DISSECTING MECHANISMS OF PURINE-RESPONSIVE GENE EXPRESSION THROUGH THE LENS OF THE *LEISHMANIA DONOVANI* PURINE TRANSPORTERS

By

M. Haley Licon

A DISSERTATION

Presented to the Department of Molecular Microbiology and Immunology
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

October 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ALBA, acetylation lowers binding affinity

BSD, blasticidin resistance gene

Ch1, chromosome 1

CDS, coding sequence

CU, polypyrimidine

DME-L, Dulbecco's Modified Eagle-Leishmania medium

eIFs, eukaryotic initiation factors

ENT, equilibrative nucleoside transporter

FAPS, fluor

Fluc, firefly luciferase

fPPG, filamentous proteophosphoglycans

GMPR, GMP reductase

IGR, intergenic region

IMPDH, IMP dehydrogenase

LdNT1, *L. donovani* purine nucleoside transporter 1

LdNT2, *L. donovani* purine nucleoside transporter 2

LdNT3, *L. donovani* purine nucleobase transporter 3

LdNT4, *L. donovani* purine nucleobase transporter 4

LGN, Leishmania Genome Network

LmTUB, L. major α-tubulin

LPG, lipophosphoglycan

NEO, neomycin resistance gene

NLuc, Nanoluciferase

nt, nucleotide

ORF, open reading frame

P-body, processing body

PAC, puromycin resistance gene

PFGE, pulsed field gel electrophoresis

Phleo, phleomycin resistance gene

PIC, pre-initiation complex

Pol I, RNA polymerase I

Pol II, RNA polymerase II

Pol III, RNA polymerase III

PM, peritrophic matrix

PSG, promastigote secretory gel

PUF, Drosophila Pumilio and *Caenorhabditis elegans* Fem-3-binding factor

PV, parasitophorous vacuole

RBD, RNA-binding domain

RBP, RNA-binding protein

Rluc, Renilla luciferase

RNA-FISH, RNA fluorescence in situ hybridizaton

RNP, ribonucleoprotein

RRM, RNA-recognition motif

rRNA, ribosomal RNA

RT-qPCR, Reverse transcription quantitative PCR

SA, splice-acceptor site

SIDER, short interspersed degenerated retroposon

SL, spliced leader

TbNT8.1, *T. brucei* purine nucleobase transporter 8.1

TSS, transcription start site

term, rRNA terminator sequence

UE1, LdNT1 upstream element

UIS, upregulated in infectious sporozoites

UMPS, UMP synthase

UTR, mRNA untranslated region

WT, wildtype

2A, *Thosea asigna* virus 2A peptide

3’ RACE, 3’ rapid amplification of cDNA ends

4E-BP, eIF4E binding-protein

4E-IP, eIF4E interacting-protein

ACKNOWLEDGMENTS

I would first like to express my sincere gratitude to my research mentor, Dr. Phillip Yates, whose genuine curiosity and enthusiasm for the scientific process have been a source of constant inspiration. He gave me the space to grow as an independent thinker and scientist, yet he was never too busy to spitball ideas, edit a rough draft, or workshop the design of an experiment. Through his example, Phil showed me how to approach research with humility, creativity, and (on more than one occasion) resilience in the face of adversity. I am incredibly to fortunate to have had his mentorship throughout this journey.

I would like to thank my faculty advisors, Drs. Scott Landfear and Buddy Ullman, as well as my committee members, Dr. Georgiana Purdy, Dr. Matthew Thayer, Dr. Michael Cohen, and Dr. Timothy Nice. All of these individuals have generously given their time and consideration to help make this project as successful as possible. Dr. Purdy in particular has been a constant mentor, advisor, and all-around academic guru since my matriculation at OHSU. For this, I am deeply grateful.

Special thanks to Dr. Stefanie Kaech Petrie and Dr. Leslie Smith, for lending their vast technical expertise with fluorescence microscopy and RNA-FISH, as well as to Dr. Robyn Kent and Corinne Fargo, who allowed me to impose on their friendship and generosity for editing support.

Many colleagues and friends have enriched my life over the last six years. Rather than attempt to list them all by name, I offer a general (but heartfelt) admission of gratitude. The relationships and memories that I take away from this period of my life will be forever cherished.

Finally, to my parents, Val and Melissa Licon: thank you for instilling in me your love of adventure and curiosity about the natural world. You filled our home with music, books, and art, and crammed my childhood with more outdoor epics than most people experience in a lifetime. Whether intentional or not, you taught me to always question the boundaries of what I thought myself capable and I would not be where I am today without your example to aspire to.

Dissecting mechanisms of purine-responsive gene expression through the lens of the *Leishmania donovani* purine transporters

M. Haley Licon
Oregon Health and Science University, 2020

To progress through their normal developmental cycle, *Leishmania* and related kinetoplastid parasites must sense and respond to extreme fluctuations in the host environment. Gene expression in kinetoplastids is divergent from other eukaryotes in that transcription is not controlled on an individual gene basis. Consequently, post-transcriptional control points such as mRNA stability and translation have been elevated as key determinants of regulation. As with higher eukaryotes, these processes are governed primarily by the interactions of *cis*-acting elements in mRNA and *trans*-acting RNA-binding proteins (RBPs). Considerable effort has therefore been devoted to the study of such elements in kinetoplastids; however, few have been implicated in the context of specific stress-response pathways.

Leishmania are obligate purine scavengers, relying upon the host to provide these essential nutrients. As such, they have evolved a robust stress response to cope with periods of purine scarcity that may arise over the course of their lifecycles. In *Leishmania donovani*, purine salvage is accomplished by four nucleoside and nucleobase transporters, LdNTs 1-4. We have shown that the *L. donovani* proteome is dramatically remodeled under purine stress, with LdNTs 1-3 ranking among the most substantially upregulated

proteins. Thus, to examine mechanisms of purine-responsive gene expression, we have used the purine transporters as a model.

In the studies described here, we performed a molecular dissection of the *LdNT* mRNA untranslated regions (UTRs) to identify the *cis*-acting elements within. For *LdNT3*, we found that a 33 nt stem-loop in the mRNA 3'-UTR serves to repress expression when purines are abundant, both via transcript destabilization and translational downregulation. In *LdNT2*, regulation was attributed to a 76 nt-long polypyrimidine tract in the 3'-UTR. Loss of this region resulted in a drastic reduction in transcript abundance and prevented translational enhancement under purine stress. Finally, we determined that *LdNT1* mRNA abundance and translation are independently controlled by adjacent *cis*-acting features, both of which are required for upregulation in response to purines. Of note, the three purine-response elements described here are the first to be identified in *Leishmania*. Their characterization points to an unexpected level of complexity in the regulation of the purine stress response and sets the stage for future efforts to identify the network of RNA-protein and protein-protein interactions involved.

Chapter 1.

Introduction

Leishmania and the leishmaniases

Leishmaniasis is a neglected tropical disease caused by the kinetoplastid parasite *Leishmania* and transmitted by blood-feeding female sand flies. The disease is endemic throughout much of the tropics, subtropics, and Mediterranean basin, where it disproportionately affects poor and marginalized communities [W.H.O., 2010]. Different *Leishmania* species cause different clinical manifestations, ranging from mild cutaneous lesions to life-threatening infection of the visceral organs. According to the WHO, these parasites affect roughly 12 million people worldwide and annually kill ~30,000, making leishmaniasis one of the planet's most dangerous infectious diseases, surpassed only by malaria in parasite-induced fatalities [Alvar, 2012].

Parasites, reservoirs, and vectors

Kinetoplastid protozoans (Trypanosomatida family) are a diverse group of uniflagellates that diverged from the opisthokonts early in eukaryotic evolution. The major distinguishing feature of these organisms is the kinetoplast, a large DNA-containing structure located inside a single mitochondrion. Among this group are both free-swimming and parasitic representatives, including *Leishmania* and the related human pathogens *Trypanosoma brucei* (African Sleeping Sickness) and *Trypanosoma cruzi*

(Chagas Disease) [Akhoundi, 2016]. Together referred to as the 'tritryps', these parasites have garnered much attention from the scientific research community for their medical and veterinary importance.

Over 20 species of *Leishmania* are known to infect humans. These are broadly classified into two groups based on geographic distribution. In the Old World (Eastern Hemisphere), relevant species include *Leishmania tropica, L. major, L. aethiopica, L. infantum,* and *L. donovani*. New World (Western Hemisphere) parasites fall primarily into either the *L. mexicana* complex (*L. mexicana, L. amazonensis, L. venezuelensis*) or the subgenus *Viannia (L. [V.] braziliensis, L. [V.] guyanensis, L. [V.] panamensis, L. [V.] peruviana*), though *L. infantum* (called *L. chagasi* in Latin America) is also present. With the exceptions of *L. donovani* and *L. tropica*, *Leishmania* are not exclusively human parasites; over 70 animal species are natural hosts for *Leishmania* [Alemayehu, 2017]. This is important from the standpoint of eradication, as both wild and domestic reservoirs are capable of perpetuating the cycle of zoonotic transmission to humans.

Endemicity of the *Leishmania* species and the diseases they cause is dependent upon the presence of competent vectors. *Leishmania* transmission by infected sand flies has been known since the early 20th century [Steverding, 2017]. Only female sand flies spread leishmaniasis, as blood feeding is required for egg development. Of the five sand fly genera, only *Lutzomyia* (New World) and *Phlebotomus* (Old World) are proven carriers of *Leishmania.* Most vector species studied to date are considered 'permissive,' in that they support the development of multiple *Leishmania* strains. However, there are several *Leishmania* species that demonstrate a close evolutionary fit with their vector, such as *L. tropica* with *Phlebotomus sergenti* or *L. major* with *P. papatasi* and *P.*

duboscqi [Dostálová, 20120]. At present, leishmaniasis is principally restricted to tropical and subtropical regions. However, as human forces such as climate change, urbanization and deforestation shape global ecology, the geographic range of disease is likely to expand as new areas become amenable to the sand fly vector [Alemayehu, 2017].

Disease in humans

There are three main clinical forms of leishmaniasis: cutaneous, mucocutaneous, and visceral. The pathology and severity of these syndromes can vary significantly depending on interactions between the host immune response, parasite species, and environmental factors.

Cutaneous leishmaniasis (CL) is the mildest and most common form of leishmaniasis, presenting as one or more painful ulcer(s) in the skin. Infection is established when parasites invade macrophages in the dermis. Over the course of weeks to months, lesions develop at or near the inoculation site, appearing first as raised papules and slowly ulcerating over time [Burza, 2018]. While these lesions are typically self-healing in immunocompetent individuals, they can result in permanent scarring and/or deformation of the affected tissue. Around one third of recovered patients will relapse, displaying pathologies similar to the original infection [Torres-Guerrero, 2017]. In rare cases, CL can progress to more serious manifestations such as disseminated or diffuse (anergic) cutaneous leishmaniasis, caused by *L. [V] braziliensis* and *L.* amazonensis, respectively. These forms are not spontaneously healing and respond poorly to treatment [Torres-Guerrero, 2017; Machado, 2019].

In South America, 1-10% of CL patients who are asymptomatic or have resolved cutaneous lesions develop a second, more debilitating condition known as mucocutaneous leishmaniasis (MCL) [Burza, 2018]. MCL is characterized by destructive lesions of the nasopharyngeal cavity that form as a result of a hyperinflammatory immune response. Invasion of the oral and nasal mucosa is relatively slow, with lesions first appearing months to years after the primary CL infection has healed [Torres-Guerrero, 2017]. However, MCL lesions are not self-healing and can lead to severe facial disfigurement or death if not treated rapidly [W.H.O., 2010]. Most cases of MCL are caused by *L. [V] braziliensis,* though other members of the *Viannia* subgenus are also involved.

