the presence of Hoechst 33342 solution (3 μ M). All images were taken with Leica DMI 6000B inverted fluorescent microscope using HCX PL Apo 40x/1.40-0.70 objective and a Leica DFC 360FX camera in addition to the Leica Application Software (LAS).

CHAPTER 3 – RESULTS

3.1 Establishing a Dual Luciferase Reporter to Study Promoters

To identify the element(s) controlling the expression of *TgUlp1* mRNA and NAT, a dual reporter assay was used. The term ‘dual’ refers to the use of two individual plasmids expressing different reporter proteins. For this study, one of the reporter plasmids constitutively expresses firefly (*Photinus pyralis*) luciferase under the control of previously characterized tubulin (TUB) promoter. The activity of FFLuc serves an internal control for transformation efficiency and expression level. The second plasmid expresses sea pansy (*Renilla reniformis*) luciferase, under the control of putative promoter sequences for *TgUlp1* mRNA and NAT. The activity of RnLuc reflects the activity of the regulatory sequence being investigated. As the region upstream from the first exon of *TgUlp1* mRNA and the region downstream from the intron 7 of *TgUlp1* locus were hypothesized to carry the sense and antisense promoter sequence, three overlapping regions upstream from exon 1 of *TgUlp1* mRNA on the sense strand (SS) and four overlapping regions downstream from intron 7 on the antisense strand (AS) were tested (Figure 3.1). These sequences were amplified by PCR (Figure 3.2) and placed upstream from the RnLuc CDS (Figure 3.3). Resultant plasmids were collectively called pPutPromRnLuc (Table B2). The successful expression of RnLuc by the putative promoter was indicative of the sequences’ ability to initiate transcription.

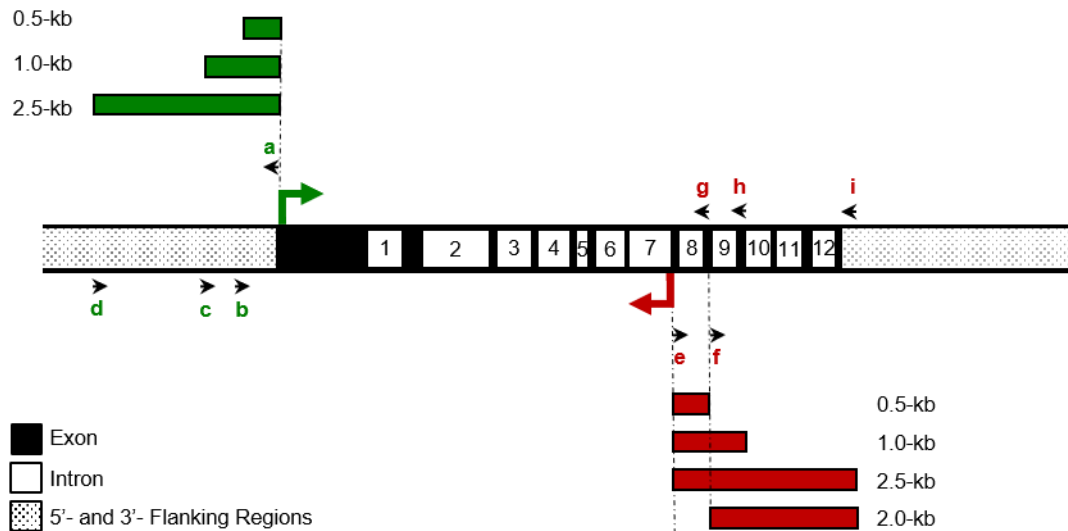


Figure 3.1 Putative Regions Tested for Promoter Activity

Schematic representation of the *TgUlp1* locus. *TgUlp1* has 13 exons indicated as black boxes and 12 introns indicated as numbered white boxes. The region upstream from the first exon of *TgUlp1* locus was hypothesized to carry the sense promoter sequence controlling the transcription of *TgUlp1* mRNA (represented by green arrow). Green bars upstream from the first exon represent putative promoter sequences on the sense strand amplified and tested for sense promoter activity. The region downstream from the intron 7 of *TgUlp1* locus was hypothesized to carry the antisense promoter sequence controlling the transcription of *TgUlp1* NAT (represented by red arrow). Red bars downstream from the intron 7 represent putative promoter sequences on the antisense strand amplified and tested for antisense promoter activity. The primers used for PCR analysis and cloning are indicated by letters and arrows and listed in Table B1.

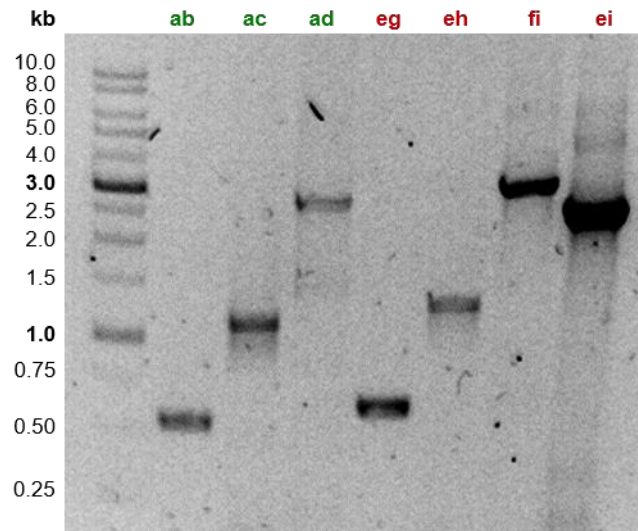


Figure 3.2 Representative Gel Image of PCR Analysis of Putative Promoters

Primer sets a-b, a-c, and a-d were used to amplify 0.5-kb, 1.0-kb, and 2.5-kb regions on the sense strand upstream from exon 1. Primer sets e-g, e-h, f-i and e-i were used to amplify 0.5-kb, 1.0-kb, 2.5-kb and 2.0-kb regions on the antisense strand downstream from intron 7. Amplicons were gel purified and cloned into plasmid construct to control the expression of RnLuc.

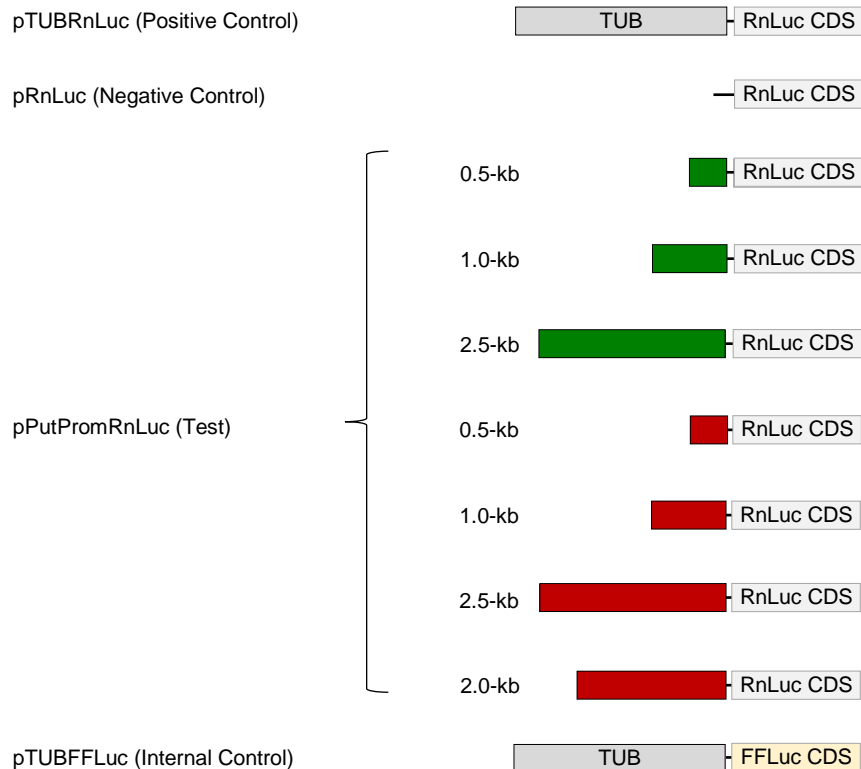


Figure 3.3 Plasmid Constructs Used in the Promoter Assays

In the plasmid pTUBRnLuc, the expression of *Renilla* luciferase (RnLuc) is under the control of tubulin (TUB) promoter. This plasmid was used as a positive control for an active promoter. The plasmid pRnLuc lacking a promoter was used as a negative control for background activity. In the test RnLuc plasmids – collectively called pPutPromRnLuc – the TUB promoter was replaced with putative promoter sequences from the *TgUlp1* locus from the sense strand (green – 0.5-kb, 1.0-kb, or 2.5-kb) or the antisense strand (red – 0.5-kb, 1.0-kb, 2.5-kb, or 2.0-kb). A successful expression of RnLuc by a putative promoter is indicative of the sequence promoter activity. The plasmid pTUBFFLuc codes for *Firefly* luciferase (FFLuc) under the control of a TUB. All RnLuc activities are normalized to FFLuc activity.

The activity of pTUBRnLuc was used to establish maximum promoter activity. The activity of the pRnLuc – no promoter – was used to establish the threshold for background activity. When equal amounts of pTUBRnLuc and pRnLuc (5 μ g) were electroporated and the activity was compared, we found that the activity of the plasmid with TUB promoter was approximately 70x greater than no promoter (Table 3.1). This establishes a wide range over which the activity of putative promoters can be compared.

Table 3.1 RnLuc Activity of Control Plasmids for Promoter Assay

Plasmid	Plasmid (μg)	Trial 1	Trial 2	Trial 3	Average
pRnLuc	5	168	205	204	192
pTUBRnLuc	5	15072	10311	15190	13524
p2.5kbSSRnLuc	5	441	413	245	366

Next, the activity of TUB promoter was compared to putative 2.5-kb SS promoter. Both promoters are approximately the same in length (~2.5-kb), thus differences in activity would be result from differences in sequences. The activity of different dilutions of TUB promoter plasmid were compared to 5 μ g of putative 2.5-kb SS promoter. As shown in Figure 3.4, the activity of 2.5-kb SS is 100x-1000x lower than TUB activity.

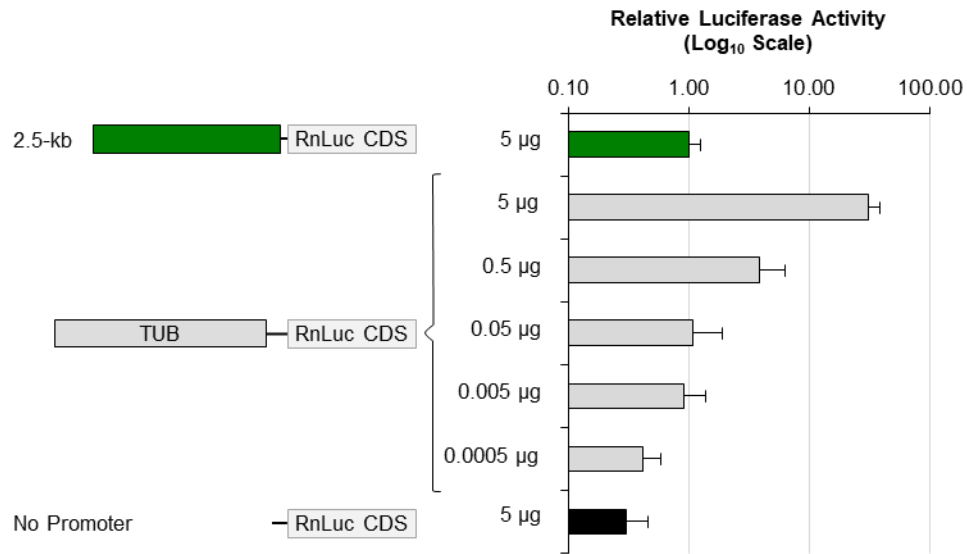


Figure 3.4 Relative Luciferase Activity of Putative 2.5-kb *TgUlp1* mRNA promoter Compared to Strong Tubulin Promoter

Relative RnLuc activity was measured and normalized to that of co-transfected FFLuc activity. The green bar represents the activity of the 2.5-kb putative promoter on the sense strand (5 μg). Gray bars represent the activity of TUB promoter in varying amounts and black bars represent activity of no promoter (5 μg). Activity of the 2.5-kb sense putative promoter was compared to 10-fold dilutions of the TUB promoter. Error bars represent SEM (n = 3).

3.2 Identifying *TgUlp1* mRNA and NAT Promoters

With the promoter assay established, the next aim was to identify the sequence controlling *TgUlp1* mRNA and NAT expression. Promoter activity was measured in tachyzoite and bradyzoite stages of *T. gondii* life cycle in both Type I and Type II strains to determine if life stage or strain had any effect on promoter activity. RnLuc and FFLuc plasmids were electroporated into freshly lysed parasites, grown for 24 hours and then collected to measure luciferase expression. Relative RnLuc/FFLuc activity of each putative promoter was compared to the activity of no promoter as opposed to the highly active TUB promoter. This allowed me to determine which putative promoter was significantly active.

In Type I tachyzoites (Figure 3.5 – solid bars), the 1.0-kb SS and the 2.0-kb AS show a significant increase in RnLuc expression when compared to no promoter. The promoter assay in Type II revealed that the same 1.0-kb SS and the 2.0-kb AS putative promoters have high RnLuc expression (Figure 3.5 – dotted bars), however only the 2.0-kb AS is statistically significant. Activity for all pPutPromRnLuc is similar between the strains which indicates that strain differences do not affect *TgUlp1* mRNA and NAT promoter activity.

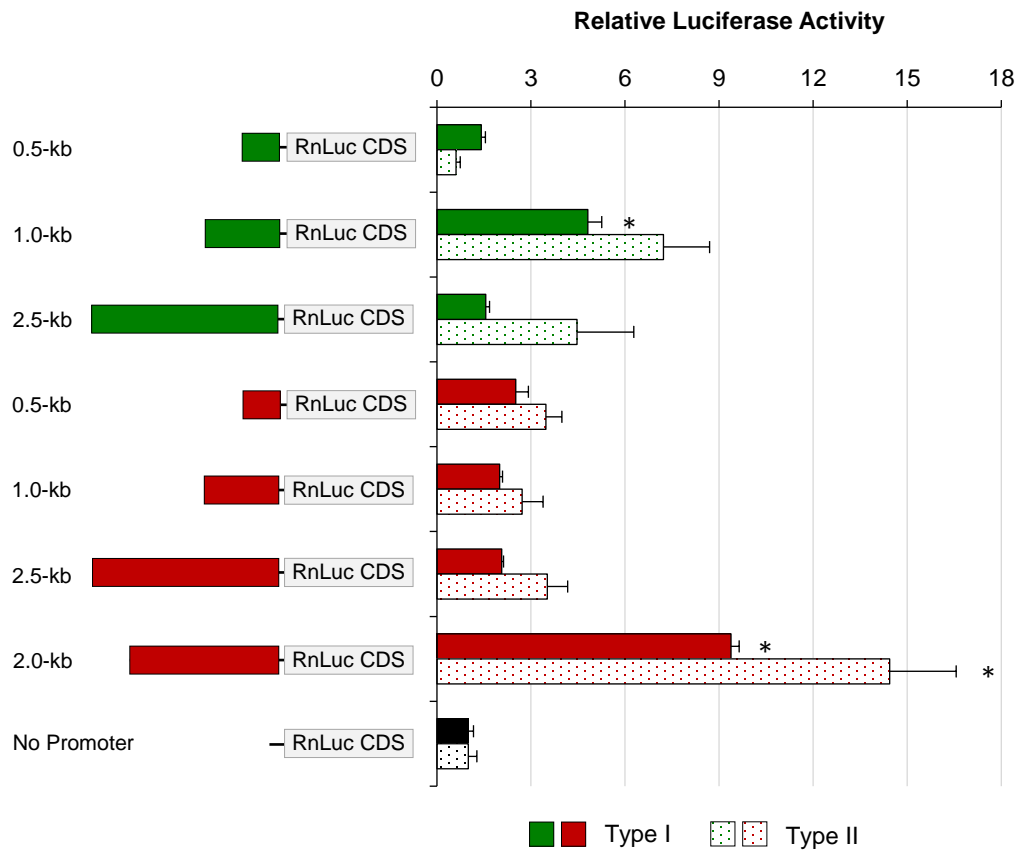


Figure 3.5 Promoter Activity in Type I and Type II *T. gondii* Tachyzoites

Relative RnLuc activity was measured and normalized to that of co-transfected FFLuc activity. Green bars represent activities of sense strand promoters, red bars represent activities of antisense strand promoters and black bars activities of represent no promoter. Solid bars represent activity in Type I, and dotted bars represent activity in Type II. The 1.0-kb region on the sense strand and 2.0-kb region on the antisense strand have the highest RnLuc expression when compared to no promoter in both Type I and Type II. Error bars represent SEM (n = 3). *P < 0.05; compared to no promoter control, unpaired two-sided t-test.

3.3 *TgUlp1* mRNA and NAT Promoter Activity in Bradyzoites

Immunofluorescence assay was performed to confirm both Type I and II strains ability and efficacy to convert to bradyzoites. Culturing the infected HFF monolayers under a high pH (~8) and low CO₂ (0.05% atmospheric level) is commonly used to convert tachyzoites to bradyzoites [54]. Tachyzoites were allowed to infect confluent HFF monolayers. After 6 hours, the media was changed to alkali conditions to allow conversion for another 24 hours. Cells were fixed and stained with FITC-conjugated *Dolichos biflorus* lectin (Dol-FITC). Dol-FITC labels the cyst walls that is specific to the bradyzoite life stage of the parasite. As shown in Figure 3.6, cyst walls were detected within 24 hours with both Type I and Type II parasites.

To evaluate whether the putative promoters were active under alkali growth conditions, the promoter assay was performed in Type I bradyzoites after 24 hr incubation in alkali media. Figure 3.7 shows that the 1.0-kb SS and 2.0-kb AS region have the highest activity for RnLuc expression. However, only the 2.0-kb AS promoter is significant when compared to no promoter. Due to the high similarity in promoter activity between in Type I and II tachyzoites, it was assumed that Type I and II bradyzoites would also have a similar pattern of activity.

Based on these promoter assays, the promoter for *TgUlp1* mRNA was identified to be located within the 1.0-kb SS region upstream from exon and the promoter for *TgUlp1* NAT was identified to be located within the 2.0-kb AS region located downstream from intron 9. Their activities by promoter assay indicates that both sense and antisense promoters are active in tachyzoites, but only the antisense promoter is active in bradyzoites.

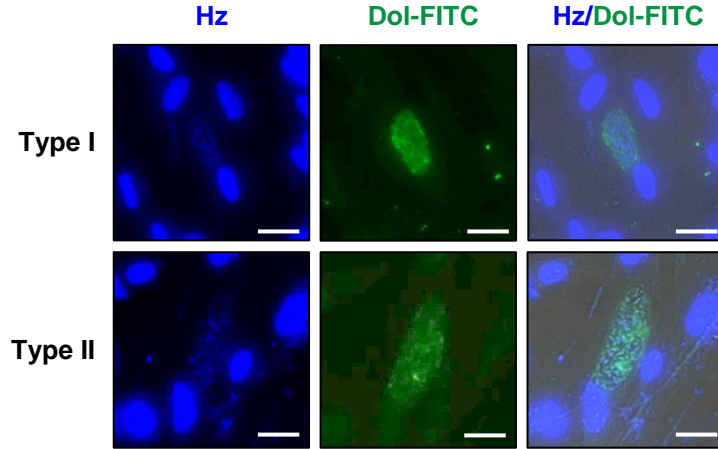


Figure 3.6 Immunofluorescence Assay of Cyst Formation in Type I and II *T. gondii*.

Tachyzoites were used to infect confluent HFF monolayers. After 6 hours, the parasites were grown under alkali conditions for another 24 hours then stained with Hoechst (Hz) and a cyst-specific FITC conjugated lectin (DoI-FITC). Slides were microscopically examined for bradyzoite conversion indicated by the presence of a cyst wall. Scale bars represent 10 μm .

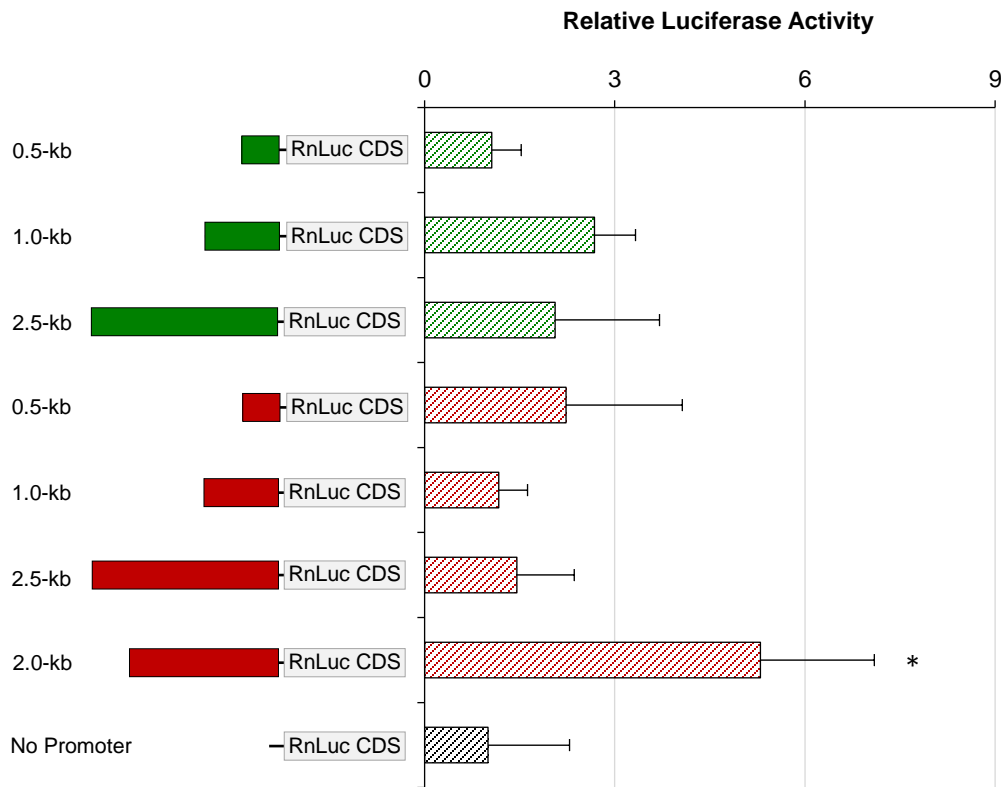


Figure 3.7 Relative activity of RnLuc Under the Control of Putative Promoters in Type I *T. gondii* Bradyzoites.

Relative RnLuc activity was measured and normalized to that of co-transfected FFLuc activity. Green bars represent activities of sense strand promoters, red bars represent activities of antisense strand promoters and black bars represent activities of no promoter. The 2.0-kb region on the antisense strand shows significant increase in RnLuc expression when compared to no promoter. Error bars represent SEM (n = 3). *P < 0.05; compared to no promoter control, unpaired two-sided *t*-test.

Our next aim was to investigate the pattern of expression when the identified promoters of *TgUlp1* mRNA and NAT were analyzed over longer incubation in alkali media. Type II was used for further study in bradyzoites due to their ability to switch to bradyzoite more easily and higher clinical relevance. Type II parasites were switched to alkali media 6 hours after electroporation and incubated in alkali media for 24hr, 48hr and 72 hr before being collected for the dual luciferase assay. As shown in Figure 3.8, the sense promoter activity decreases slightly but does not show any significant changes over time. The antisense promoter shows a steep decrease in activity at 24 hours which remains consistent up to 72 hours.