Commonly known as kala-azar (Hindi for 'black fever'), visceral leishmaniasis (VL) constitutes the third and most dangerous form of leishmaniasis. In 2018, over 95% of new VL cases occurred in just 10 countries, with the majority reported from Brazil, East Africa, and India [W.H.O., 2010]. VL is caused by members of the *L. donovani* complex - *L. donovani* in Asia and Africa and *L. infantum* in the Mediterranean Basin and South America [Chappuis, 2007]. These parasites disseminate throughout the body via the reticuloendothelial system, infiltrating the bone marrow, spleen and liver. As a result, VL commonly presents with persistent irregular fevers, weight loss, and spleno- and hepatomegaly. VL is typically fatal within two years if left untreated [Burza, 2018].

Disease management

As there is no effective vaccine to preempt transmission of *Leishmania* parasites, disease management is predicated exclusively on chemotherapy. A handful of drugs are

currently used to treat leishmaniasis, each demonstrating unique benefits, limitations, and variable efficacy against different clinical forms of the disease.

Pentavalent antimonials such as sodium stibogluconate (Pentostam) and meglumine antimoniate (Glucantime) have served as the first line of defense against leishmaniasis for over 70 years. The mechanism of action of these compounds is not well understood, though it is probably related to their interaction with thiol-containing molecules, particularly those involved in defense against oxidative stress [Haldar, 2011]. When used appropriately, antimonials are effective against all three clinical forms of leishmaniasis; however, these drugs frequently cause severe adverse reactions, forcing patients to abandon treatment before the parasite is cleared. Moreover, the dosing regimen is long, requiring daily intramuscular injection over many weeks, and treatment compliance failure is common. These factors, compounded by decades of general misuse, have contributed to the emergence of antimony resistance in hyperendemic areas [Frezad, 2009; Chakravarty, 2010]

Second-line drugs include pentamidine and the antifungal agent Amphotericin B (AmpB). Like the antimonials, the mechanism of action for pentamidine is not precisely known in *Leishmania,* though it is thought to disrupt kinetoplast replication by binding to the kinetoplast DNA (kDNA) [Yang, 2016]. Resistance and toxicity have all but ended the use of pentamidine for VL, though it is still employed with variable success against CL and MCL caused by both Old and New World *Leishmania* species [Chakravarty, 2010]. AmpB binds to ergosterol in the *Leishmania* plasma membrane, leading to pore formation and ion leakage [Stone, 2016]. On the Indian subcontinent where antimony resistance is prevalent, AmpB has been elevated as a first-line treatment for VL

[Chakravarty, 2010; Mishra, 1992]; however, more widespread use is limited due to low bioavailability and acute toxicity [Ghorbani, 2018]. A liposomal AmpB formulation (AmBisome®) has proven to be equally as effective and minimally toxic but is unavailable in the developing world due to its exorbitant price. Additionally, AmBisome treatment involves long-term IV administration and constant patient monitoring, both of which are challenging in endemic areas where the necessary medical infrastructure can be lacking [Wortmann, 2010].

The third-line drug miltefosine was originally developed as an anti-cancer agent but was later found to have strong trypanocidal activity *in vitro*. Early clinical trials with miltefosine yielded promising results and the compound was approved in India in 2002 as the first orally available antileishmanial [Ganguly, 2002]. Miltefosine is widely heralded for its relatively mild adverse effects and high efficacy against VL. When administered at the recommended dose, cure rates as high as 95% have been reported in adults [Sundar, 2000; Sunyoto, 2018]. However, like AmBisome®, miltefosine is prohibitively expensive in most *Leishmania* endemic countries and the majority of patients are forced to defer to cheaper antimonial compounds despite their relative toxicity [Sunyoto, 2018]. In addition, concern has been raised over the possible teratogenicity and long half-life of the drug, which could potentially encourage clinical resistance even when used appropriately [Chakravarty, 2010; Ghorbani, 2018].

In summary, *Leishmania* parasites are responsible for widespread human suffering and death throughout much of the developing world. While there are a handful options for treating leishmaniasis, they are far from satisfactory. Existing drugs are plagued by high cost, toxic side effects, difficult routes of administration and resistance is

steadily increasing. Thus, new leishmanicidal compounds are desperately needed. Exploring the unique aspects of kinetoplast biology may reveal pathways and targets that are amenable to therapeutic exploitation.

The *Leishmania* lifecycle: An exercise in adaptation

Leishmania are digenetic organisms, requiring passage through both an insect vector and a vertebrate host to complete their developmental cycle (Figure 1.1A). The cycle begins when a female sand fly ingests infected host cells during blood feeding. Intracellular parasites (i.e. amastigotes) emerge into the midgut lumen and rapidly differentiate into flagellated, extracellular promastigotes. Over the next several days, these parasites undergo a complex series of morphological transitions, with each form playing a specific role and occupying a distinct compartment within the midgut (Figure 1.1 B). Intravector development culminates in the maturation of infective metacyclic forms, which are deposited into the skin of a new vertebrate host when the sand fly takes her next blood meal. Though other mononuclear phagocytes, particularly neutrophils, have been implicated during early stages of infection, macrophages are the host cell of choice for *Leishmania.* Parasites that gain access to the macrophage interior are funneled into the host endocytic pathway and eventually reside in a membrane-bound structure called the parasitophorous vacuole (PV). Within PVs, internalized metacyclic parasites undergo a final transition back to tissue-resident amastigotes that multiply and metastasize to cause pathology.

The physiological conditions within the vertebrate host and the sand fly vector differ significantly and each presents unique set of challenges that *Leishmania* must

overcome for colonization (Figure 1.1A). The PV, for instance, is a hostile compartment derived from the macrophage phagolysosome. As such, its interior is characterized by an acidic pH, temperatures hovering around 37°C, and an abundance of lysosomal acid hydrolases [Antoine, 1990; Sturgill-Koszycki, 1994; Antione, 1998]. The environment within the sand fly is ambient by comparison, with increased pH and lower temperature, though parasites still face sustained proteolytic attack from the digestive enzymes of the vector [Santos, 2008; Dostálová, 2012]. Conditions also fluctuate within an individual host over the course of infection. For example, intracellular amastigotes may experience a range of temperatures as host cells migrate throughout the body [Sunter, 2017]. In the midgut, the availability of salvageable nutrients changes progressively as the bloodmeal is digested, absorbed, and eventually excreted by the sand fly. At all stages, *Leishmania* parasites are subjected to the immunological reactions of their respective hosts [Dostálova, 2012].

As a consequence of this unique lifestyle, pathways involved in sensing and responding to environmental stress are of preeminent importance in *Leishmania*. Indeed, adaptation is so intimately intertwined with the *Leishmania* lifecycle that several stressors are themselves triggers for differentiation, signaling the transition from one developmental stage to the next. Although we still know very little of how *Leishmania* detect environmental change and translate that signal to an intracellular response, examination of these essential pathways is expected to reveal new, parasite-specific targets for drug design. In this section, I describe the *Leishmania* lifecycle in detail, highlighting the important role played by stress in lifecycle progression and development.

Figure 1.1

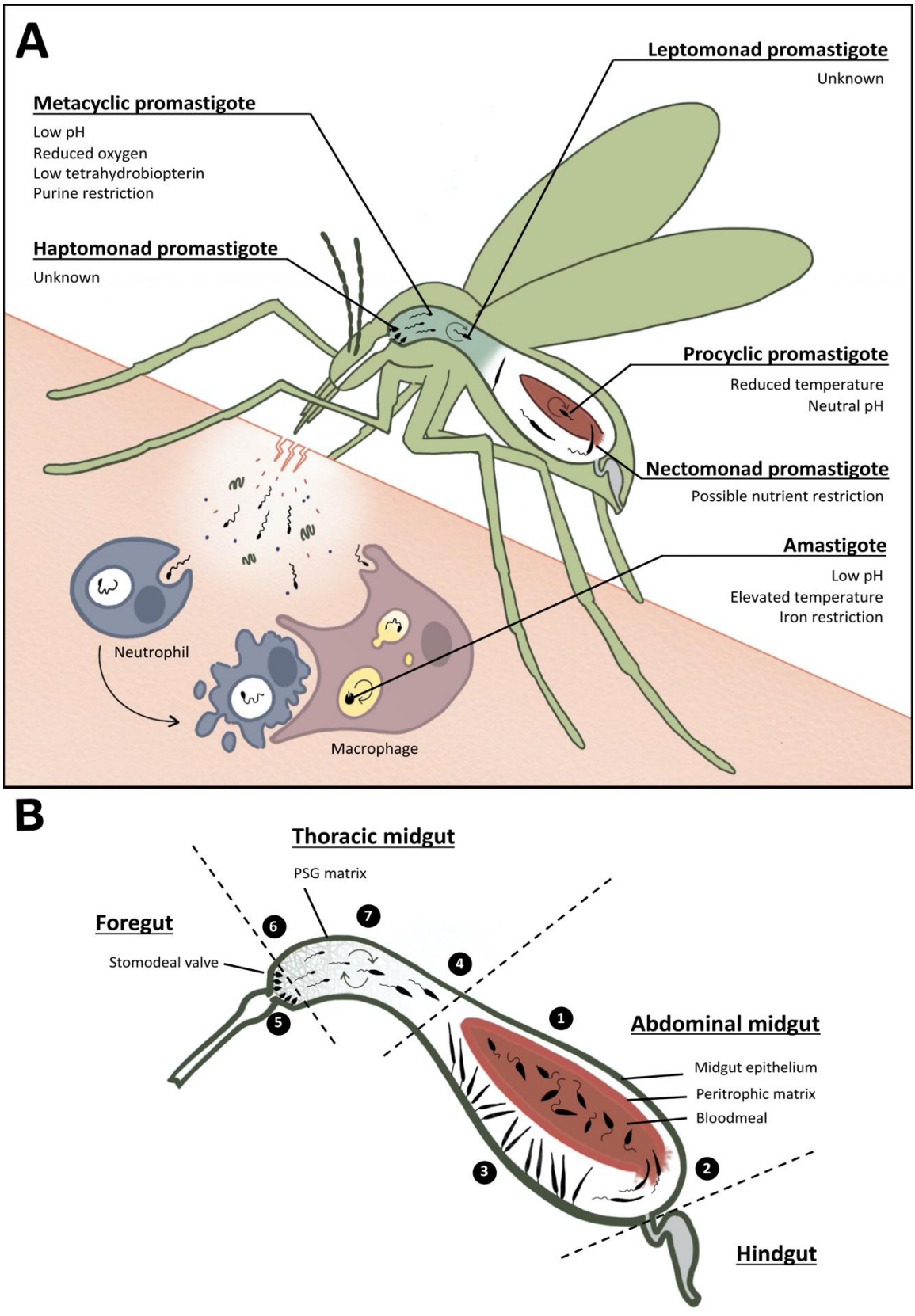

A
Leptomonad promastigote
Unknown
Metacyclic promastigote
Low pH
Reduced oxygen
Low tetrahydrobiopterin
Purine restriction
Haptomonad promastigote
Unknown
Procyclic promastigote
Reduced temperature
Neutral pH
Nectomonad promastigote
Possible nutrient restriction
Amastigote
Low pH
Elevated temperature
Iron restriction
Neutrophil
Macrophage
B
Thoracic midgut
PSG matrix
Foregut
Stomodeal valve
Abdominal midgut
Midgut epithelium
Peritrophic matrix
Bloodmeal
Hindgut
1
2
3
4
5
6
7

Figure 1.1: The *Leishmania* lifecycle. A) Overview of *Leishmania* development in both the insect and vertebrate host. Individual lifecycle stages are labeled with their respective differentiation triggers. Proliferative forms are indicated with a circular arrow. B) For emphasis, the major events of intravector development (numbered below) are shown in greater detail. Additional information regarding stage identification and the timeline of development is provided in Table 1.1.

Sand flies become infected with *Leishmania* by ingesting amastigote-containing macrophages in the bloodmeal. Likely triggered by a combination of pH and temperature change, intracellular amastigotes egress from host cells and differentiate to procyclic promastigotes in the midgut lumen (1). The bloodmeal is encased in a chitinous peritrophic matrix (PM). Towards the end of digestion, procyclic forms become nectomonad promastigotes which traverse the PM and migrate in the anterior direction (2). Although the stimulus for this transition is not definitively known, nectomonad development is suspected to reflect an adaptive response by the parasite to a dwindling blood nutrient pool. To avoid excretion with the digested bloodmeal, nectomonads attach to the midgut epithelium (3) before continuing their anterior migration. In the thoracic midgut nectomonads give rise to leptomonad promastigotes (4), the main replicative stage within the vector and the source of promastigote secretory gel (PSG). Around the same time, haptomonads emerge and attach to the stomodeal valve (5), damaging the structure to promote reflux of infectious parasites at the next blood feeding. Differentiation signals are not known for either stage. To prepare for transmission, leptomonads give way to infective metacyclic promastigotes (6); however, this transformation can be reversed by sequential bloodmeals (7). Although metacyclogenesis is attributed to a broad range of exogenous stressors, only purine restriction has been shown to promote metacyclic differentiation *in vivo*.

Metacyclic promastigotes are injected into the skin of a new vertebrate host during blood feeding. Neutrophils, followed later by macrophages, are recruited to the infection site by chemical signals released as part of the host wound healing response, PSG, and other components of the infectious inoculum. Metacyclic forms enter host macrophages either via uptake of infected, apoptotic neutrophils (i.e. the Trojan Horse model) or by directly engaging the macrophage surface for receptor-mediated phagocytosis. Within, parasites inhabit a parasitophorous vacuole (PV) derived from the phagolysosomal compartment of the host cell. The cycle is completed when internalized metacyclic parasites differentiate back to the amastigote form, triggered primarily by the low pH and elevated temperatures that characterize the PV.

Life within the sand fly

Intravector development is a complex process involving multiple differentiation events. For the majority of *Leishmania* species, this sequence occurs entirely within the confines of the sand fly midgut [Dostálová, 2012]. There are five major vector-resident forms: the procyclic promastigote, the nectomonad promastigote, the leptomonad/

retroleptomonad promastigote, the haptomonad promastigote, and the metacyclic promastigote [Rogers, 2002; Serafim, 2018]. All display a similar morphology with an elongated, tapered cell body and single apical flagellum [Sunter, 2017].

Historically, detailed examination of these stages has proven technically challenging. With the exception of the procyclic promastigote, there are no established methods for maintaining pure cultures of the various midgut forms *in vitro*. Moreover, their identification is based solely on subtle morphological differences; there are no molecular markers for the distinct stages [Sunter, 2017; Dostálová, 2012]. Nonetheless, each is thought to represent a response to a specific microenvironment encountered within the sand fly midgut. Based on their spatial and temporal distribution, the various promastigote forms have been placed in specific developmental sequence with precursor-product relationships between them [Rogers, 2002; Gossage, 2003]. For a concise summary of promastigote morphologies, timing, and localizations, refer to Table 1.1.

Sand flies become infected by ingesting amastigote-containing macrophages during blood feeding. Ingestion of the bloodmeal causes numerous changes in the sand fly, including formation of the peritrophic matrix (PM). Composed of chitin and glycoproteins secreted by the intestinal epithelium, the PM is a fibrillar structure that surrounds the bloodmeal, protecting the sand fly from its contents while still allowing limited diffusion of digestive enzymes [Secundino, 2005]. Within 12-18 hours, ingested amastigotes give rise to the first of the midgut-adapted stages, the procyclic promastigote. This transition is thought to be triggered by a combination of pH and temperature shift within the vector [Sunter, 2017]. The procyclic promastigotes are weakly motile but

Table 1.1

Stage	Morphology	Timing of appearance	Function	Host (sub)compartment
Procyclic promastigote	Cell body 6.5 - 11.5 μm; flagellum shorter than cell body	12 – 18 hr; peak ~48 hr	• Replication within bloodmeal	Abdominal midgut; lumen of PM
Nectomonad promastigote	Needle-like cell body >12 μm	36 hr; peak ~60 hr	• Escape from PM • Anterior migration • Midgut attachment	Abdominal midgut; lumen of PM, attached to intestinal epithelium throughout
Leptomonad promastigote	Cell body 6.5 - 11.5 μm; flagellum longer than cell body	3-4 days; peak ~day 5	• Main replicative stage • Secretion of PSG	Thoracic midgut
Haptomonad promastigote	Broad, leaf-shaped (size and flagellar length varies)	4 days	• Stomodeal valve attachment	Thoracic midgut; anchored to stomodeal valve
Metacyclic promastigote	Cell body <8 μm long and 1μm wide; long flagellum	4 days; peak ~day 10-14	• Transmission to vertebrate host	Thoracic midgut; enriched in PSG plug
Amastigote	Ovoid cell body; no protruding flagellum	2-3 days (post-invasion)	• Replicate within vertebrate host	Parasitophorous vacuole (phagolysosome-like)

Table 1.1: Characteristics, roles, and host environments inhabited by the developmental stages of the *Leishmania* lifecycle. The timing of lifecycle progression varies somewhat by species; timing indicated here is based on observations from *L. mexicana* [Rogers, 2002; Gossage, 2003].

highly proliferative, characterized by a cell body length between 6.5-11.5 μm and a short flagellum [Rogers, 2002]. Supported by nutrients in the bloodmeal, these forms replicate within the confines of the PM, leading to a rapid expansion in cell density during the first 18-24 hours of infection [Rogers, 2002; Gossage, 2003].

The PM plays dual, opposing roles over the course of *Leishmania* development [Dostálová, 2012]. In one respect, it is thought to benefit *Leishmania* early in infection by buffering against the proteolytic activity of the host; preventing PM formation *in vivo* leads to parasite lysis and loss of the initial midgut infection [Pimenta, 1997]. Conversely, at later stages, it presents a physical barrier that *Leishmania* must traverse to avoid being ejected from the vector during defecation [Pimenta, 1997; Secundino, 2005; Walters, 1989]. Escape from the PM is accomplished by nectomonad promastigotes. These non-replicating forms appear roughly 36 hours post-infection and are readily distinguished from their procyclic predecessors by their long, needle-like shape and higher motility [Walters, 1989; Rogers, 2002]. As with the procyclic promastigote, nectomonad development is not well understood; however, similar morphologies have been observed *in vitro* in response to certain types of nutrient stress, such as purine starvation [Martin, 2014]. Additionally, the nectomonad transcriptome shows an upregulation in stress- and starvation-response genes [Inbar, 2017], indicating that nutrient depletion may be an important trigger for differentiation. Aided by high flagellar activity, nectomonads rupture the peritrophic sac and migrate toward the anterior portion of the abdominal midgut, where they anchor to the intestinal wall by intercalating their flagella between the microvilli [Killick-Kendrick, 1974; Walters; 1989]. Attachment is mediated by surface lipophosphoglycans (LPG), which form species-specific interactions

with lectins expressed on the host epithelium [Sacks, 2000; Svárovská, 2010; Kamhawi, 2004]. In addition to PM escape, this step is critical to preventing excretion with the digested bloodmeal and is thought constitute a major determinant of parasite-vector specificity [Kamhawi, 2004; Pimenta, 1994].

Around 3-4 days post-feeding, nectomonads detach from the intestinal wall and undergo a third round of differentiation, giving rise to the leptomonad promastigote [Rogers, 2002]. This change occurs roughly coincident with defecation of the bloodmeal; however, the specific environmental signals for leptomonad development are not known [Walters, 1989]. Leptomonads are morphologically similar to the procyclic (first) promastigote stage but are distinguished by flagellar length, possessing a flagellum that is longer than the cell body [Rogers, 2002]. Importantly, leptomonads represent the main replicative stage within the vector, responsible for the massive population expansion required to generate a mature, transmissible infection [Serafim, 2017]. Over the next several days, parasite numbers steadily rise, peaking around day 6 [Rogers, 2002]. At the same time, anterior migration continues, with leptomonad parasites eventually reaching the stomodeal valve that separates the thoracic midgut from the foregut (refer to Figure 1). A small subset of these differentiate to broad, leaf-shaped haptomonad promastigotes, which attach to the valve via hemidesmosome-like structures [Walters, 1989]. While the exact role of these forms has not been definitively proven, they are thought to cause damage to the stomodeal valve, facilitating reflux of the midgut contents at the next bloodmeal [Schlein, 1991].

The other major function of the leptomonad promastigote is the production of promastigote secretory gel (PSG). Composed primarily of filamentous

proteophosphoglycans (fPPG), PSG plays a critical role in *Leishmania* transmission, affecting sand fly feeding behavior to favor inoculation of the vertebrate host [Rogers, 2004; Rogers, 2007; Rogers, 2009]. Briefly, throughout leptomonad development, PSG accumulates behind the stomodeal valve to fill the thoracic midgut. As a result, the so-called 'blocked fly' is unable to feed properly and must regurgitate the PSG plug into the bite wound in order to obtain a blood meal [Ilg, 1996; Rogers, 2002]. It is within this matrix that leptomonads complete their transition to infective metacyclic promastigotes in a process known as metacyclogenesis. Metacyclics are non-dividing, highly motile forms, preadapted for survival within the vertebrate host. This stage is identified by its small size (<8 μm in length and 1 μm wide) and a flagellum that is longer than the cell body [Rogers, 2002]. Although they are detectable in low numbers as early as 4 days post-feeding, metacyclic parasites do not become the dominant morphology until around day 10, at which point they account for ~50% of the total parasite burden in the midgut [Rogers, 2002]. As demonstrated by careful dissection of infected sand flies, the majority (~75%) of metacyclic promastigotes are contained within the plug, strategically positioned for transmission [Rogers, 2002].

The molecular basis for metacyclogenesis is still a topic of open investigation. In early studies by Sacks and colleagues, parasites isolated from stationary-stage cultures were shown to be significantly more infective than those undergoing exponential growth [Sacks, 1984], suggesting that metacyclogenesis may be part of a generalized response by the parasite to adverse environmental conditions. In agreement with this, a functioning autophagic pathway is required for metacyclogenesis in *Leishmania* [Bestiero, 2016]. A variety of stressors have since been shown to promote metacyclogenesis *in vitro*,

including acidic pH, low oxygen, glucose restriction, and reduced tetrahydrobiopterin [Bates, 1993; Mendez, 1999; Saini, 2016; Cunningham, 2001]. However, none of these effects were recapitulated in the context of an actual sand fly infection. Of particular interest, purine starvation is also implicated in differentiation. Feeding infected sand flies on supplemental adenosine was sufficient to inhibit metacyclogenesis *in vivo* [Serafim, 2012]. *Leishmania* are purine auxotrophs and exclusively acquire these nutrients from the host milieu. It is therefore possible that purine restriction within the midgut contributes to differentiation; however, the actual signals encountered in the sand fly are not known.

Classically, metacyclic promastigotes were considered to be terminally differentiated and incapable of further transformation until passage to the vertebrate host. This was based on observations from experimental sand fly infections, typically established using a large number of parasites administered in a single bloodmeal. [Rogers, 2002; Gossage, 2003; Dostálová, 2012; Bates, 2008]. However, recent evidence from Serafim and colleagues suggests that the metacyclic stage may be more plastic than previously appreciated. By sequentially blood-feeding infected sand flies, the authors demonstrated that metacyclic parasites could be induced to revert to a leptomonad-like form (termed the retroleptomonad promastigote) and reenter the cell cycle. These forms were found to undergo metacyclogenesis much like the leptomonad promastigote. This had profound impacts on disease transmissibility, quadrupling the number of infective metacyclic promastigotes in twice-fed sand flies over those only receiving a single bloodmeal [Serafim, 2018]. Importantly, female sand flies typically require one bloodmeal per 6-day gonotrophic cycle, several of which occur during the course of *Leishmania* infection [Killick-Kendrick, 2002]. Thus, this sequential-feeding model is

likely representative of *Leishmania* development in the field, wherein iterative rounds of vector feeding, replication, and metacyclogenesis serve to amplify parasite numbers within the gut.