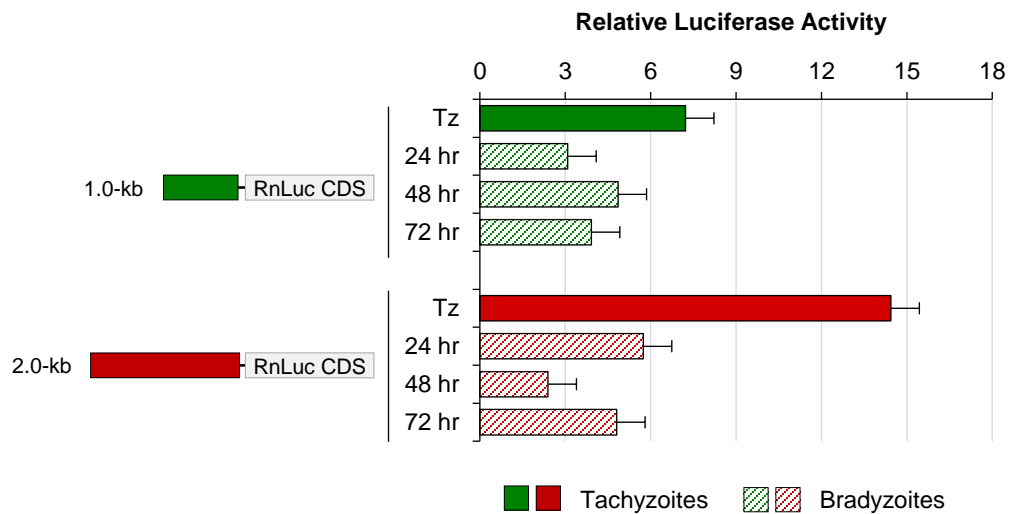


Figure 3.8 Change in Sense and Antisense Promoter Activity in Type II Bradyzoites

Relative RnLuc activity was measured and normalized to that of co-transfected FFLuc activity. Solid bars represent expression of the promoters in tachyzoites (Tz) and striped bars represent expression of the promoters after incubation in alkali media. Error bars represent SEM (n = 3).

3.4 *TgUlp1* mRNA and NAT Expression

The promoter assay suggested that transcription of *TgUlp1* mRNA and NAT is differentially regulated between tachyzoites and bradyzoites. To further confirm this, RT-qPCR was performed in Type II parasites to determine the level of *TgUlp1* mRNA and NAT expression. Total RNA was collected from freshly lysed parasites for tachyzoite samples, and after incubation in alkali media for 24 hours for bradyzoite samples. Both *TgUlp1* mRNA and NAT were detected in tachyzoites. After conversion to bradyzoites, *TgUlp1* mRNA was detected to be 10x lower but *TgUlp1* NAT was 60x higher (Figure 3.9).

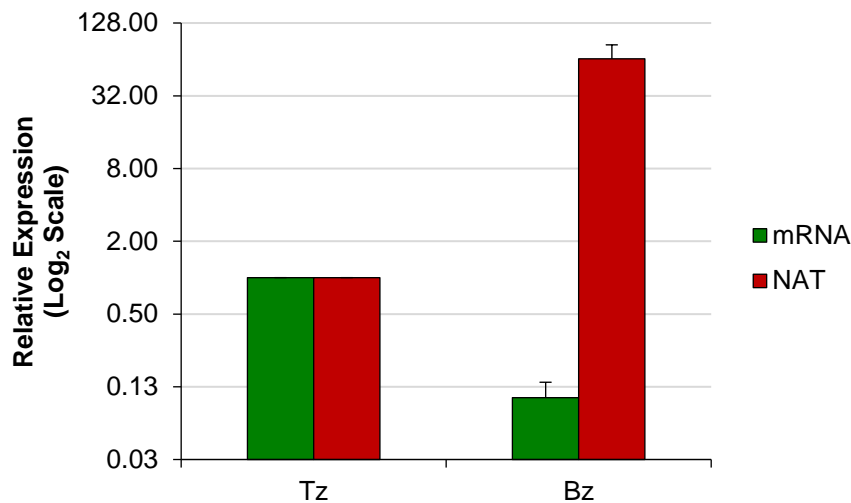


Figure 3.9 Expression of *TgUlp1* mRNA and NAT in Tachyzoites and Bradyzoites

Bradyzoite samples were collected after growing for 24 hours in alkali media. The cDNA reaction mixtures were diluted 1/1000 dilution for tachyzoite samples and 1/10 dilutions for bradyzoite samples. Relative *TgUlp1* mRNA and NAT expression levels in comparison to GAPDH expression levels were measured and calculated using $2^{-\Delta\Delta C_t}$ protocol. RT-qPCR were analyzed using both biological and technical triplicates. Error bars represent SEM.

3.5 The Effect of *TgUlp1* NAT on Gene Expression

3.5.1 Structural Analysis of *TgUlp1* NAT

It has been hypothesized that *TgUlp1* NAT is a precursor for short regulatory RNAs belonging to miR60 family [53], which are capable of regulating the expression of *TgUlp1* mRNA expression. The RNA structure prediction algorithm Mfold [55] was used to generate secondary structures for a further analysis of *TgUlp1* NAT as the substrate of Dicer. Being the key enzyme of RNAi, Dicer recognizes hairpin (stem-loop) structures of transcripts and cleaves them to yield short double-stranded RNA products of approximately 21-23 bps with 3'-overhang to guide RNA-silencing complexes. Based on the most stable predicted structure shown in Figure 3.10, various locations with hairpin structures could serve as the substrates to produce miRNA. The sequences of the hairpin structures are shown in Table 3.2. The sequence from the potential miRNA were used to predict interactions with the sense transcript. For example, hairpin #1 will give a miRNA product, whose upper strand can recognize *TgUlp1* transcript at the nucleotides +4786 to +4806 in the intron 2. The lower strand can recognize *TgUlp1* transcript at the nucleotides +2809 to +2828 in the intron 2.

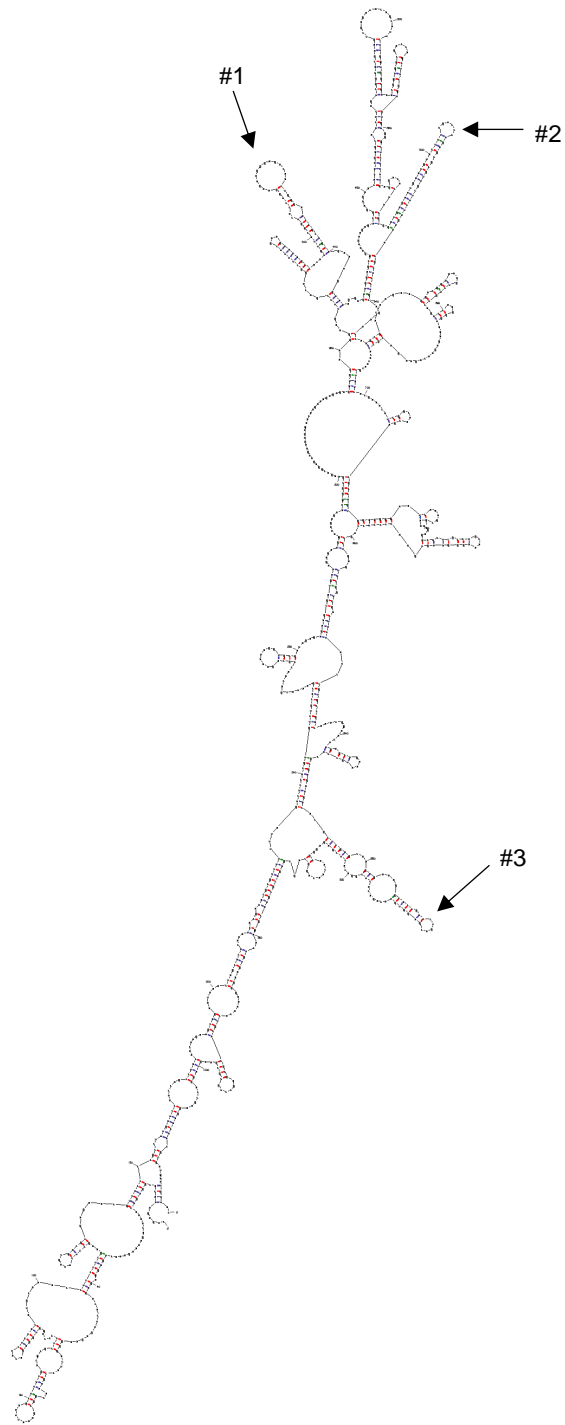


Figure 3.10 *TgUlp1* NAT Structure Prediction

Predicted secondary structure of *TgUlp1* NAT using Mfold. It contains many hairpin loops (indicated by black arrows) that maybe processed by Dicer to generate miRNA.

Table 3.2 Stem Loops Predicted to be Cleaved by Dicer

Yellow highlights indicate Watson-Click base pairing, and green highlights indicate wobble (G:U) pair. Green and red bases are from step loop and black bases are from *TgUlp1* transcript.

Stem Loop Sequence	Matches on <i>TgUlp1</i> Locus
<p>Hairpin #1 (45nts)</p> <pre> AGAA 5' UGUUACGGUCACCUGGA G : G 3' ACGACGCCAA GGGUCA AG ACAG </pre>	<p>Exon 1 +2809 to +2828 5' CTCCTCAGCCTTCGGCTCTT 3' 3' ACGACGCCAAGGUCAACAG 5'</p> <p>Intron 2 +4786 to +4806 5' TTCTGCGAGAAAACGAAAGA 3' 3' AAGAAGGUCCACUGGCAUUGU 5'</p> <p>Intron 2 +5183 to +5203 5' TTTTCTACGCGCGCGTCAGA 3' 3' AAGAAGGUCCACUGGCAUUGU 5'</p> <p>Intron 3 +5635 to +5655 5' CTCTCTAGGTGAGCATCAGC 3' 3' AAGAAGGUCCACUGGCAUUGU 5'</p> <p>Intron 6 +7051 to +7070 5' CTCTTACCCTTGCTTGGC 3' 3' ACGACGCCAAGGGCAACAG 5'</p> <p>Intron 6 +7096 to +7116 5' TGGTTCTGCAGGAGCGTAACA 3' 3' AAGAAGGUCCACUGGCAUUGU 5'</p> <p>Exon 8 +7771 to +7791 5' TTCTTCTACGCCAAGCTGACG 3' 3' AAGAAGGUCCACUGGCAUUGU 5'</p> <p>Exon 9 +8237 to +8256 5' CACTGGACTCTCGCCGTCGT 3' 3' ACGACGCCAAGGGCAACAG 5'</p> <p>Intron 9 +8275 to +8295 5' TGGATCCACGTGACAGTTTCT 3' 3' AAGAAGGUCCACUGGCAUUGU 5'</p>

<p>Hairpin #2 (46 nts)</p> <pre> U G 5' GACGAGAAUGAACGUACUUU U : : : C 3' UUGUUCUUCGUUGC GCGAGA C C C </pre>	<p>5'-Flanking +2061 to +2080</p> <pre> 5' AAAAAACG GCCTTTC CCGTG 3' 3' UUUCAUGCAAGUAAGAGCAG 5' </pre> <p>Intron 1 +4079 to +4098</p> <pre> 5' ATCTTTCCTTCCTTCTCGCC 3' 3' UUUCAUGCAAGUAAGAGCAG 5' </pre> <p>Intron 5 +6552 to +6571</p> <pre> 5' CCTCTACTTTCATGCTCGTC 3' 3' UUUCAUGCAAGUAAGAGCAG 5' </pre> <p>Intron 6 +7073 to +7092</p> <pre> 5' ACATTTCCGTCCTTCTTGTC 3' 3' UUUCAUGCAAGUAAGAGCAG 5' </pre>
<p>Hairpin #3 (57 nts)</p> <pre> C 5' CUCGAG CAA GAG AAAG GCACCGUC C : G 3' GAGCUCA GCUC UGUGG AG A CAG ACAGACA C </pre>	<p>5'-Flanking +2149 to +2172</p> <pre> 5' GAATTTTCCCTTCTCTTGTCCTGG 3' 3' CUGCCACGGAAAGAGAAACGAGC 5' </pre> <p>Intron 9 +8410 to +8433</p> <pre> 5' TGGGTGTCTTGCCCTTACTGCAG 3' 3' CUGCCACGGAAAGAGAAACGAGC 5' </pre> <p>Intron 12 +9573 to +9596</p> <pre> 5' GATGGTGAGTGTCTTTTGCGGGTG 3' 3' CUGCCACGGAAAGAGAAACGAGC 5' </pre> <p>3'-Flanking +10614 to +10641</p> <pre> 5' TCGATCTGCCGCGTCTCTCTCTCCT 3' 3' AGGUGUACAGACACCGGACACUCGA 5' </pre>

3.5.2 NAT and its Derivatives on the Expression of *RnLuc* Transcripts

A dual reporter system was used to determine if *TgUlp1* NAT and its derivatives have a regulatory effect. For this study, both of the reporter plasmids constitutively and individually express FFLuc or RnLuc under the control of TUB promoter to ensure a high level of expression. While the FFLuc activity serves as an internal control for transformation efficiency and expression level, the RnLuc activity will reflect the regulatory effect of *TgUlp1* NAT and its derivatives. The *RnLuc* transcripts were engineered to carry various portions of *TgUlp1* mRNA containing predicted miRNA binding sites (25 nts), referred to as site A or B (Figure 3.11). These miRNA binding sites were placed at the 3'-UTR of the RnLuc transcript. Both constructs were electroporated along with single stranded NAT (ssNAT) and ssNAT digested with RNase III (NAT-R). NAT-R was confirmed to produce short RNA using gel electrophoresis. It was observed that ssNAT was able to lower RnLuc expression by $55 \pm 2.4\%$ and $40 \pm 1.8\%$ for constructs A and B respectively (Figure 3.12). NAT-R was also able to lower RnLuc expression by $47 \pm 1.2\%$ and $41 \pm 1.6\%$ for constructs A and B respectively.