Development in the vertebrate host

In contrast to the complex developmental program taking place inside the vector, the intracellular phase of the *Leishmania* lifecycle is relatively simple. Upon host cell invasion, metacyclic promastigotes undergo a single differentiation event to become tissue-resident amastigotes. This stage is responsible for maintaining the infection and causes disease pathology. Nonetheless, similar to the promastigote forms in the sand fly midgut, amastigogenesis represents an adaptive response by the parasite to unique environmental challenges presented by the vertebrate host. Here again, stress plays a critical role in lifecycle progression.

Transmission occurs when an infected sand fly takes a blood meal. Sand flies are pool-feeders and equipped with a barbed proboscis which they repeatedly insert into the skin to lacerate capillaries in the dermis. Neutrophils, followed later by macrophages, rapidly infiltrate the bite wound as part of the host response to tissue damage [Phillipson, 2019]. At the same time, PSG produced by leptomonads in the thoracic midgut promotes reflux by the infected sand fly (i.e. the 'blocked fly' hypothesis), introducing metacyclic parasites into the skin [Rogers, 2004]. A mixture of chemoattractant and immunomodulatory factors, including PSG and sand fly salivary proteins, are also present in the inoculum. Together, these shape the immunological profile of the infection

site to promote *Leishmania* survival and invasion [Van Zandbergen, 2002; Abdeladhim, 2014; Rogers, 2009].

Although macrophages are the definitive host cell for *Leishmania*, neutrophils are the first phagocytes they encounter upon transmission, attracted by stress signals released from damaged and necrotic tissue in the bite wound. Metacyclic promastigotes invade these cells early on and are able to resist intracellular killing [Laufs, 2002]. Interestingly, internalized parasites do not appear to differentiate or replicate inside neutrophils [van Zanderbergen, 2004]; however, the ability of *Leishmania* to successfully colonize a vertebrate host is reduced by neutrophil depletion, suggesting that this cell type may serve as an important reservoir for *Leishmania* during the first hours or days of infection [Peters, 2008]. Characteristics of the neutrophil compartment inhabited by metacyclic parasites are not known [Moradin, 2012].

Following the immediate neutrophil response, macrophages are recruited to clear cellular debris from the infection site. *Leishmania* gain access to the macrophage interior in one of two ways. In the so-called 'Trojan horse' model, macrophages engulf apoptotic, infected neutrophils containing the metacyclic parasites. Parasites that enter via this route do so without activating the macrophage antimicrobial response, instead triggering release of the potent anti-inflammatory mediator TGF-β [van Zanderbergen, 2004]. Thus, it is thought that early-responding neutrophils may provide a conduit for immunologically 'silent' macrophage invasion. Alternatively, extracellular parasites can directly engage receptors on the macrophage membrane to enter via phagocytosis [Moradin, 2012]. In this case, defensive molecules expressed by *Leishmania*, including

LPG and surface metalloprotease gp63, disarm the microbicidal machinery of the activated host cell [reviewed in Walker, 2013].

Regardless of their method of entry, metacyclic promastigotes exist only transiently inside the macrophage, giving rise to amastigote forms within two to five days of invasion [Antoine, 1998]. Amastigotes are identified by their rounded cell body and a short flagellum that barely protrudes from the flagellar pocket [Sunter, 2017]. Phagosomes containing the internalized parasites fuse with other endocytic vacuoles and lysosomes in the cytoplasm, resulting in the production of a PV. This compartment shares many characteristics with the mature macrophage phagolysosome, including a pH around 4.2 to 5.7 and an abundance of lysosomal acid hydrolases; however, both metacyclic- and amastigote-stage *Leishmania* subvert reactive oxygen generation within the PV by preventing assembly of host NADPH oxidase on the vacuolar membrane [Antione, 1990 and 1998; Sturgill-Koszycki, 1994; Russell, 1992; Lodge, 2006a and b]. Consequently, acidic pH and elevated temperatures (like those encountered in the human body) are classically regarded as the primary cues for amastigote differentiation. In all *Leishmania* species tested to date, these stimuli are together sufficient to induce amastigogenesis in the absence of a host cell (i.e. in axenic culture) [Zilberstein, 1994]. The resultant amastigote-like forms demonstrate the same rounded morphology and increased infectivity as their lesion-derived counterparts [Somanna, 2002; Debrabant, 2004]. Encouragingly, on a transcriptomic level, axenic and 'true' amastigotes are more similar to one another than either is to the cultured promastigote stage [Fiebig, 2015]. And yet, these two forms are still transcriptionally divergent, with ~13% of cellular mRNAs

differing in abundance, suggesting that additional signals may contribute to amastigogenesis in the context of a natural infection [Holzer, 2006; Fiebig, 2015].

One possible candidate is iron uptake. *Leishmania* salvage iron from the PV lumen via the ferrous iron transporter LIT1 (*Leishmania* iron transporter 1) [Huynh, 2006]. At the same time, macrophages express the iron efflux pump Nramp1 on the PV membrane, reducing the lumenal concentration of this nutrient as part of the host response to intracellular pathogens [Gruenheid, 1997]. As shown by Mittra and colleagues, iron depletion in the culture medium results in LIT1 upregulation and increased iron uptake by *L. amazonensis*, triggering axenic amastigogenesis [Mittra, 2013]. Importantly, this effect was observed at neutral pH and low (26°C) temperature, leading the authors to conclude that iron sensing is a potent stimulus of amastigote differentiation, capable of overriding other environmental signals [Mittra, 2013]. Nonetheless, precisely how representative these iron-induced amastigotes are of those formed during vertebrate infection remains to be definitively proven, as their transcriptomic and proteomic profiles were not characterized. Thus, the full complement of signals for amastigote differentiation demands further study.

In summary, the *Leishmania* lifecycle involves a complex sequence of differentiation events, triggered by fluctuations in the host environment. Each developmental stage is specifically adapted to the host compartment it inhabits and each plays an integral role in lifecycle progression. As a consequence, mechanisms of stress tolerance and adaptation are of the utmost importance to *Leishmania*. Several stressors, including pH shift, temperature change, and nutrient restriction, have been shown to induce developmental transformation *in vitro*. However, for most stages, the conditions

driving differentiation are poorly characterized. Less still is known about the downstream mechanisms that translate these external triggers into transcriptomic, proteomic, and/or metabolomic changes within the parasite. Some of these pathways are expected to overlap and intersect, representing the general *Leishmania* stress response. Others are likely stimulus-specific, tailored to facilitate adaptation to a narrow range of conditions. In any case, identifying the molecular players involved will provide key insights into lifecycle progression and survival of this important human pathogen.

Challenges in conquering the *Leishmania* genome: A historical perspective

As kinetoplastid parasites branched from the eukaryotic lineage early in evolutionary history, *Leishmania* and related *Trypanosoma* species possess a number of unique biological features [Akhoundi, 2016]. Preeminent among these is their method of gene expression (Figure 1.2). In the nuclei of kinetoplastid parasites, protein-coding genes are arranged end-to-end in long, polycistronic arrays. Transcription by RNA polymerase II (Pol-II) initiates bidirectionally from DNA stretches separating oppositely oriented, divergent arrays and terminates when two transcription units converge. The multi-gene primary message is then processed into individual transcripts via 5' *trans*-splicing of a conserved, pre-capped leader sequence (the spliced leader, SL) and 3' polyadenylation. Thus, changes in gene expression underlying both progression of the *Leishmania* lifecycle and acute stress tolerance are mediated exclusively though post-transcriptional mechanisms.

As a consequence of their noncanonical genomes, kinetoplastid parasites have historically proven challenging to study. Promoterless constitutive transcription precludes

Figure 1.2

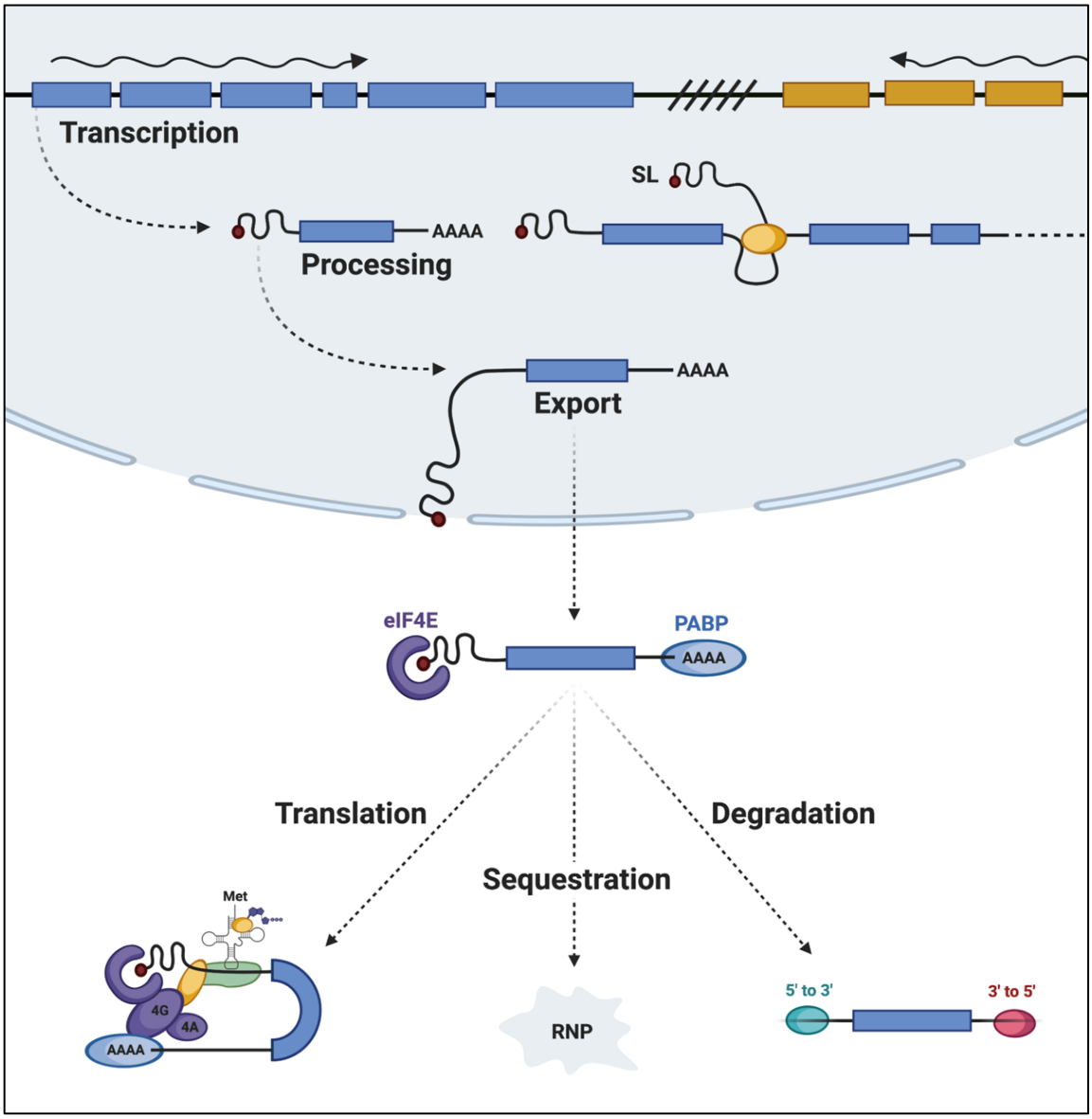


Figure 1.2: General overview of kinetoplastid gene expression. From top to bottom: In kinetoplastid parasites, protein-coding genes (colored boxes) are arranged end-to-end as polycistronic transcription units. Pol-II transcription (wavy arrow) initiates constitutively and in the absence of a canonical eukaryotic promoter. Transcription terminates when oppositely oriented units converge or at regions transcribed by other polymerases (hash marks). Polycistronic primary messages are co-transcriptionally processed into single-gene mRNAs via 5' *trans*-splicing with a pre-capped leader sequence (the spliced leader, SL) and 3' polyadenylation. Transcripts that are exported from the nucleus are stabilized by associations with cap-binding protein eIF4E and PABP. As in other eukaryotes, cytoplasmic transcripts may either be translated (Figure 1.4), degraded (Figure 1.5) or, in rare cases, sequestered in RNP granules.

the use of many classical genetic approaches. In *Leishmania* and *T. cruzi*, the RNA interference (RNAi) pathway is not conserved, eliminating this tool as a means of functional genomic analysis [Ullu, 2004]. Additionally, due to their unusual dependence on post-transcriptional regulation, changes in mRNA and protein abundance rarely track together in kinetoplastid parasites. Only with completion of the tritryp genomes in the early 2000s did the full complexity of kinetoplastid gene expression and regulation begin to yield to human understanding. Here, I outline a brief history of *Leishmania* genetics, highlighting several of the major advances of the last few decades.

Early genomic characterization

Genetic studies in *Leishmania* and related kinetoplastid parasites were long hampered by the lack of a manipulatable genetic system. Well into the 1980s, even the most basic structural information about the genome was unknown. This is largely attributable to the fact that the chromosomes of kinetoplastid flagellates do not condense significantly at any point during the life cycle, rendering them refractory to conventional cytogenetic techniques. Only with the advent of pulsed field gel electrophoresis (PFGE) did simple genetic characterization become somewhat feasible [Lighthall, 1992]. Early studies relied entirely on direct visualization of individual chromosomes, which produce a banding pattern when separated in an electric field. The resultant “molecular karyotypes” enabled investigators to discriminate between strains and answer very basic questions about genome size and ploidy. However, accurately interpreting these patterns was challenging due to comigration of multiple chromosomes within a limited size range [Lighthall, 1992]. As a result, reports of both ploidy and chromosome number varied

substantially - among early studies in *Leishmania*, estimates of chromosome number ranged from 23 to 96 [Spithill, 1987; Samaras, 1987; Pages, 1986; Scholler, 1986; Galindo, 1989].

The Leishmania Genome Network

In the early 1990s, the global parasitological research community coalesced around the common goal of producing a high-resolution genome map for *Leishmania*. The *Leishmania major* Friedlin strain (MHOM/IL/81/Friedlin; referred to as LmjF) was selected as the reference organism and, with the support of the World Health Organization (WHO), the *Leishmania* Genome Network (LGN) was established to coordinate the effort [Uliana, 2006]. An accurate description of the *Leishmania* genome was eventually reached using a combination of PFGE and systematic hybridization analysis. In pioneering work conducted by Wincker and colleagues, 244 individual loci probes were hybridized against the *L. infantum* (Old World) molecular karyotype to reveal a total of 36 physical linkage groups. These corresponded to the complete set of *L. infantum* chromosomes which range from 0.3 – 2.5 megabases (Mb) in size [Wincker, 1996]. Characterization of the New World *L. (Leishmania)* and *L. (Viannia)* subgenera followed shortly after, revealing a total 34 and 35 chromosomes, respectively [Britto, 1998]. Importantly, these reports uncovered a surprising degree of conservation in overall chromosome structure and gene order (i.e. synteny) across Old and New world *Leishmania* species alike, validating the use of a single reference genome as a working model for the entire genus. A similar approach was extended to the LGN reference strain

to produce a first-generation physical map of the entire *L. major* Friedlin genome [Ivens, 1998].

The next major benchmark of the LGN was achieved with sequencing of the first and smallest chromosome (Chr1), followed by Chr3 and Chr4 shortly thereafter [Myler, 1999; Myler, 2001]. Despite comprising only a small fraction of the total *Leishmania* genome, analysis of the 285-kb Chr1 sequence exposed several remarkable genetic features unique to kinetoplastid parasites. Consistent with other kinetoplastids, the study found no evidence for introns within any of the 79 putative open reading frames encoded on Chr1. Additionally, all the intergenic regions contained one or more polypyrimidine-rich tracts, known to constitute at least part of the signal for 5' processing of all mRNAs. By far, the most interesting aspect of Chr1 was its organization. The sequence was found to be composed of two distinct units, with 29 genes tandemly arranged on one DNA stand and, on the other, the remaining 50 organized head-to-tail in the opposite direction. While other, smaller studies in related kinetoplastid parasites had hinted at a model of polycistronic arrangement, Myler and colleagues were the first to capture two opposing "head-to-head" clusters of protein-coding genes. In addition, these studies demonstrated that, unlike the regulatory operons of lower prokaryotes, in *Leishmania,* functionally related genes do not cluster together [Myler, 1999; Myler, 2001].

Completion of the genome

Through the combined efforts of over 100 LGN researchers distributed across 26 international institutions, sequencing of the entire *L. major* genome was completed in 2005 [Ivens, 2005]. Based on the results of this study, the haploid *L. major* genome was

found to be 32.8-megabases in total (~60% GC) and encode a predicted 911 RNA genes, 39 pseudogenes, and 8272 protein-coding genes. Together, these are arranged into 133 transcription units which span up to 1259 kb in length and contain tens to hundreds of functionally unrelated genes. Around this same time, the complete genomic sequences from *T. brucei* and *T. cruzi* also became available [Berriman, 2005; Luchtan, 2004]. Remarkably, despite having diverged 200 – 500 million years ago, the genomes of *Leishmania* and related kinetoplastid parasites were found to be highly syntenic [Ivens, 2005; El-Sayed, 2005]. Between the tritryp parasites, ~75% of putative protein-coding genes were found to be conserved. Among these, nearly all (94%) remain in the same genetic context.

Completion of the *Leishmania* genome project marked a significant turning point in the field. By establishing a framework for genome content and organization, this landmark accomplishment significantly improved the genetic tractability of kinetoplastid parasites and set the stage for the large number of mechanistic molecular and genetic studies that followed.

The key processes and players of kinetoplastid gene expression

Chromatin structure

The first layer of genome regulation is established at the level of chromatin structure. Kinetoplastids possess a standard nucleosome composed of histones H2A, H2B, H3, and H4, as well as one variant of each. These include variant H2A.Z, which is evolutionarily conserved from yeast to mammals, as well as three others (H2B.V, H3.V, and H4.V) that are unique to kinetoplastid flagellates [Ivens, 2005; Siegel, 2009].

Leishmania also encode linker histone H1 is also encoded, though the *Leishmania* ortholog lacks an N-terminal globular domain that is present in most other eukaryotes [Ivens, 2005; Masina, 2007]. While the dynamic nature and regulation of chromatin structure are still incompletely understood for kinetoplastid parasites, evidence supports a role for H1 in chromosomal compaction that is tied to infectivity in *L. major* [Masina, 2007]. The *Leishmania* genome also encodes a number of enzymes with predicted roles in histone modification and epigenetic regulation, including acetyltransferases, methyltransferases, and histone deacetylases [Ivens, 2005].

Transcription

Canonically, transcription in eukaryotes is accomplished by three RNA polymerases (Pol-I, -II, -III) with each responsible for a specific subset of cellular transcripts [reviewed in Cramer, 2008]. Pol-I transcribes most ribosomal RNAs (rRNA), Pol-II produces protein-coding mRNAs and small nuclear RNAs, and Pol-III transcribes transfer RNA, the 5S rRNA, and other small messages. These multimeric complexes contain 14, 12, and 17 subunits, respectively, and share a general structure comprised of ten core subunits with additional polymerase-specific components associating at the periphery. *Leishmania* possess all three polymerases, having orthologs to most of the core subunits, though they lack many of the polymerase-specific proteins [Ivens, 2005]. Interestingly, though conserved in kinetoplastid genomes, the large Pol-II subunit lacks the C-terminal heptad repeat found in other eukaryotes [Ivens, 2005]. In higher organisms, this region is differentially phosphorylated throughout the transcription cycle, licensing regulating initiation, elongation, and termination [Cramer, 2008]. The absence

of these regions is consistent with transcription not being a primary means of gene regulation for kinetoplastids.

Compared to other eukaryotes, transcription of protein-coding genes is highly unusual in *Leishmania* and other kinetoplastids. Pol-II initiates bidirectionally from stretches of DNA separating divergent gene clusters, termed 'strand-switch' regions (Figure 1.3). While these regions are generally characterized by an unusually high AT base composition and the relative absence of predicted secondary structure, they lack canonical promoter elements, such as the TATA box found in archaea and other eukaryotes [Martinez-Calvillo, 2003; Tosato, 2001; Ivens, 2005]. Likewise, kinetoplastid parasites possess relatively few sequence-specific Pol-II transcription factors [Ivens, 2005]. Instead, nucleosome compaction and modification are thought to serve as the primary determinants of transcription initiation in these organisms. A number of epigenetic markers are associated with transcription start sites (TSSs), including H3K9/K14 and H4K10 acetylation [Thomas, 2009; Siegel, 2009], H3K4 trimethylation [Wright, 2010], and the transcriptional activator BDF3 (bromodomain factor 3) [Siegel, 2009]. Histone variants H2A.Z and H2B.V are also enriched at TSSs [Siegel, 2009]. It is known that nucleosomes formed with H2A.Z are less stable than those containing the core histone H2A, suggesting that this variant may facilitate TSS firing by maintaining a particularly labile chromatin conformation over Pol-II initiation sites [Jin, 2007].

Transcription proceeds polycistronically and terminates where gene arrays converge or at loci transcribed by other RNA polymerases. In all kinetoplastid flagellates characterized to date, termination sites are enriched in a hypermodified thymidine residue, β-D-glucosyl-hydroxymethyluracil, commonly referred to as base J [Van