The promoter assay suggests that both *TgUlp1* mRNA and NAT transcription is active in tachyzoites. Thus, it is possible that *TgUlp1* mRNA and NAT form dsRNA which may confer a regulatory effect on *TgUlp1* mRNA expression. To test this, we electroporated constructs A and B with long double stranded NAT (dsNAT, 1167bps). In Figure 3.12, we see that dsNAT did not have an effect on RnLuc expression.

Predicted mRNA Binding Site Sequence	Location on <i>TgUlp1</i> mRNA
5' CGUCCUGGACACAAGGACGACUAUCU 3'	Exon 1/2 Junction (+1165 to +1190) (Site A)
5' UGCCAUGCAUGGAAAGAAAUGUGUG 3'	3' Flanking Region (+2700 to +2725) (Site B)

Figure 3.11 Predicted Binding Sites on *TgUlp1* mRNA Used To Detect Self-Regulation By *TgUlp1* NAT

Two sequences on the *TgUlp1* mRNA were tested for self-regulation. Site A is in the CDS spanning the exon 1/2 junction (+1165 to +1190). Site B is in the 3'-UTR (+2700 to +2725). The sites were cloned downstream of the RnLuc CDS in pTUBRnLuc so that the 3'-UTR of the RnLuc transcript would carry the either Site A or B sequences. These constructs were used in a dual luciferase assay.

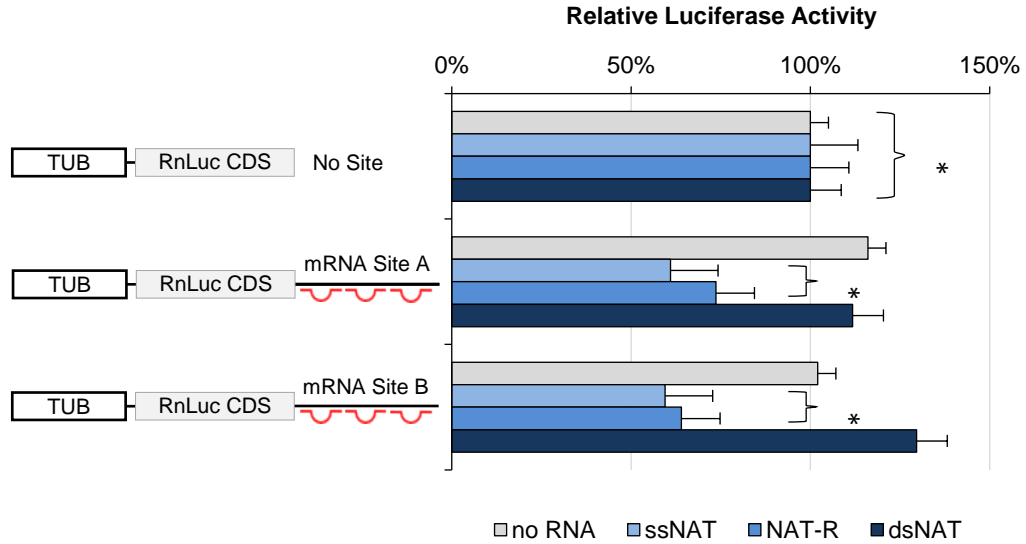


Figure 3.12 Effect of *TgUlp1* NAT on Engineered RnLuc Expression

Relative RnLuc activity was obtained and normalized to that of co-transfected FFLuc activity. Construct A and B were electroporated with and without 5 μ g of *in vitro* transcribed single stranded (ss), double stranded (ds) or RNase III Digested (-R) *TgUlp1* NAT to determine its' effect on RnLuc expression. pTUBRnLuc (no site) was used to establish basal RnLuc activity without any binding sites. ssNAT and NAT-R were able to lower the expression of both constructs RnLuc carrying predicted miRNA binding sites from *TgUlp1* mRNA. Error bars represent SEM (n = 3). *P < 0.05; compared to no RNA, unpaired two-sided t-test.

3.6 Effect of *TgUlp1* NAT on the level of *TgUlp1* mRNA

The dual luciferase assay suggested that *TgUlp1* NAT could be a precursor for miR60 family that regulates the expression of *RnLuc* transcripts. To determine the effect of *TgUlp1* NAT on the expression of mRNA, ssNAT was electroporated into Type II parasites and mRNA expression was measured by RT-qPCR. Figure 3.13 describes the experiment flow. Lysed parasites were electroporated with (+) and without (mock, -) ssNAT. RNA samples were collected after 24 hours to determine the immediate effect of ssNAT. Parasites were subcultured twice, and RNA was collected after each subculture to observe changes in *TgUlp1* mRNA expression. In Figure 3.14, we observed that after mock electroporation, mRNA level dropped approximately 10x and stayed consistently low for 120 hours post electroporation. When ssNAT is electroporated, mRNA expression is unaffected compared to mock (Figure 3.14). We also measured *TgUlp1* NAT level after mock electroporation. In Figure 3.15, we see mock electroporation caused an increase NAT expression by approximately 10x which is consistent for 120 hours post electroporation. Due to the changes in *TgUlp1* mRNA and NAT level caused by electroporation process itself, we were unable to determine if NAT had an effect on the level of *TgUlp1* mRNA.

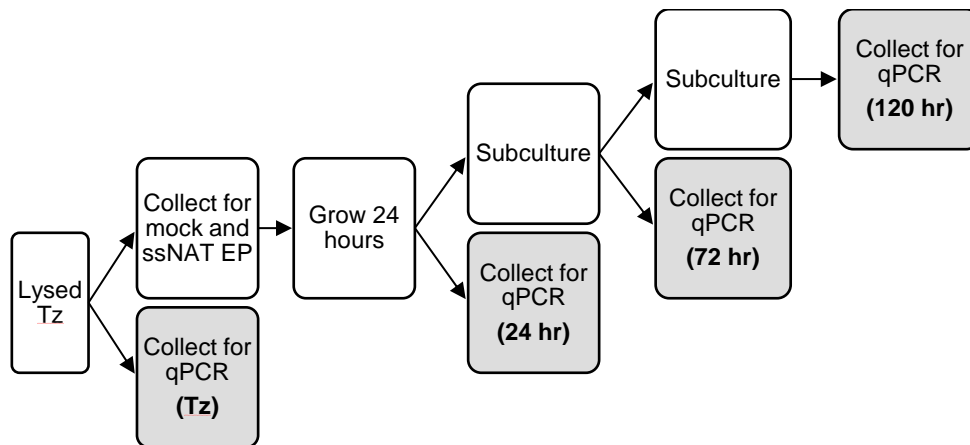


Figure 3.13 Experiment Flow for Detecting Effect of *in vitro* NAT

Lysed parasites were electroporated with (+) and without (mock, -) ssNAT. RNA samples were collected after 24 hours to determine the immediate effect of ssNAT. Parasites were subcultured twice, and RNA was collected after each subculture to observe changes in *TgUlp1* mRNA expression. Tz refers to tachyzoites and EP refers to electroporation.

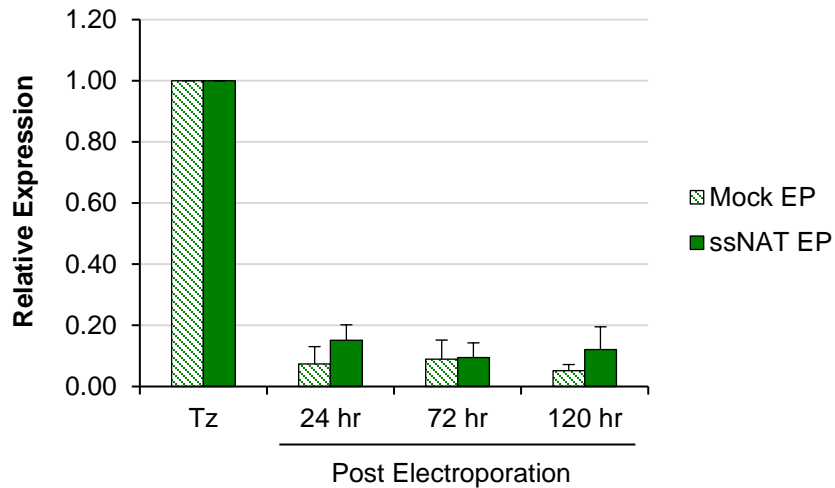


Figure 3.14 *TgUlp1* mRNA Expression After Mock and *in vitro* NAT Electroporation

Striped bar represents mRNA expression after mock electroporation and solid bars represent mRNA expression after *in vitro* NAT electroporation. Relative *TgUlp1* mRNA expression levels in comparison to GAPDH expression levels were measured and calculated using $2^{-\Delta\Delta C_t}$ protocol. RT-qPCR were analyzed using both biological and technical triplicates. Error bars represent SEM.

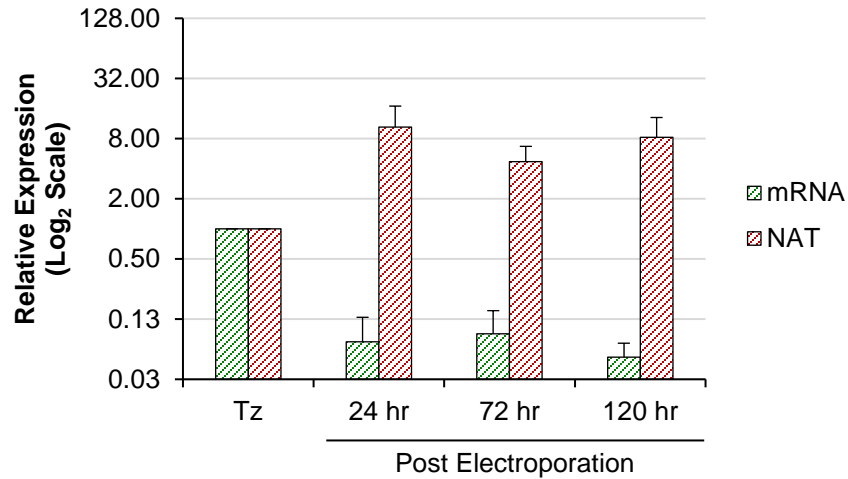


Figure 3.15 *TgUlp1* mRNA and NAT Expression After Mock Electroporation

Striped green bar represents *TgUlp1* mRNA expression after mock electroporation and striped red bars represent *TgUlp1* NAT expression after *in vitro* NAT electroporation. Relative *TgUlp1* mRNA and NAT expression levels in comparison to GAPDH expression levels were measured and calculated using $2^{-\Delta\Delta C_t}$ protocol. RT-qPCR were analyzed using both biological and technical triplicates. Error bars represent SEM.

3.7 *TgUlp1* mRNA and NAT Expression in *TgDicer-KO* and *TgAgo-KO*

To further evaluate whether RNAi is involved in the expression of *TgUlp1* mRNA and NAT, RT-qPCR analysis was performed in *TgDicer-KO* and *TgAgo-KO* strains. *TgDicer-KO* was created by a previous student by inserting the YFP HX gene in frame with the *TgDicer* gene. *TgDicer-KO* is incapable of digesting lncRNA, such as the hypothesized *TgUlp1* NAT. *TgAgo-KO* was created by replacing the entire *TgAgo* gene with hypoxanthine-xanthine guanine phosphoribosyl transferase (HXGPRT) as a selectable marker. *TgAgo-KO* strain has reduced gene silencing ability. RT-qPCR performed show that *TgUlp1* NAT levels is approximately 7x higher in both KO strains compared to its parental strain. *TgUlp1* mRNA levels in both KO strains are approximately 10x lower (Figure 3.16).

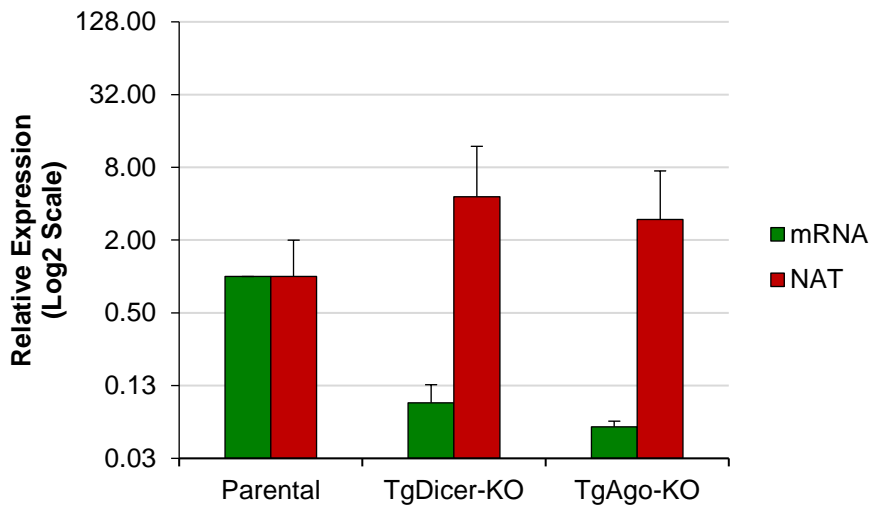


Figure 3.16 *TgUlp1* mRNA and NAT expression in *TgDicer-KO* and *TgAgo-KO* compared to its' parental

Relative *TgUlp1* mRNA and NAT expression levels in comparison to GAPDH expression levels were measured and calculated using $2^{-\Delta\Delta C_t}$ protocol. RT-qPCR were analyzed using both biological and technical triplicates. Error bars represent SEM.

CHAPTER 4 – DISCUSSION

4.1 Promoters for *TgUlp1* mRNA and NAT

To establish the promoter assay, we used the highly expressed β -tubulin (TUB) promoter controlling the activity of the RnLuc reporter in setting an upper limit of detection, and the RnLuc construct without a promoter in setting a lower limit of background detection. There are multiple advantages for using RnLuc and FFLuc. First, the assay is highly reliable. Both enzymes are stable for more than one hour while providing a luminescent signal using independent substrates: coelenterazine for RnLuc and D-luciferin for FFLuc. Second, both enzymes have a linear range covering eight orders of magnitude, which allows for the detection of approximately 0.1 femtogram (approximately 10^{-21} mole) of enzyme [56]. Therefore, the assay provides highly sensitive measurement of gene expression analysis. The reliability, stability and sensitivity of dual luciferase assays have allowed us to establish an assay to identify the promoters for *TgUlp1* mRNA and NAT.

In eukaryotes, two core components constitute a functional promoter. These DNA sequences often referred to as core promoter and proximal promoter [57]. The core promoter contains the RNA polymerase binding site, TATA box, and transcription start site (TSS). The proximal promoter contains sequences that bind transcription factors and is found approximately 250-bps upstream from the TSS. The sequences and locations of these promoter elements contain a certain degree of conservation. Therefore, mapping and functional analysis are essential in gaining a better understanding of these elements. To study the promoter controlling the expression of *TgUlp1* mRNA and NAT, we began by choosing an arbitrary 2.5-kb region upstream from exon 1 to map the putative promoter

controlling *TgUlp1* mRNA, and 2.5-kb region downstream from intron 7 to map *TgUlp1* putative promoter controlling *TgUlp1* NAT. Starting with the 2.5-kb 5'-flanking region of *TgUlp1* locus as the expression of *TgUlp1* is well characterized (Figure 3.1), we compared the activity of this region to TUB promoter and found that the putative *TgUlp1* mRNA promoter activity was 100x-1000x lower. This expression level agrees with microarray analysis published by ToxoDB for *TgUlp1* (TGGT1_2144700) and *TgTUB* (TGGT1_266960), as *TgUlp1* is shown to express at a lower level than *TUB* [58]. Further mapping of the region identified the promoter for *TgUlp1* mRNA to be the 1.0-kb region upstream from exon 1 on the sense strand. Although it is shorter than the TUB promoter we used, it agrees with the minimal and function promoter controlling TUB [59]. The promoter of *TgUlp1* NAT was identified to be the 2.0-kb region between from intron 9 (+5768) to 3'-flanking region (+7920) on the antisense strand (Figure 4.1).

So far, well-characterized motifs such as TATA box have not been observed in *T. gondii* core promoters [60]. However, there is evidence that other motifs are important for transcription in *T. gondii* such as the initiator element (Inr). Inr is a core promoter, commonly found in genes that lack TATA box and the most common sequence found at the TSS of eukaryotic genes [61]. Using ElemeNT, we identified Inr in the promoter for *TgUlp1* mRNA and NAT (Table B6). We can also identify other important motifs such as GAGACG and TGCATGC in the identified promoter for *TgUlp1* mRNA and NAT as shown in Table B7 and Table B8. GAGACG has been identified as a critical element in many *T. gondii* promoters [62]. In fact, it has been suggested that this motif might be the *T. gondii* equivalent to TATA box seen in other eukaryotes [32]. TGCATGC is the putative binding motif for apicomplexan AP2 (ApiAP2) family of transcription factors and

conserved in other members of the apicomplexan phylum including *T. gondii* [63]. This motif is found in hundreds of *T. gondii* promoters. Inr, GAGACG, and TGCATGC motifs were found almost exclusively within the identified promoter regions for *TgUlp1* mRNA and NAT.

TgUlp1 NAT 2.0-kb promoter exhibits the highest RnLuc activity but is not immediately upstream from intron 7 of the antisense transcript detected, as shown in Figure 4.1 [53].

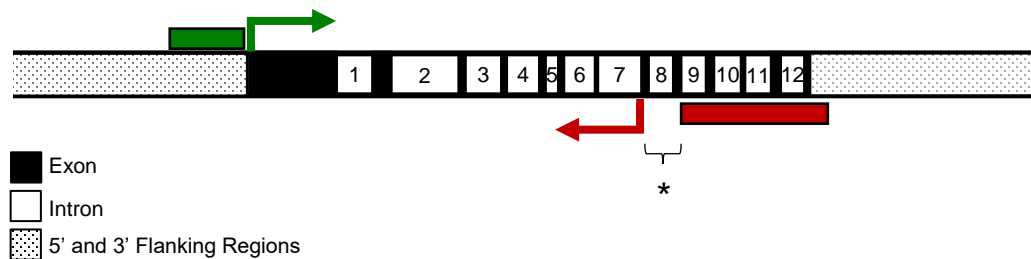


Figure 4.1 Identified Promoters of *TgUlp1* mRNA and NAT

Green and red bar represents the promoter region for *TgUlp1* mRNA and NAT respectively. Green and red arrows represent the transcription of *TgUlp1* mRNA and NAT respectively. * indicates a 500-bps region between the detected *TgUlp1* NAT and its identified promoter.

Two possible explanation to this finding are: one, this 500-bps region, indicated by a star (Figure 4.1), contains regulatory elements such as silencers which can bind repressor proteins to negatively regulate transcription. To characterize this controlling elements, one can perform mutation analysis of the nucleotide sequence of the region. A dual luciferase assay, such as those used in the study, can be modified for the study. The second

explanation is due to our lack of information on the TSS features. Currently the full length of *TgUlp1* NAT has not yet been characterized, and its 5'-end could start in within this starred region. For future work, one can map the 5'-end of *TgUlp1* NAT by using a technique called primer extension [64]. This technique requires a radiolabeled primer complementary to a region on the transcript. The primer anneals to the transcript and reverse transcriptase is used to synthesize cDNA until it reaches the 5'-end of the transcript. The transcript-cDNA hybrid is denatured, and cDNA product is analyzed on a sequencing gel. Rapid amplification of cDNA ends (RACE) is another technique that can be used to obtain the full length sequence of a transcript [65]. Similar to primer extension, this technique generates cDNA using a gene specific primer and reverse transcriptase. After first strand cDNA synthesis, the original transcript template is treated with RNase and an oligonucleotide adapter is linked to the 3'-end of the cDNA. Thereafter, PCR is performed using a gene specific primer and primer for the adapter to amplify the region with the 5' unknown sequence. PCR products are then cloned into a vector for sequencing.

4.2 Expression Of *TgUlp1* mRNA And NAT During Its Asexual Cycle

T. gondii pathogenicity comes from its ability to differentiate between tachyzoites and bradyzoites to evade the host immune response. Thus, a better understanding of the molecular mechanisms that drive stage conversion between tachyzoites and bradyzoites is necessary to manage transmission and pathogenesis of *T. gondii*. Many studies have documented genes that are exclusively expressed in either tachyzoites or bradyzoites, suggesting that different expression of gene play a major role in coordinating transitions in *T. gondii* [66]. However, studies in differential regulation in *T. gondii* have focused on protein-coding genes. In this study, I focus on antisense gene regulation and observed that

TgUlp1 NAT is differentially regulated at the transcriptional level. In tachyzoites, both identified promoters are active which indicates *TgUlp1* mRNA and NAT transcription may be occurring at the same time. This was further confirmed by RT-qPCR which detects the presence of both sense and antisense transcripts in tachyzoites. In bradyzoites we observe a lower sense promoter activity bradyzoites in agreement with lower *TgUlp1* mRNA expression. However, despite a decrease in antisense promoter activity, RT-qPCR shows us that *TgUlp1* NAT expression is higher in bradyzoites compared to tachyzoites. This indicates that although antisense promoter might be less active in bradyzoites, the antisense transcript turnover is slower. In addition, time course experiment showed no change in promoter activity with longer incubation in alkali conditions. Taken together, the promoter assay and RT-qPCR suggested that the expression of *TgUlp1* NAT and mRNA are differentially regulated at the transcriptional and post- transcriptional level *via* promoter activity and transcript turnover.

Although the promotor assay has been useful in discriminating between active and nonactive promoter sequences, the discrepancy between promoter activity and transcript suggested other mechanisms are involved. For example, differences in transcript stability between *TgUlp1* mRNA and NAT and RnLuc would affect our measurements for promoter activity. In addition, distal regulatory elements and chromatin modification are a common feature of the eukaryotic genome which are responsible for transcriptional regulation [67]. In a reporter system, the promoter is stripped of this genomic context which may lead to non-specific promoter activity.

4.3 *TgUlp1* NAT and its Regulatory Effect

ncRNA molecules have generally been discovered by large-scale sequencing projects that were carried out for various model organisms, such as humans, mice and worms. Similarly, large-scale sequencing projects of Apicomplexan parasites have confirmed the presence of long and short ncRNAs, including NATs and miRNA [30, 31]. However, the discovery of *TgUlp1* NAT was fortuitous. During the study of *TgUlp1* mRNA as a potential target of miR60 family, NAT was discovered by gene specific RT-PCR [53]. The fact that *TgUlp1* is the first gene confirmed as the target of miR60 family led us to hypothesize that *TgUlp1* NAT, as a lncRNA species, is a regulatory RNA. To test this hypothesis, we adopted two strategies. One was an *in silico* analysis of *TgUlp1* NAT as the precursor of miRNAs. The analysis of *TgUlp1* NAT showed several potential miRNAs whose sequence are highly complementary to *TgUlp1* mRNA (Table 3.2).

Our second strategy was to use a dual luciferase system in determining the regulatory effect of *TgUlp1* NAT (ssNAT) and its RNase-III digested products (NAT-R). To create the reporter transcript for this experiment, we placed the sequence of predicted binding site at the end of RnLuc CDS and in the 3'-UTR. Such a reporter construct is commonly used in the study of siRNA and miRNA silencing function [68, 69]. We detected that ssNAT and NAT-R were able to downregulate RnLuc expression. This indicates that as a lncRNA, *TgUlp1* NAT could yield short regulatory ncRNAs. It is highly likely that the short regulatory ncRNAs have the structural features of miRNAs. It should be noted that these short regulatory ncRNAs were previously shown to have similar down-regulation effect as those of miR60 family [53]. As both sense and antisense promoters are active in tachyzoites, it is also possible that both transcripts could form perfect-paired double-

stranded RNA that could affect the regulation of *TgUlp1* mRNA. However, we did not detect that the long double-stranded RNA of NAT (dsNAT) has silencing activity. It would be interesting to further test whether the RNase III products of long double-stranded RNA of NAT can exhibit similar activity as those of ssNAT and NAT-R. Consequently, it would imply that *TgUlp1* NAT and its counter mRNA form dsNAT and serve as a siRNA precursor.