Figure 1.3

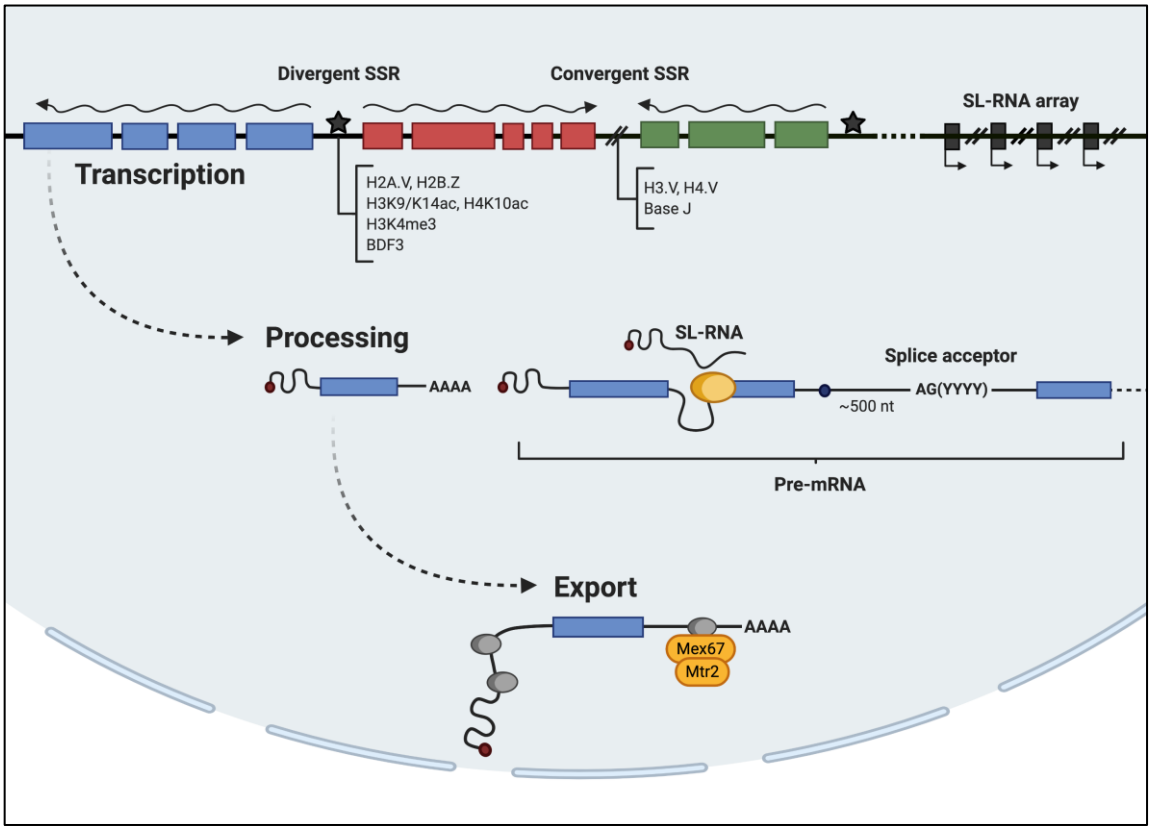


Figure 1.3: Events in the nucleus. Transcription: With the exception of the SL precursor (SL-RNA), most Pol-II transcription start sites (TSSs, stars) are located in divergent strand switch regions (SSRs) and lack recognizable eukaryotic promoter elements. Instead, the TSS is marked by a unique epigenetic signature, including enrichment for histone variants H2A.V and H2B.V, H3K9/K14 and H4K10 acetylation, H3K4 trimethylation, and bromodomain factor 3 (BDF3). Polycistronic transcription (wavy arrow) terminates at convergent SSRs where Pol-II readthrough is prevented by the presence of histone variants H3.V and H4.V and β-D-glucosyl-hydroxymethyluracil (Base J). SL-RNA is the only Pol-II-transcribed message that is expressed under the control of defined promoter and terminator sequences (90º arrows and double hash marks, respectively). Processing: mRNA 5' processing is directed by a splice acceptor site that closely resembles the consensus sequence for *cis*-splicing in other eukaryotes. 5' and 3' processing reactions are mechanistically coupled such that *trans*-splicing by the kinetoplastid spliceosome (yellow oval) dictates the position and efficiency of polyadenylation for the upstream gene. In *L. major*, the polyadenylation site (black circle) is generally found 500 nt from the downstream splice acceptor. Export: Kinetoplastids rely on the universally conserved Mex67/Mtr2 shuttling complex to export mRNAs to the cytoplasm. RBPs (grey ovals) responsible for recruiting Mex67/Mtr2 to targets for export are not known.

Luenen, 2012]. Base J prevents readthrough past the end of individual transcription units, though its relative importance and genomic distribution varies substantially between the tritryp parasites. For instance, whereas base J depletion results in genome wide defects in transcription termination and cell death in *Leishmania*, it is not required for viability in *T. brucei* [Reynolds, 2014; Van Luenen, 2012; Schulz, 2016]. Instead, in both *T. brucei* and *T. cruzi,* base J appears to contribute to developmental gene expression, being either significantly downregulated or entirely absent during certain lifecycle stages [van Leeuwen, 1997; Ekanayake, 2007]. Loss of base J also results in increased TSS firing in *T. cruzi,* an observation that has not been recapitulated in either *T. brucei* or *Leishmania*, suggesting that this epigenetic marker plays functionally divergent roles in different kinetoplastids [Ekanayake, 2011]. Pol-II termination sites are also enriched in the histone variants H3.V and H4.V [Siegel, 2009]. As H3.V-null *L. major* parasites are not compromised in transcriptional termination, this variant is not believed to be an important regulator of readthrough in *Leishmania* [Anderson, 2013]. However, similar to base J, loss of H3.V leads to transcriptional readthrough at specific genomic loci in *T. brucei*. This phenotype is intensified by subsequent base J depletion, suggesting that the two marks act synergistically in this organism [Reynolds, 2014]. In any case, the precise mechanism of transcriptional termination in kinetoplastids is not known.

Processing

The polycistronic products of Pol-II transcription are resolved into single-gene mRNAs via two co-transcriptional process. In the first, a conserved mini-exon sequence (i.e. the spliced-leader, SL) is *trans*-spliced onto what will become the 5’ end of each

individual message. Though the length (39-nt in *L. donovani*) and sequence of the SL varies slightly between kinetoplastids, it is completely invariant within a given species. The SL is donated by a non-polyadenylated precursor known as the SL-RNA. SL-RNA is pre-capped by 7-methylguanosine and modified at the first four bases, thus bestowing all mRNAs with a 'cap4' structure [Perry, 1987; Bangs, 1992]. Interestingly, the SL-RNA is the only Pol II-transcribed message under the control of discrete promoter and terminator elements [Gilinger, 2001]. The *L. donovani* genome encodes ~68 copies of SL-RNA, arranged as a cluster of repeated units wherein each is assigned its own promoter [Lypaczewski, 2018; Saito, 1994].

Trans-splicing is mediated by a spliceosome similar to that which facilitates the removal of introns in other eukaryotes [Liang, 2003]. Also parallel to *cis*-splicing, SL addition involves two transesterification reactions, though the branched intermediate forms a Y instead of a lariate structure [Murphy, 1986]. The 5' end processing signal, known as the 'splice-acceptor,' is loosely defined, consisting of polypyrimidine tract immediately followed by an AG dinucleotide [Matthews, 1994]. Efficiency of *trans*-splicing varies based on the length, composition, and position of the polypyrimidine tract [Siegel, 2005]. Most genes typically favor a dominant SA site, though processing is remarkably promiscuous and the vast majority of ORFs have two or more [Kolev, 2010; Camacho, 2019].

The second required processing reaction in the nucleus is 3' polyadenylation. Unusually, kinetoplastids lack a clear consensus motif for polyadenylation, such as the AAUAAA signal sequence in higher eukaryotes. Instead, polyadenylation and *trans*-splicing of adjacent genes are mechanistically coupled such that selection of a 5' splice-

acceptor site influences the position of polyadenylation for the upstream gene [LeBowitz, 1993; Matthews, 1994]. In *L. major,* polyadenylation was shown to occur ~500 nucleotides upstream of splicing [LeBowitz, 1993]

mRNA Export

The mechanisms by which kinetoplastid parasites export nuclear mRNAs to the cytoplasm are incompletely understood. In higher eukaryotes, mRNAs are decorated with accessory proteins that facilitate translocation through the nuclear pore complex throughout transcription and processing. Export is universally dependent on a heterodimeric shuttling factor known as Mex67/Mtr2 [Kohler, 2007]. Although Mex67/Mtr2 can itself bind weakly to nucleic acid, it primarily associates via adapter RNA-binding proteins (RBPs) [Köhler, 2007]. Additionally, in yeast and metazoans, mRNA synthesis is coupled to export though the multiprotein transcription/export (TREX) complex. All three of the tritryp parasites express Mex67 and Mtr2 and RNAi against either *T. brucei* homologue results in nuclear accumulation of polyadenylated transcripts, suggesting that these organisms also to rely on the conserved Mex67/Mtr2 export system [Schwede, 2009; Dostálová, 2013]. A TREX subunit (Sub2) is also present in kinetoplastids and appears to associate with sites of active Pol-II transcription in *T. cruzi*, consistent with a role in the RNA transcription/export pathway [Serpeloni, 2011a, b]. Yet additional TREX components are seemingly absent from tritryp genomes. Additionally, kinetoplastids lack homologues of most accessory RBPs involved in recruiting Mex67/Mtr2 in eukaryotes [Serpeloni, 2011a]. Thus, how the mRNA shuttling complex associates with its target transcript remains an open question.

Translation

mRNAs are typically conscribed to one of a few fates in the cytosol (Figure 1.2). Highly unstable transcripts may be directly targeted for degradation without translation. In rare cases, some mRNAs are shielded from both translation and decay by targeting to cytoplasmic ribonucleoprotein granules (RNPs), where they are sequestered for later use. More commonly, binding by eukaryotic translation initiation factors (eIFs) triggers ribosome assembly and protein synthesis.

As with higher eukaryotes, translation initiation in kinetoplastid parasites is a complex event requiring a large number of eIFs [reviewed in Gerbauer, 2004]. Briefly, translation begins with the formation of a 43S pre-initiation complex comprised of the small 40s ribosomal subunit, a methionine-loaded initiator tRNA, and a handful of eIFs (Figure 1.4). Important among these are GFP-coupled eIF2, which is implicated in global translational regulation, and eIF3, which is responsible for mRNA recruitment. At the same time, mRNAs are marked for translation initiation by binding of a second multiprotein complex known as eIF4F. This assembly is composed of three additional initiation factors: eIF4E, which binds the mRNA 5' cap, the RNA helicase eIF4A, and the scaffold protein eIF4G, which is thought to recruit the 43S pre-initiation complex through association with eIF3 [Lamphear, 1995]. Kinetoplastid parasites possess multiple homologues of the various eIF4F components (see Chapter 1, Section 5).

Upon binding to the transcript, the small 40S ribosomal subunit scans along the untranslated region in a 5'-to-3' direction until encountering an AUG start codon. The large 60S ribosomal subunit then joins the complex at the translation start site to form a complete 80S ribosome and commence translation. Subsequent recruitment of additional

Figure 1.4

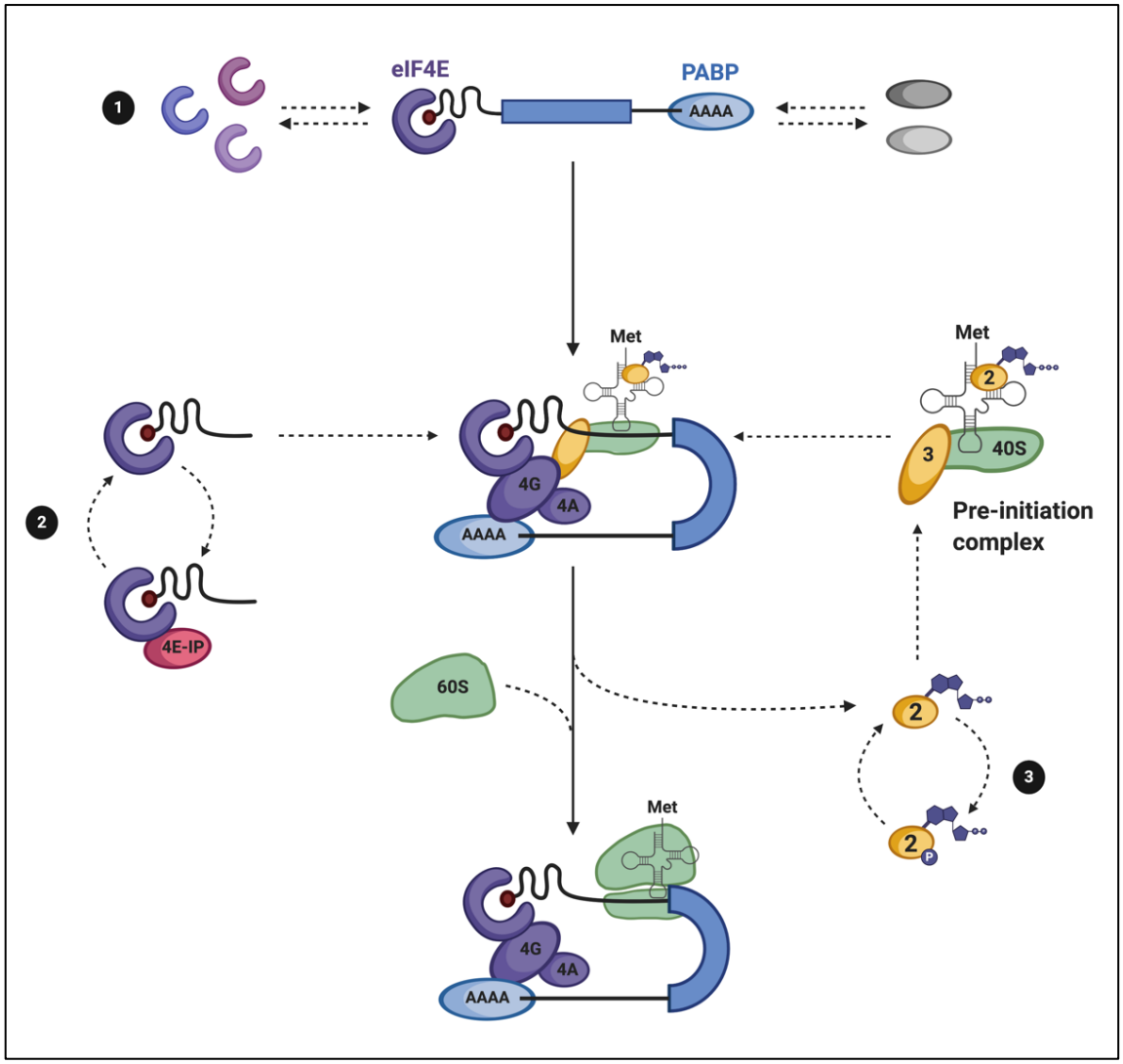


Figure 1.4: Translation initiation and control in kinetoplastids. Initiation: mRNAs are marked for translation by the eIF4F complex (subunits in dark purple). Cap-binding eIF4E tethers the complex to the transcript. RNA helicase eIF4A (4A) aids in translation by unwinding RNA secondary structure. eIF4G (4G) serves as a scaffold for complex assembly and circularizes mRNAs by interacting with PABPs at the 3' end. The pre-initiation complex, or PIC, is recruited by interaction of eIF3 (3) with eIF4G. The small (40S) ribosomal subunit scans along the mRNA 5' UTR until a start codon is reached, triggering GTP hydrolysis on eIF2 (2) and GDP-eIF2 release. The large (60S) ribosomal subunit then binds to form a functional 80S ribosome. Regulation: 1) Kinetoplastids express multiple homologs of the eIF4F subunits (for simplicity, only eIF4E is shown) and several PABPs. Homologs vary in their interactions, binding capabilities, and expression levels, and are thought to associate with distinct mRNA pools. 2) 4E-interacting proteins (4E-IPs) competitively bind to eIF4E, blocking eIF4G and formation of the eIF4F complex. 3) Phosphorylation of GDP-eIF2 blocks GTP exchange, preventing PIC reconstitution.