When we analyzed the predicted secondary structure of *TgUlp1* NAT, we found multiple stem loops that could be substrates for Dicer. We analyzed the predicted products against mRNA and found multiple sites which may be targeted by *TgUlp1* NAT derivatives to regulate expression. To identify if the loops we predicted were actually miRNA precursors, one can perform northern blot of *in vitro* NAT-R products. By probing for the stem loop sequences predicted as Dicer cleavage sites, we can confirm if *TgUlp1* NAT produces short regulatory RNA. In addition, one can sequence NAT-R products and look for a match on the *TgUlp1* transcript. If a match between the sequenced NAT-R products and *TgUlp1* transcript is found, it will provide a more specific and likely target to test for self regulation of *TgUlp1*. To determine if Dicer is involved in processing *TgUlp1* NAT endogenously, one could also perform northern blot in TgDicer-KO and probe for *TgUlp1* NAT.

Since ssNAT and NAT-R were able to downregulate the expression of RnLuc carrying predicting binding sites derived from the *TgUlp1* mRNA, we speculated that *TgUlp1* NAT would directly affect the level of *TgUlp1* mRNA. However, we are unable to conclude if *TgUlp1* NAT has a direct effect on mRNA levels. Instead, we observed that the stress of electroporation alone affected *TgUlp1* transcript levels; mRNA level

decreased, and NAT levels increased and remained consistent up to 120 hours. This indicated *TgUlp1* NAT may play a role in the parasite's stress response. To further study the effect and role of *TgUlp1* NAT in the response, one can alter the expression of *TgUlp1* NAT by repressing the transcription of *TgUlp1* NAT using CRISPR interference (CRISPRi) [70]. This repurposed CRISPR/Cas9 system requires an inactive version of Cas9 (dCas9) which was created by two point mutations in its RuvC-like (D10A) and HNH nuclease (H840A) domains [71]. These mutations allow dCas9 to bind double stranded DNA, but it no longer has endonuclease activity. Directed by a single guide RNA, the dCas9 will bind at the promoter we identified by the dual luciferase assay. Once bound, it would interfere with transcription initiation and lead to *TgUlp1* NAT knockdown. Interrupting the expression of *TgUlp1* NAT will allow us to measure its effect on the mRNA by RT-qPCR. If *TgUlp1* NAT is essential to the regulation of *TgUlp1*, we would expect to see change in mRNA levels when NAT decreases. Previous studies using CRISPRi show that this system is good for studying overlapping transcription units [71], such as in the case of *TgUlp1* mRNA and NAT. Traditional knockdown methods such as a RNAi have several limitations. First, knocking down antisense transcript require exogenous siRNAs to carry sense transcript sequences. Second, unlike protein-coding genes, many lncRNAs primarily localize in the nucleus. Even though RNAi machinery has been found to be active in the nucleus, siRNAs against nuclear lncRNAs have often proven to be less effective [71].

4.4 Dicer and Ago Knockout Strains of *Toxoplasma gondii*

Dicer and Argonaute are key ribonuclease of the silencing pathways mediated by small ncRNAs, including siRNA and miRNA. Logically, if *TgUlp1* NAT is a precursor

short regulatory ncRNA, it would depend on Dicer and Ago to regulate mRNA expression. We expected to detect a higher expression of *TgUlp1* mRNA in both knockout strains. However, we detected that TgDicer-KO and TgAgoKO have a much lower level of *TgUlp1* mRNA and a higher level of *TgUlp1* NAT in comparison to the parental strain, indicating that Dicer and Ago are involved in maintaining *TgUlp1* mRNA.

A study by Napoli et al., [72] looked at self-regulation of the oncogene *c-MYC* by its NAT. They found that, the *c-Myc* NAT is processed by Dicer to produce short regulatory RNA and that Dicer knockdown resulted in a lower mRNA level and higher NAT level. They speculated that because DICER is also involved in chromatin modifications, *c-Myc* sense and antisense transcripts were regulated by altering the chromatin state. To test this, they treated Dicer knockout cells with HDAC inhibitor which induced acetylation of histones at *c-Myc* locus, further reducing mRNA and increasing NAT levels. This provides evidence that altering chromatin states regulate *c-Myc* NAT transcription, which in turn may negatively regulate *c-Myc* mRNA expression. Although my study did not look at epigenetic regulation of *TgUlp1* NAT, studying this relationship would provide more insight on antisense regulation in *T. gondii*.

Interestingly, we see a very similar pattern of expression for *TgUlp1* mRNA and NAT when we compare RT-qPCR results between bradyzoites, mock electroporation, TgDicer-KO, and TgAgoKO. In each experiment, there was a higher level of *TgUlp1* NAT and lower level of mRNA. This suggests that *TgUlp1* NAT may play a role with managing the parasites stress response. Previous work in studied the relationship between bradyzoite and Ago and found that knockout of Argonaute expression resulted in an increase in

bradyzoite formation [73]. This suggests that RNAi and *TgUlp1* NAT play a role in bradyzoite formation.

RT-qPCR shows a drastic decrease in *TgUlp1* mRNA level following stress and during bradyzoite life stage. Consequently, this would also affect TgUlp1 protein levels and ultimately the SUMOylation pathway in *T. gondii*. Previous work provides evidence that SUMOylation plays an important role in host cell invasion and cyst genesis [52]. If *TgUlp1* NAT does regulate mRNA expression, it will provide more insight into the role of NAT in gene expression and pathogenesis in *T. gondii*. In addition, studies in other organisms provide evidence that NAT also exert their function in a *trans*-acting manner. Therefore, in addition to the *cis*-acting mechanism hypothesized and tested in this study, *TgUlp1* NAT may also function in *trans* to regulate gene expression through various ways.

CONCLUSION

NATs are widespread in eukaryotes and are being recognized as important regulators of gene expression. Although NATs are a common feature in Apicomplexan transcriptome, very little is known about their regulation, function, or its implications in cell biology. A NAT species was fortuitously discovered during the study of *TgUlp1* mRNA as a potential target of miR60 family. Studying the mechanisms controlling the function of *TgUlp1* NAT will provide more insight on the role NATs play in *T. gondii*.

The objective of this study was to identify the elements controlling the expression of *TgUlp1* NAT and elucidate its mechanism of function. Using a dual luciferase assay, we identified the promoter controlling the transcription of *TgUlp1* mRNA and NAT. Although lacking a TATA box, the identified promoters contain many motifs that have been identified in *T. gondii* promoters as important for initiating transcription. Further work revealed that the promoters for *TgUlp1* mRNA and NAT are active in tachyzoites but only NAT promoter is active in bradyzoites. RT-qPCR showed *TgUlp1* mRNA was lower, but *TgUlp1* NAT higher in bradyzoites compared to tachyzoites. Taken together, the data suggests that the expression of *TgUlp1* NAT and mRNA are differentially regulated at the transcriptional level, *via* promoter activity and transcript turnover. Furthermore, this implies stage-specific expression of transcripts and therefore may provide insight on the role NATs have in tachyzoite-bradyzoite differentiation.

TgUlp1 is the first gene confirmed as the target of miR60 family, which led us to hypothesize that *TgUlp1* NAT is a regulatory RNA. Using a dual luciferase, we observed that when *TgUlp1* NAT was *in vitro* processed by RNase III, the products retain the ability to lower the expression of engineered reporters carrying *TgUlp1* mRNA sequences. This

suggests the involvement of RNAi. To further evaluate whether the Dicer and Argonaute, key enzymes in RNAi, are required for the processing and function of *TgUlp1* NAT *in vivo*, RT-qPCR analysis was performed in TgDicer-KO and TgAgo-KO. We observed a lower level of *TgUlp1* mRNA and higher level of *TgUlp1* NAT in both strains compared to the parental. This indicates that Dicer and Ago are involved in maintaining *TgUlp1* mRNA, We were unable to determine if *in vitro* *TgUlp1* NAT had a direct affect on mRNA but did observe that electroporation alone caused decrease in mRNA and increase in NAT levels. The similar transcript levels between bradyzoites and mock electroporation suggest that *TgUlp1* NAT may play a role in the parasite's stress response. Consequently, these changes in *TgUlp1* mRNA and NAT levels would also affect TgUlp1 protein levels and ultimately the SUMOylation pathway in *T. gondii*. Previous work in *T. gondii* show that SUMOylation plays an important role in host cell invasion and cyst genesis. If *TgUlp1* NAT does regulate mRNA expression, it would ultimately affect the parasites ability to invade and form cysts. Therefore, studying the role of *TgUlp1* NAT plays in the parasite will provide more insight into the pathogenesis of *T. gondii*.

REFERENCES

- [1] Kaikkonen, M. U., Lam, M. T., & Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular research*, 90(3), 430-440.
- [2] Marchese, F. P., Raimondi, I., & Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome biology*, 18(1), 206.
- [3] Latgé, G., Poulet, C., Bours, V., Josse, C., & Jerusalem, G. (2018). Natural antisense transcripts: Molecular mechanisms and implications in breast cancers. *International journal of molecular sciences*, 19(1), 123.
- [4] Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA biology*, 10(6), 924-933.
- [5] Wight, M., & Werner, A. (2013). The functions of natural antisense transcripts. *Essays in biochemistry*, 54, 91-101.
- [6] Zhang, Y., Liu, X. S., Liu, Q. R., & Wei, L. (2006). Genome-wide in silico identification and analysis of *cis* natural antisense transcripts (*cis*-NATs) in ten species. *Nucleic acids research*, 34(12), 3465-3475.
- [7] Mercer, T. R., Neph, S., Dinger, M. E., Crawford, J., Smith, M. A., Shearwood, A. M. J., ... & Filipovska, A. (2011). The human mitochondrial transcriptome. *Cell*, 146(4), 645-658.
- [8] Brantl, S. (2012). Small regulatory RNAs (sRNAs): key players in prokaryotic metabolism, stress response, and virulence. *Regulatory RNAs* (pp. 73-109). Springer, Berlin, Heidelberg.
- [9] Shearwin, K. E., Callen, B. P., & Egan, J. B. (2005). Transcriptional interference—a crash course. *TRENDS in Genetics*, 21(6), 339-345.
- [10] Lin, S., Zhang, L., Luo, W., & Zhang, X. (2015). Characteristics of antisense transcript promoters and the regulation of their activity. *International journal of molecular sciences*, 17(1), 9.

- [11] Palmer, A. C., Egan, J. B., & Shearwin, K. E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription*, 2(1), 9-14.
- [12] Sado, T., Hoki, Y., & Sasaki, H. (2005). Tsix silences Xist through modification of chromatin structure. *Developmental cell*, 9(1), 159-165.
- [13] Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & development*, 25(10), 1010-1022.
- [14] Tufarelli, C., Stanley, J. A. S., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G., & Higgs, D. R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nature genetics*, 34(2), 157.
- [15] Zhao, Y., Sun, H., & Wang, H. (2016). Long noncoding RNAs in DNA methylation: new players stepping into the old game. *Cell & bioscience*, 6(1), 45.
- [16] Zinad, H. S., Natasya, I., & Werner, A. (2017). Natural Antisense Transcripts at the Interface between Host Genome and Mobile Genetic Elements. *Frontiers in microbiology*, 8, 2292.
- [17] Salameh, A., Lee, A. K., Cardó-Vila, M., Nunes, D. N., Efstathiou, E., Staquicini, F. I., ... & Lauer, R. C. (2015). PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. *Proceedings of the National Academy of Sciences*, 112(27), 8403-8408.
- [18] Beltran, M., Puig, I., Peña, C., García, J. M., Álvarez, A. B., Peña, R., ... & de Herreros, A. G. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition. *Genes & development*, 22(6), 756-769.
- [19] Zong, X., Nakagawa, S., Freier, S. M., Fei, J., Ha, T., Prasanth, S. G., & Prasanth, K. V. (2016). Natural antisense RNA promotes 3' end processing and maturation of MALAT1 lncRNA. *Nucleic acids research*, 44(6), 2898-2908.

- [20] Sun, J., Wang, X., Fu, C., Wang, X., Zou, J., Hua, H., & Bi, Z. (2016). Long noncoding RNA FGFR3-AS1 promotes osteosarcoma growth through regulating its natural antisense transcript FGFR3. *Molecular biology reports*, 43(5), 427-436.
- [21] Wang, G. Q., Wang, Y., Xiong, Y., Chen, X. C., Ma, M. L., Cai, R., ... & Pang, W. J. (2016). Sirt1 AS lncRNA interacts with its mRNA to inhibit muscle formation by attenuating function of miR-34a. *Scientific reports*, 6, 21865.
- [22] Britto-Kido, S. D. A., Costa, J. R., Pandolfi, V., Marcelino-Guimarães, F. C., Nepomuceno, A. L., Vilela Abdelnoor, R., ... & Kido, E. A. (2013). Natural antisense transcripts in plants: a review and identification in soybean infected with *Phakopsora pachyrhizi* SuperSAGE Library. *The Scientific World Journal*, 2013.
- [23] Bhan, A., & Mandal, S. S. (2015). LncRNA HOTAIR: A master regulator of chromatin dynamics and cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1856(1), 151-164.
- [24] Li, J. T., Zhang, Y., Kong, L., Liu, Q. R., & Wei, L. (2008). Trans-natural antisense transcripts including noncoding RNAs in 10 species: implications for expression regulation. *Nucleic acids research*, 36(15), 4833-4844.
- [25] Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R., & Zhu, J. K. (2005). Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*, 123(7), 1279-1291.
- [26] Rosikiewicz, W., & Makałowska, I. (2016). Biological functions of natural antisense transcripts. *Acta Biochimica Polonica*, 63(4), 665-673.
- [27] Carthew, R. W., & Sontheimer, E. J. (2009). Origins and mechanisms of miRNAs and siRNAs. *Cell*, 136(4), 642-655.
- [28] Okamura, K., Ishizuka, A., Siomi, H., & Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes & development*, 18(14), 1655-1666.

- [29] Kim, K., & Weiss, L. M. (2004). *Toxoplasma gondii*: the model apicomplexan. *International journal for parasitology*, 34(3), 423-432.
- [30] Siegel, T. N., Hon, C. C., Zhang, Q., Lopez-Rubio, J. J., Scheidig-Benatar, C., Martins, R. M., ... & Scherf, A. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. *BMC genomics*, 15(1), 150.
- [31] Radke, J. R., Behnke, M. S., Mackey, A. J., Radke, J. B., Roos, D. S., & White, M. W. (2005). The transcriptome of *Toxoplasma gondii*. *BMC biology*, 3(1), 26.
- [32] Weiss, L. M., & Kim, K. (Eds.). (2011). *Toxoplasma gondii: the model apicomplexan. Perspectives and methods*. Elsevier.
- [33] Ferguson, D. J. (2009). *Toxoplasma gondii*: 1908-2008, homage to Nicolle, Manceaux and Splendore. *Memorias do Instituto Oswaldo Cruz*, 104(2), 133-148.
- [34] Lim, L., & McFadden, G. I. (2010). The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1541), 749-763.
- [35] Dubey, J. P., & Beattie, C. P. (1988). *Toxoplasmosis of animals and man*. CRC Press, Inc..
- [36] Black, M. W., & Boothroyd, J. C. (2000). Lytic cycle of *Toxoplasma gondii*. *Microbiol. Mol. Biol. Rev.*, 64(3), 607-623.
- [37] Dubey, J. P., Lindsay, D. S., & Speer, C. A. (1998). Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clinical microbiology reviews*, 11(2), 267-299.
- [38] Zhang, M., Joyce, B. R., Sullivan, W. J., & Nussenzweig, V. (2013). Translational control in *Plasmodium* and *Toxoplasma* parasites. *Eukaryotic cell*, 12(2), 161-167.

- [39] Dupont, C. D., Christian, D. A., & Hunter, C. A. (2012, November). Immune response and immunopathology during toxoplasmosis. In *Seminars in immunopathology* (Vol. 34, No. 6, pp. 793-813). Springer-Verlag.
- [40] Montoya, J. G., & Liesenfeld, O. (2004). Toxoplasmosis. *The Lancet*, 363(9425), 1965–1976.
- [41] Howe, D. K., & Sibley, L. D. (1995). *Toxoplasma gondii* comprises three clonal lineages: correlation of parasite genotype with human disease. *Journal of infectious diseases*, 172(6), 1561-1566.
- [42] Saeij, J. P., Boyle, J. P., & Boothroyd, J. C. (2005). Differences among the three major strains of *Toxoplasma gondii* and their specific interactions with the infected host. *Trends in parasitology*, 21(10), 476-481.
- [43] Sibley, L. D., & Boothroyd, J. C. (1992). Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature*, 359(6390), 82.
- [44] Sibley, L. D., Khan, A., Ajioka, J. W., & Rosenthal, B. M. (2009). Genetic diversity of *Toxoplasma gondii* in animals and humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1530), 2749-2761.
- [45] Ramazi, S., Zahiri, J., Arab, S. S., & Parandian, Y. (2016). Computational prediction of proteins sumoylation: a review on the methods and databases. *J. Nanomed. Res*, 3(5).
- [46] Newman, H. A., Meluh, P. B., Lu, J., Vidal, J., Carson, C., Lagesse, E., ... & Matunis, M. J. (2017). A high throughput mutagenic analysis of yeast sumo structure and function. *PLoS genetics*, 13(2), e1006612.
- [47] Hay, R. T. (2005). SUMO: a history of modification. *Molecular cell*, 18(1), 1-12.
- [48] Mossessova, E., & Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Molecular cell*, 5(5), 865-876.

- [49] Flotho, A., & Melchior, F. (2013). Sumoylation: a regulatory protein modification in health and disease. *Annual review of biochemistry*, 82, 357-385.
- [50] Li, X. C., Zeng, Y., Sun, R. R., Liu, M., Chen, S., & Zhang, P. Y. (2017). SUMOylation in cardiac disorders-a review. *Eur Rev Med Pharmacol Sci*, 21(7), 1583-1587.
- [51] Hay, R. T. (2013). Decoding the SUMO signal. *Biochem. Soc. Trans.*, 41 (2013), 463-473
- [52] Braun, L., Cannella, D., Pinheiro, A. M., Kieffer, S., Belrhali, H., Garin, J., & Hakimi, M. A. (2009). The small ubiquitin-like modifier (SUMO)-conjugating system of *Toxoplasma gondii*. *International journal for parasitology*, 39(1), 81-90.
- [53] Crater, A. K., Roscoe, S., Fahim, A., & Ananvoranich, S. (2018). *Toxoplasma* ubiquitin-like protease 1, a key enzyme in sumoylation and desumoylation pathways, is under the control of non-coding RNAs. *International journal for parasitology*, 48(11), 867-880.
- [54] Soete, M., Camus, D., & Dubrametz, J. F. (1994). Experimental induction of bradyzoite-specific antigen expression and cyst formation by the RH strain of *Toxoplasma gondii* in vitro. *Experimental parasitology*, 78(4), 361-370.
- [55] Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, 31(13), 3406-3415.
- [56] Sherf, B. A., Navarro, S. L., Hannah, R. R., & Wood, K. V. (1996). Dual-luciferase reporter assay: an advanced co-reporter technology integrating firefly and Renilla luciferase assays. *Promega Notes*, 57(2).
- [57] Butler, J. E., & Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & development*, 16(20), 2583-2592.

- [58] Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., ... & Brunk, B. P. (2007). ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic acids research*, 36(suppl_1), D553-D556.
- [59] Soldati, D., & Boothroyd, J. C. (1995). A selector of transcription initiation in the protozoan parasite *Toxoplasma gondii*. *Molecular and cellular biology*, 15(1), 87-93.
- [60] Yamagishi, J., Wakaguri, H., Ueno, A., Goo, Y.K., Tolba, M., Igarashi, M., ... & Xuan, X., (2010). High-Resolution Characterization of *Toxoplasma gondii* Transcriptome with a Massive Parallel Sequencing Method. *DNA Research*, 17(4), 233-243.
- [61] Nakaar, V., Bermudes, D., Peck, K. R., & Joiner, K. A. (1998). Upstream elements required for expression of nucleoside triphosphate hydrolase genes of *Toxoplasma gondii*. *Molecular and biochemical parasitology*, 92(2), 229-239.
- [62] Mercier, C., Lefebvre-Van Hende, S., Garber, G. E., Lecordier, L., Capron, A., & Cesbron- Delauw, M. F. (1996). Common *cis*-acting elements critical for the expression of several genes of *Toxoplasma gondii*. *Molecular microbiology*, 21(2), 421-428.
- [63] Gaji, R. Y., Behnke, M. S., Lehmann, M. M., White, M. W., & Carruthers, V. B. (2011). Cell cycle-dependent, intercellular transmission of *Toxoplasma gondii* is accompanied by marked changes in parasite gene expression. *Molecular microbiology*, 79(1), 192-204.
- [64] Carey, M. F., Peterson, C. L., & Smale, S. T. (2013). The primer extension assay. *Cold Spring Harbor Protocols*, 2013(2), 164– 173.
- [65] Yeku, O., & Frohman, M. A. (2011). Rapid amplification of cDNA ends (RACE). *Methods Mol Biol*, 703,107–122
- [66] Behnke, M. S., Radke, J. B., Smith, A. T., Sullivan Jr, W. J., & White, M. W. (2008). The transcription of bradyzoite genes in *Toxoplasma gondii* is controlled by autonomous promoter elements. *Molecular microbiology*, 68(6), 1502-1518.

- [67] Yoon, J. H., Kim, J., & Gorospe, M. (2015). Long noncoding RNA turnover. *Biochimie*, *117*, 15-21.
- [68] Jin, Y., Chen, Z., Liu, X., & Zhou, X. (2013). Evaluating the microRNA targeting sites by luciferase reporter gene assay. *Methods Mol. Biol*, *936*, 117-127.
- [69] Clément, T., Salone, V., & Rederstorff, M. (2015). Dual luciferase gene reporter assays to study miRNA function. *Methods Mol Biol*, *1296*, 187-198
- [70] Dominguez, A. A., Lim, W. A., & Qi, L. S. (2016). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature reviews Molecular cell biology*, *17*(1), 5.
- [71] Goyal, A., Myacheva, K., Groß, M., Klingenberg, M., Duran Arqué, B., & Diederichs, S. (2017). Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. *Nucleic acids research*, *45*(3), e12-e12.
- [72] Napoli, S., Piccinelli, V., Mapelli, S. N., Pisignano, G., & Catapano, C. V. (2017). Natural antisense transcripts drive a regulatory cascade controlling c-MYC transcription. *RNA biology*, *14*(12), 1742-1755.
- [73] Hossain, A. (2017) Investigation of the argonaute protein variants in *Toxoplasma gondii* and the contribution of argonaute to RNA silencing (Masters thesis). University of Windsor, Windsor, Canada.
- [74] Skariah, S., McIntyre, M. K., & Mordue, D. G. (2010). *Toxoplasma gondii*: determinants of tachyzoite to bradyzoite conversion. *Parasitology research*, *107*(2), 253-260.
- [75] Haussecker, D., & Proudfoot, N. J. (2005). Dicer-dependent turnover of intergenic transcripts from the human β -globin gene cluster. *Molecular and cellular biology*, *25*(21), 9724-9733.

APPENDICES

Appendix A

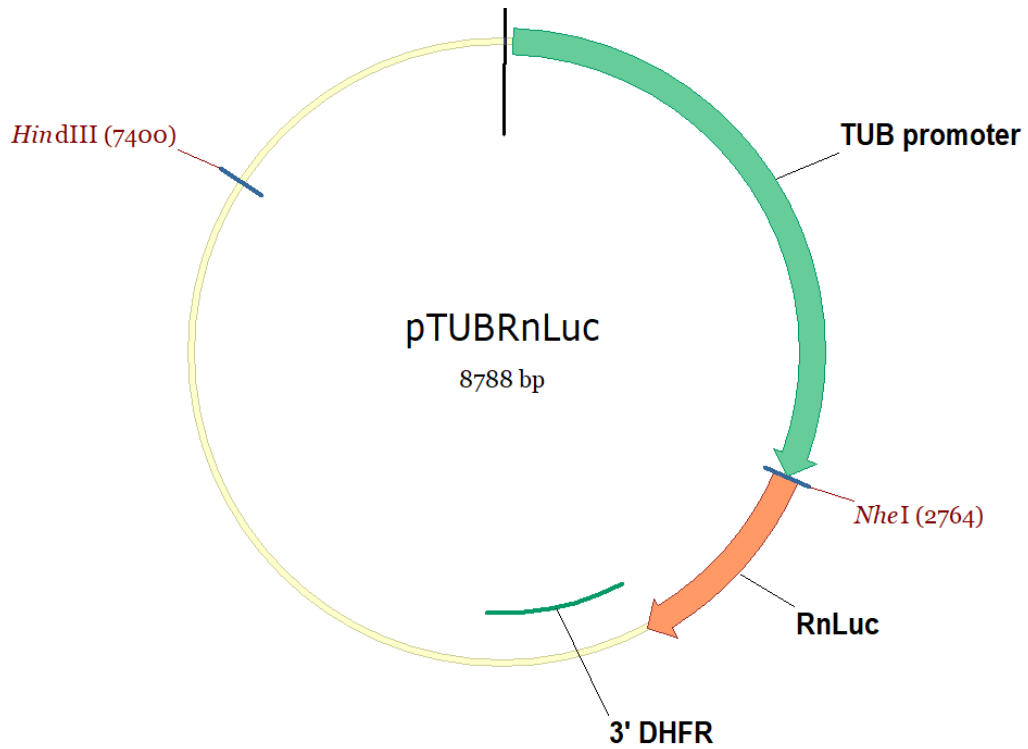


Figure A1 pTUBRnLuc

This diagram shows the plasmid coding for RnLuc under the control of TUB promoter. It was used as a positive control in dual luciferase assays for both identifying promoters and gene silencing. pTUBRnLuc was digested with NheI and HindIII to remove the tubulin (TUB) promoter (2717bps). The linearized product was ligated to putative promoter sequences that were amplified by PCR.