ribosomes results in formation of a polysome. In addition to bridging to the pre-initiation complex, eIF4G can also interact with PABPs at the mRNA 3' terminus, circularizing the polysome-associated transcript [Wells, 1998]. This association is thought to both protect the mRNA from degradation and facilitate translational reinitiation by allowing terminating ribosomes to reassemble on the same message.

mRNA decay

Owing to their unique genomic organization, kinetoplastid parasites are incapable of individual transcriptional regulation; yet the steady-state levels of mRNAs produced from adjacent genes can differ substantially. In some cases, high-level expression is achieved by amplification within the genome. However, since gene copy number cannot be differentially regulated, this approach is reserved for abundant housekeeping proteins like tubulin and histones [Manful, 2011]. The relative efficiency of pre-mRNA processing can also contribute to steady-state transcript level, though this is unlikely to be a major point of differential control since changes in *trans*-splicing of one gene would impact polyadenylation of the upstream ORF. Thus, transcript abundance is primarily determined by mechanisms operating at the level of mRNA turnover.

Several routes initiate mRNA decay. All mature mRNAs are protected from degradation by two common features: the 5' cap and the poly(A) tail. Loss of either one of these exposes the transcript to cytoplasmic exoribonucleases. For most transcripts, the process begins at the 3' end with deadenylation [Manful, 2011]. Three main deadenylases are known in eukaryotes: the CCR4/CAF1/NOT complex, the PAN2/PAN3 heterodimer, and the poly(A) ribonuclease PARN. All three are present in the genomes of tritryp

parasites [Clayton, 2012]. Once reduced to fewer than ~25 nt in length, the poly(A) tail can no longer support binding of PABP and its protective effect is lost [Lowell, 1992]. At this point, transcripts are susceptible to digestion by the exosome, a nine-subunit exoribonuclease complex that mediates 3' to 5' decay in both the cytoplasm and nucleus [Clayton, 2016]. More often, however, deadenylation triggers removal of the mRNA 5' cap. In higher eukaryotes, this is catalyzed by the DCP2 decapping enzyme and enhanced by a handful of accessory protein factors. Paradoxically, while several of the decapping accessory factors are found in tritryp genomes, kinetoplastids lack a DCP2 homolog [Clayton, 2012; Schwede, 2009]. Instead, 5' cap removal is accomplished by a unique ApaH-like phosphatase ALPH1, though how ALPH1 is specifically recruited to its targets is not known [Kramer, 2017].

Decapped and deadenylated transcripts are degraded primarily in the 5'-3' direction by a single exoribonuclease known as XRN1. Although kinetoplastid parasites possess four homologues of XRN1, XRNA appears to be responsible for the majority of mRNA turnover in the cytosol [Li, 2006]. In rare cases, decay is initiated without prior removal of the poly(A) tail, either via direct decapping or endonucleolytic cleavage. These deadenylation-independent pathways are thought to be reserved for a small subset of highly unstable transcripts in kinetoplastid parasites [Li, 2006; Schwede, 2009; Manful, 2011].

In eukaryotes, components of the 5'-3' mRNA decay pathway typically aggregate in two types of cytoplasmic granules: processing bodies (P-bodies) and stress granules. P-bodies are nonmembrane-bound RNP structures that function as specialized centers of mRNA decay, although they are not required for degradation to occur [reviewed in Luo,

2018]. P-bodies can vary in composition but typically contain RBPs that function in translational repression, deadenylation, decapping, and 5'-3' digestion. Notably, they do not contain subunits of the exosome [Luo, 2018]. Although P-bodies are constitutively present, their formation is induced by cellular stress [Kramer, 2013].

Stress granules are thought to function primarily in translational repression by sequestering and storing non-translating messages. Like P-bodies, stress granules are comprised of mRNAs and some components of the mRNA degradation machinery, including XRNA. Also like P-bodies, the stress granule protein profile can vary with respect to stress and cell-type, though the latter are distinguished by their larger size, the absence of deadenylation enzymes, and the presence of some translation initiation factors [Buchan, 2010]. Mammalian stress granules also contain ribosomal subunits, although this may not be the case in kinetoplastids [Kramer, 2013]. Importantly, mRNAs have been observed cycling between P-bodies, stress granules, and polysomes in a number of model eukaryotes, indicative of the dynamic nature of these structures. [Brengues, 2005; Buchan, 2010].

Figure 1.5

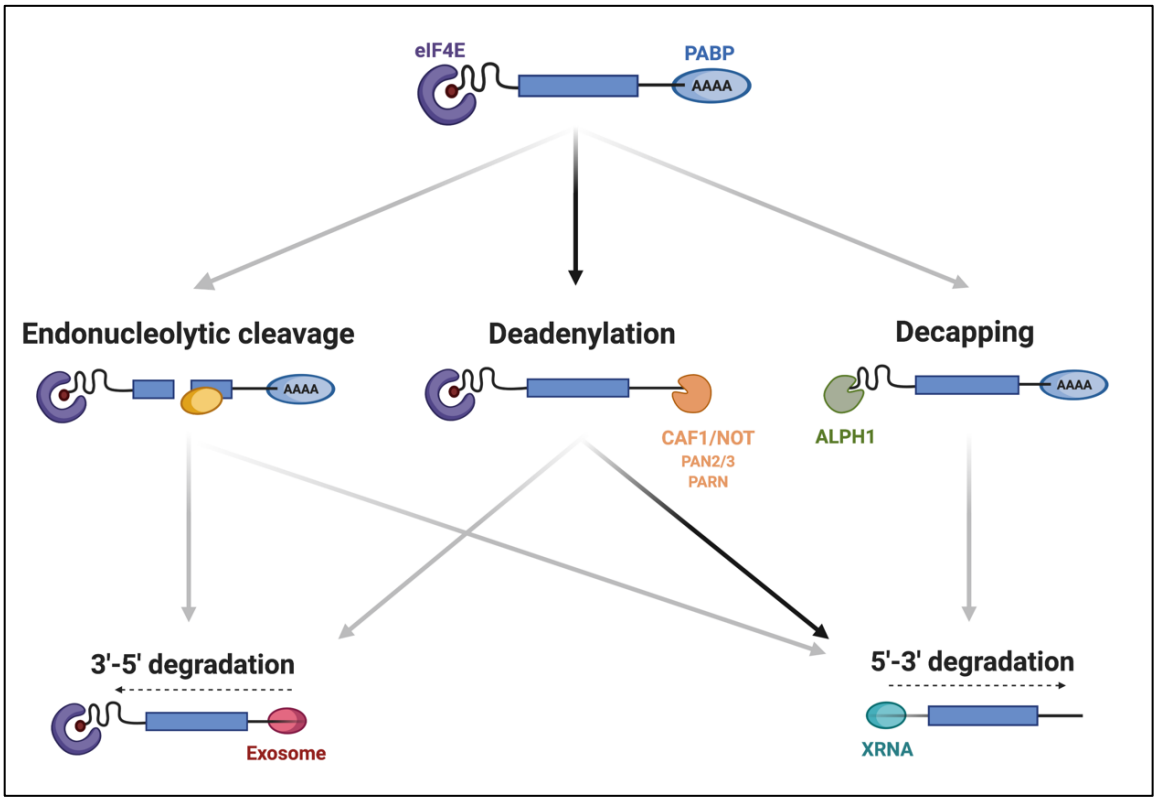


Figure 1.5: Pathways of mRNA decay. Black arrows trace the most common route to degradation. In the cytoplasm, messages are stabilized by association with cap-binding protein eIF4E and PABPs. For most transcripts, degradation is initiated by removal of the poly(A) tail. The CAF1/NOT complex is responsible for the majority of deadenylation, with PAN2/PAN3 and poly(A) ribonuclease (PARN) contributing to a lesser degree. Deadenylated transcripts are susceptible to degradation from the 3' end by the exosome but most are subsequently decapped by ALPH1 and degraded in the 5' to 3' direction by XRNA. In rare cases, decay initiates without prior deadenylation, either via deadenylation-independent decapping or internal cleavage by an endoribonuclease (yellow oval).

RNA-binding proteins: Master regulators in kinetoplastid parasites

To compensate for their lack of individual transcriptional control, *Leishmania* and related kinetoplastid parasites rely heavily on post-transcriptional processes to implement changes in gene expression. These mechanisms can be generally classified as having either global or specific effect as well as by the levels at which they operate (i.e. mRNA abundance and localization, translation, protein stability, etc). With the obvious exception of protein modification, RNA-binding proteins (RBPs) affect all stages of post-transcriptional control; thus, RBPs and the *cis*-acting mRNA sequences with which they interact are key determinants of adaptation in kinetoplastid flagellates.

RBPs can bind to any number of features found in mature transcripts, including common elements like the 5' cap and poly(A) tail as well as specific *cis*-acting sequences in coding and non-coding regions alike. However, most mRNA-encoded regulatory elements have been found in 3'-UTRs [Clayton, 2016]. RBPs influence all stages of mRNA metabolism and each mRNA is likely bound by multiple RBPs that compete and coordinate to determine mRNA fate. Additionally, individual RBPs are known to bind multiple mRNA targets, coordinating expression of multi-gene cohorts, or regulons [Keen, 2007; Noé, 2008; Trenaman, 2019]. And yet, despite their central importance in kinetoplastids, only a handful of RBPs and fewer *cis*-acting elements have been ascribed roles in the context of specific stress-response pathways.

I describe below the main control points of kinetoplastid gene expression, highlighting the potential roles of RBPs in each. As post-translational modification is not a primary focus of the studies described in this thesis, I have limited the scope of this discussion to mRNA-centric mechanisms. In the final portion of this section, I have

outlined the major families of RBPs in kinetoplastid parasites and provided several specific examples. It should be noted that much of what is known about gene regulation in these organisms is based solely on work conducted in *T. brucei.* This is due to their relative genetic tractability among the tritryp parasites: tools for RNAi-mediated gene knockdown and tetracycline-inducible expression are available in *T. brucei* but not *Leishmania* or *T. cruzi.* Although the tritryp parasites differ in various aspects of their respective biologies (i.e. host range, vector specificity, intra- vs extracellular lifestyle, etc.), roughly 6,200 genes are conserved between them. Moreover, their genomes are highly syntenic, with ~94% of conserved genes remaining in the same genetic context [El-Sayed, 2005]. Thus, it is thought that many core processes, including gene regulation, are well conserved.