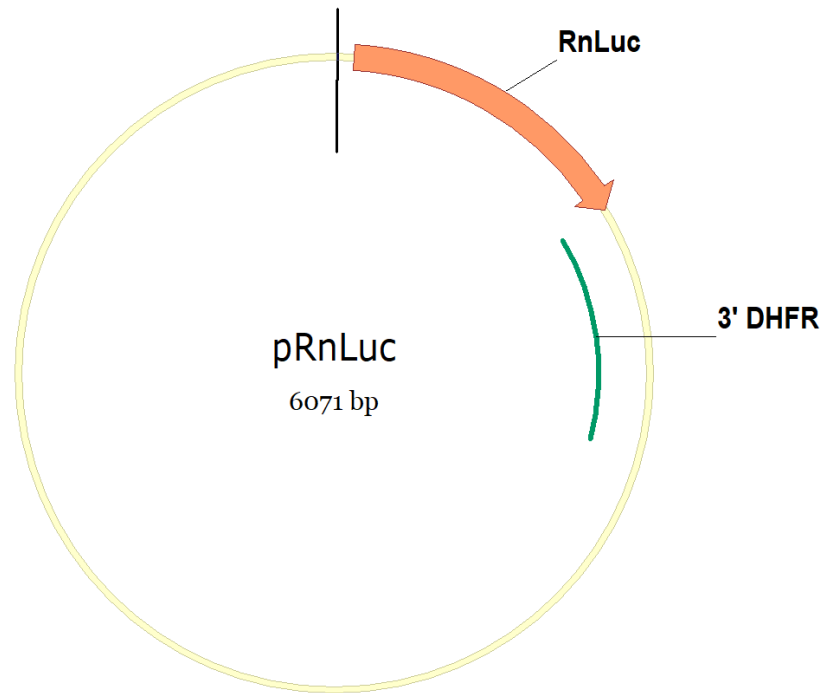


Figure A2 pRnLuc

This diagram shows the plasmid coding for RnLuc without a promoter. It was used as a negative control for background expression in dual luciferase assays for identifying promoters.

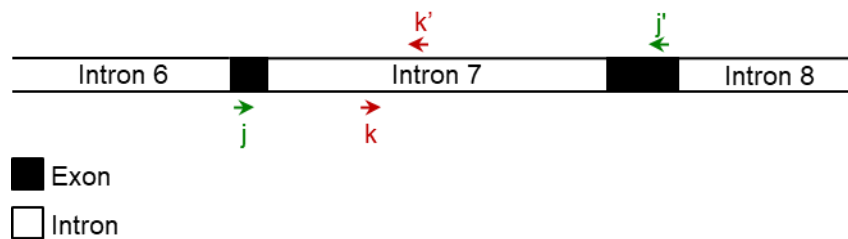


Figure A3 Diagram Showing Primers Used for *TgUlp1* RT-qPCR.

Schematic diagram of intron 6 to intron 8 from the *TgUlp1* locus. Arrows indicate location of primer binding site. Green arrows are *TgUlp1* mRNA primers and red arrows are *TgUlp1* NAT primers used for RT-qPCR. Primer sequences are listed in Table B4.

Appendix B

Table B1 Primers Used to Construct Plasmids For The Promoter Assay

Digestion sites are highlighted in red

Primer Name	Sequence
pr_RV_senseUlp1 (a)	caccttggagccat gctagc ggccacacgagaggggaaaag
pr_FW-senseUlp1 (b)	gtcgacggtatcgat aagctt ctgcatcggttgcgcct
pr_FW-senseUlp1p1k (c)	gtcgacggtatcgat aagctt caaccgagcgggtctggc
pr_FW-senseUlp1p2.5k (d)	gtcgacggtatcgat aagctt agagcagaagagggagcc
pr_RV_antisenUlp (e)	caccttggagccat gctagc gaccttcttcttcaacac
pr_RV_antisenUlpAt1k (f)	caccttggagccat gctagc gaggtgagacaatggatc
pr_FW-antisenUlp0.5k (g)	gtcgacggtatcgat aagctt gatcattgtctcacctc
pr_FW-antisenUlp1k (h)	gtcgacggtatcgat aagctt cgttgcagaacttgcg
pr_FW-antisenUlp2k (i)	gtcgacggtatcgat aagctt gaggtcaaatgaccacgg

Table B2 pPutPromRnLuc Used in the Promoter Assay

Promoter	Plasmid Name
0.5-kb SS	pRnpromoterUlp0.5k_SS
1.0-kb SS	pRnpromoterUlp1k_SS
2.5-kb SS	pRnpromoterUlp2.5k_SS
0.5-kb AS	pRnpromoterUlp0.5k_NAT
1.0-kb AS	pRnpromoterUlp1.0k_NAT
2.5-kb AS	pRnpromoterUlp2.5k_NAT
2.0-kb AS	pRnpromoterUlp2k_NAT

Table B3 Primers Used to Make Templates For *in Vitro* RNA

T7 promoter sequence for *in vitro* transcription are highlighted in red

Primer Name	Sequence
TgUlp1intron6Fw	ggctgaacgacgaagttatc
TgUlp1intron7T7Rv	taatacgaactcactataggaacagaggcgaagtcgtaggt
TgUlp1intron6T7Fw	taatacgaactcactataggctgaacgacgaagttatc

Table B4 Primers Used for RT-qPCR

Primer Name	Sequence
RT_PCR Ulp1_Fw (j)	cgtaacaagaagcaacgcgc
RT_PCRUlp1_Rv (j')	cgaacagaggcgaagtcgta
qAntiUlp1Fw (k)	tgggcgaagacggagaaga
qAntiUlp1Rv (k')	ttccaggtgaccgtaacatgtg
qPCR_GAPDH_Fw	ggtgttccgtgctgcat
qPCR_GAPDH_Rv	gcctttccgccgacaat

Table B5 Amount of Plasmid Electroporated Based on Experiment

	Plasmid	µg
Establishing Promoter Assay	pRnLuc	5
	p2.5kbSSRnLuc	5
	pTUBRnLuc	5
		0.5
		0.05
		0.005
		0.0005
<i>In vitro</i> NAT Electroporation	pTUBRnLuc	5
	pTUBRnLuc_SiteA	5
	pTUBRnLuc_SiteB	5

Table B6 Initiator Element Identified in *TgUlp1* Promoters

Identified *TgUlp1* sense and antisense promoter sequences were analyzed by ElemeNT for Initiator Element (Inr), YYANWYY, where Y = C/T, W = A/T, N = A/C/G/T. Only matches with PWM score >0.2 and complete consensus match are displayed.

<i>TgUlp1</i> Promoter	Sequence	<i>TgUlp1</i> Locus	PWM Score
1.0-kb mRNA Promoter	5' TG CCACTTC TCGTTGCGTTC 3' 3' ACGGTGAAGAGCAACGCAAG 5'	5'-Flanking (-469 to -450)	1.0000
	5' GAGAGAAGTTG TCACATC TG 3' 3' CTCTCTTCAACAGTGTAGAC 5'	5'-Flanking -1019 to -1000	0.3636
	5' GCTTTG TCACTTT TCCCCTC 3' 3' CGAAACAGTGAAAAGGGGAG 5'	5'-Flanking -29 to -10	0.4800
2.0-kb NAT Promoter	5' TATG GGTCTGA GTCGCGCGT 3' 3' ATACCCAGACTCAGCGCGCA 5'	Intron 9 (+6051 to +6070)	0.3410
	5' CTC GGATTGA GTGCCTCGCG 3' 3' GAGCCTAACTCACGGAGCGC 5'	Intron 10 +6211 to +6230	0.6499
	5' TGT GGTTTGA AAGTCCCCTACT 3' 3' ACACCAAACCTTTCAGGGTGA 5'	Intron 10 +6511 to +6530	0.2954
	5' CATTGA AGAGTGG TGCATCC 3' 3' GTAACCTTCTCACCACGTAGG 5'	Exon 11 +6581 to +6600	0.6000
	5' GG GAAGTGAA CAAGGACGGG 3' 3' CCCTTCACTTGTTCCCTGCCC 5'	Intron 11 +6701 to +6720	0.8000
	5' GTG GGTGTGG AACCAGGTGT 3' 3' CACCCACACCTTGGTCCACA 5'	Intron 11 +6881 to +6900	0.4545
	5' GG AGAGTGA CAAGCGAAGAG 3' 3' CCTCTCACTGTTTCGCTTCTC 5'	3'-Flanking +7441 to +7460	0.4800
	5' GAGAAAA AGAATGA AGCAG 3' 3' CTCTTTTTTCTTACTTCGTC 5'	3'-Flanking +7501 to +7520	0.3600
5' AC GAAGTGA CACGGGCCGGG 3' 3' TGCTTCACTGTGCCCGGCC 5'	3'-Flanking +7721 to +7740	0.8000	

Table B7 GAGACG Motif Identified in *TgUlp1* Promoters

Sequence from identified *TgUlp1* sense and antisense promoters were analyzed for GAGACG motif using Vector VNTI.

<i>TgUlp1</i> Promoter	Sequence	<i>TgUlp1</i> Locus
1.0-kb mRNA Promoter	5' AAGCGTG GAGACG CAGAGAA 3' 3' TTCGCACCTCTGCGTCTCTT 5'	3'-Flanking -229 to -210
2.0-kb NAT Promoter	5' AGTCGCGCGTCTCGGTTCTC 3' 5' TCAGCGC GCAGAG CCAAGAG 3'	Intron 9 +6061 to +6080
	5' GATCGTGTGCGTCTCGAGCG 3' 5' CTAGCACAC GCAGAG CTCGC 5'	Intron 10 +6421 to +6440
	5' TTTTCGCGTCTCTTCGGTGAT 3' 3' AAAGC GCAGAG AAGCCACTA 5'	Intron 10 +6451 t + 6470
	5' CTTTTGCTCGCGTCTCCGGA 3' 3' GAAAACGAGC GCAGAG GCCT 5'	3'-Flanking +7641 to +7660

Table B8 TGCATGC Motif Identified in *TgUlp1* Promoters

Sequence from identified *TgUlp1* sense and antisense promoters were analyzed for TGCATGC motif using Vector VNTI.

<i>TgUlp1</i> Promoter	Sequence	<i>TgUlp1</i> Locus
1.0-kb mRNA Promoter	N/A	N/A
2.0-kb NAT Promoter	5' TCTCTGCATGCATATCGCTC 3' 3' AGAGA CGTACGT ATAGCGAG 5'	Intron 9 +6011 to + 6030
	5' GGAAGCATGCACAGAGAAG 3' 3' CCCTT CGTACGT GTCTCTTC 5'	Intron 10 +6281 to +6300
	5' GCTTCCGCGCATGCAGGCTC 3' 3' CGAAGGCG CGTACGT CCGAG 5'	Exon 13 +7381 to + 7400

VITA AUCTORIS

NAME: Ambreen Fahim

PLACE OF BIRTH: Karachi, Pakistan

YEAR OF BIRTH: 1995

EDUCATION: Vincent Massey Secondary School
Windsor, ON
2009-2013

University of Windsor, Windsor, ON
2013-2017
B.Sc. Biochemistry and Biotechnology

University of Windsor, Windsor, ON
2017-2019
M.Sc. Chemistry and Biochemistry