Regulation of mRNA export

After transcription, mRNA export represents the first possible control point for gene expression among eukaryotes. For example, in yeast the transcripts encoding members of the Hsp70 gene family are selectively sent to the cytoplasm during heat shock [Saavedra, 1996]. However, it is unclear how translocation across the nuclear envelope is monitored and regulated in kinetoplastids, or if it is subject to differential control. Low levels of incompletely spliced (oligocistronic) mRNA were observed in the cytosol of trypanosomes and recently it was demonstrated that *T. brucei* can initiate mRNA export co-transcriptionally before the 3' end has been synthesized [Kramer, 2012; Goos, 2019]. This is in stark contrast to other eukaryotes, which restrict export to mature, processed messages. Still, the observation that unprocessed transcripts are ~eight-fold

enriched in *T. brucei* nuclear fractions suggests that some quality control mechanisms exist [Kramer, 2012]. Interestingly, when splicing is inhibited in trypanosomes, unprocessed mRNAs are seen to accumulate in granules on the cytoplasmic side of the nuclear envelope [Kramer, 2012; Goos, 2019]. These foci are remarkably similar to stress granules in terms of their protein composition and are enriched in a number of RBPs [Goos, 2019]. Based on these data, it was proposed that kinetoplastid parasites may regulate mRNA export at the point of emergence from rather than during recruitment to the nuclear pore complex; however, further study is required to discern whether this process is differentially controlled.

Regulation of mRNA decay

For most transcripts, removal of the poly(A) tail is the first and rate-limiting step in degradation [Manful, 2011]. Thus, deadenylation represents a key point in determining mRNA fates within the cell. All three of the main cytosolic deadenylases described in eukaryotes (i.e. the [CCR4]/CAF1/NOT complex, the PAN2/PAN3 heterodimer, and PARN) are present in kinetoplastid parasites. Each has unique characteristics and properties, suggesting that they play specialized and non-overlapping roles in mRNA turnover.

In yeast and metazoans, CCR4/CAF1/NOT accounts for the majority of cytosolic deadenylation. NOT1 is the only essential component of this complex, serving as a scaffold to support the assembly of the other subunits, and exoribonuclease activity is attributed to both CAF1 and CCR4 [Maillet, 2000; Collart, 1993; Bai, 1999]. Though kinetoplastids lack CCR4, the other components are conserved. In *T. brucei,* the majority

of transcripts are stabilized by CAF1 depletion, suggesting that CAF1/NOT is also central to constitutive mRNA turnover in trypanosomes [Schwede, 2008; Fadda, 2013].

PAN2/PAN3 is nonessential in yeast and absent altogether from the genomes of several parasitic protozoa, including *Plasmodium falciparum* and *Giardia lamblia* [Boeck, 1996; Schwede, 2009]. In most eukaryotes, PAN2/PAN3 is believed to function primarily in initiating deadenylation before passing transcripts to the CAF1/NOT complex [Sachs, 1987; Yamashita, 2005]. Nonetheless, some evidence suggests that PAN2/PAN3 may play a more significant role in kinetoplastid parasites. RNAi knockdown of PAN2 caused variable levels of growth inhibition in bloodstream-form *T. brucei* with some clones ceasing replication entirely, indicative of an essential function [Schwede, 2009]. Additionally, PAN2 was implicated in degrading several life stage-specific mRNAs in *T. brucei,* though the significance of this is unclear [Schwede, 2009].

Like the PAN2/PAN3 complex, PARN does not constitute a major route of deadenylation in higher eukaryotes. However, PARN regulates development in both *Xenopus* and higher plants by targeting specific mRNA subsets at various stages of growth [Kim, 2006; Reverdatto, 2004]. These observations have led many to speculate about potential regulatory roles for PARN in other organisms including kinetoplastids. The *T. brucei* genome encodes three different *PARN* genes, all of which are expressed in both vertebrate- and insect-resident stages; however, simultaneous RNAi knockdown of all three homologues had little to no effect on parasite growth*,* indicating that PARN activity is not required for viability [Utter, 2011]. Interestingly, PARN-1 overexpression selectively destabilized a small subset of *T. brucei* mRNAs, including a number of

developmentally-regulated transcripts. [Utter, 2011]. Still, the role of PARN in trypanosome differentiation remains to be rigorously tested.

Despite their central role in mRNA turnover, it is not clear how and when the various deadenylases are recruited to individual transcripts in kinetoplastids. It is tempting to speculate how RBPs fit into this process. For instance, an RBP could regulate mRNA decay by promoting or preventing association with any one of the deadenylases described above. Indeed, such recruitment factors were characterized in several model eukaryotes, including *Drosophila, Xenopus,* yeast, and humans [Chicoine, 2007; Kim, 2006; Goldstrohm, 2006; Hook, 2007; Lykke-Anderson, 2004]. While a number of RBPs have been localized to cytoplasmic P-bodies (i.e. specialized sites of mRNA decay) in kinetoplastids, so far there is no evidence that they bind directly with components of the deadenylation machinery.

In rare cases, mRNA degradation is initiated without prior removal of the poly(A) tail [Garneau, 2007]. This can be accomplished in one of two ways. In the first, mRNAs are rendered susceptible to 5' digestion by XRNA via deadenylation-independent decapping. There are few examples of direct decapping among model eukaryotes [Badis, 2004; Muhlrad, 2005] and none described in kinetoplastid parasites. However, identifying the proteins that interact with decapping enzyme ALPH1 *in vivo* could be informative in this regard. Alternatively, transcripts can be targeted for degradation by endonucleolytic cleavage. This route is functionally analogous to RNAi, exposing the interior of the message to attack by either XRNA or the exosome. The best-known example of this is found in *Leishmania* in a family of small degenerated retroelements known as the Short Interspersed Degenerated Retroposons (SIDERs).

Encoded in the 3'-UTRs of highly unstable transcripts, SIDER elements are targeted for site-specific cleavage within a 79-nucleotide signature sequence [Bringaud, 2007; Müller, 2010]. Importantly, cleavage is dependent on a small RBP, which facilitates endoribonuclease recruitment to the SIDER-containing transcript [Azizi, 2017]. Few other instances of deadenylation-independent decay have been described in kinetoplastids, though given their particular dependence on post-transcriptional regulation, it would not be surprising if these pathways play a more substantial role in kinetoplastid parasites than is currently appreciated.

Regulation of translation

Like higher eukaryotes, kinetoplastids regulate translation on a global scale through modification or disruption of the core translation machinery. A well-characterized example of this is phosphorylation of the translation initiation factor eIF2 [reviewed in Gebauer, 2004]. As described above, GTP-coupled eIF2 forms part of the 43S pre-initiation complex (PIC). At the start of translation, GTP is hydrolyzed to GDP. Thus, in order for subsequent rounds of translation to occur, active eIF2-GTP must be reconstituted by the GTP-exchange factor, eIF2B. Under stress conditions, this exchange is reversibly blocked by phosphorylation of eIF2 on its α subunit, preventing new PIC formation and halting translation initiation (Figure 1.4, 3). eIF2α phosphorylation has been observed in all three of the tritryp parasites and is induced in response to a number of exogenous stressors as well as during developmental transition [Tonelli, 2011; Lahav, 2010; Abhishek, 2017; reviewed in Kramer, 2013].

Kinetoplastid genomes encode an unusually large number of eIF4F components (Figure 1.4, 1), including 2 homologues of eIF4A, 5 of eIF4G, and 4-6 of cap-binding protein eIF4E [Freire, 2017]. Interestingly, these factors display variations in expression level, interaction partners, and, in the case of eIF4E, cap-binding capability [Freire, 2014; Dhalia, 2005]. While the relative importance of each eIF4F subunit homolog is incompletely understood, evidence suggests that they play a role in regulation by interacting with distinct mRNA subsets. In *L. infantum,* for example, eIF4E-1 is substantially upregulated in amastigotes, indicative of a stage-specific function [Yoffe, 2006]. Similarly, one of the six eIF4E homologs expressed in *T. brucei*, eIF4E-5, is essential for insect-stage parasites but is dispensable in bloodstream forms. RNAi-depletion of this protein results in a motility defect, perhaps indicative a pathway-specific function [Freire, 2014]. Adding a further level of complexity to this system, multiple PABPs are known in kinetoplastids [Kramer, 2012; da Costa Lima, 2010]. In both *T. brucei* and *Leishmania*, different PABP paralogs show distinct nuclear and/or cytosolic localizations in response to stress, suggesting that they have evolved divergent functions [Kramer, 2013]. Importantly, as both eIF4E and PABP recognize conserved molecular patterns common to all mRNAs (i.e. the 5' cap structure and poly(A) tail, respectively), the association of these different homologs with specific gene cohorts is most likely mediated through interactions with additional RBPs that bind transcripts in a sequence-specific manner.

In higher eukaryotes, an additional mechanism for regulating global protein synthesis involves displacement of eIF4E from other components of the translation initiation machinery (Figure 1.4, 2). eIF4E facilitates mRNA recruitment to the 43S PIC

through its interaction with scaffold protein eIF4G. This interaction is mediated by a 7-residue motif that is also present on a class of small regulatory proteins known as eIF4E-binding proteins (4E-BPs). Under stress, eIF4E is displaced from eIF4G by competitive binding with 4E-BP, preventing translation initiation [reviewed in Kamenska, 2014]. While no canonical 4E-BP homologs are present in the genomes of kinetoplastid parasites, proteins possessing similar function have been identified. Termed 4E-interacting proteins, or 4E-IPs, these factors contain the 7-reside consensus motif required for eIF4E binding [Terrao, 2018; Tupperwar, 2020]. Additionally, 4E-IPs demonstrate specificity for particular eIF4E homologs and stage-specific binding to their respective targets, providing yet another mechanism for controlling translation of distinct mRNA cohorts in kinetoplastid parasites [Terrao, 2018; Tupperwar, 2020; Zinoviev, 2011].

Finally, in addition to these pathways, RBPs can negatively regulate protein synthesis by sequestering mRNAs away from the translational machinery in the cytoplasm. This is exemplified by stress granules, which function to stabilize transcripts under cellular stress [Kramer, 2013]. A large number of proteins with predicted RNA-binding domains localize to stress granules in *T. brucei* [Fritz, 2015]. Conversely, mRNAs may be stored in granules under steady-state and released for translation when needed. While this phenomenon is well-documented in the distantly related apicomplexan parasite *Plasmodium* [Lindner, 2013], no such mechanism has yet been described in kinetoplastid parasites.

The families of RBPs in kinetoplastid parasites

RBPs are generally classified by their RNA-binding domains (RBDs). There are four major families of RBPs found in kinetoplastid parasites: RNA-recognition motif (RRM) proteins, CCCH zinc finger proteins, PUF proteins, and ALBA domain proteins. Recent evidence suggests that kinetoplastids also possess a large number of RBPs that lack a classical or recognizable RBD [Nandan, 2017]. For simplicity, those are omitted from this discussion.

The RRM domain is one of the most abundant and ubiquitously conserved in nature [reviewed in Afroz, 2015]. Approximately 90 amino acids long, RRM domains consist of two alpha helices packed against a four-stranded beta sheet, which typically serves as the main surface for RNA interaction. These domains are most often involved in sequence-specific binding to single-stranded RNA, though they can also bind to DNA. As in other eukaryotes, RRM proteins are highly represented in the genomes of kinetoplastid parasites, with ~75 and 80 identified in the *Trypanosoma* species and *L. major,* respectively [De Gaudenzi, 2005].

CCCH zinc finger proteins are generally composed of two or more finger-like protrusions which, similar to RRM, bind single-stranded RNA in a sequence-specific manner [Hudson, 2004]. The best studied among this group are the TIS11 family in mammals, which preferentially bind AU-rich elements in the 3'-UTRs of their target mRNAs [Carballo, 1998]. CCCH proteins are implicated in all stages of mRNA processing, stability, and decay. *In silico* screening predicted that there are 54 non-redundant proteins containing one or more CCCH-domains encoded in the *L. major* genome [Kramer, 2010].

Figure 1.6:

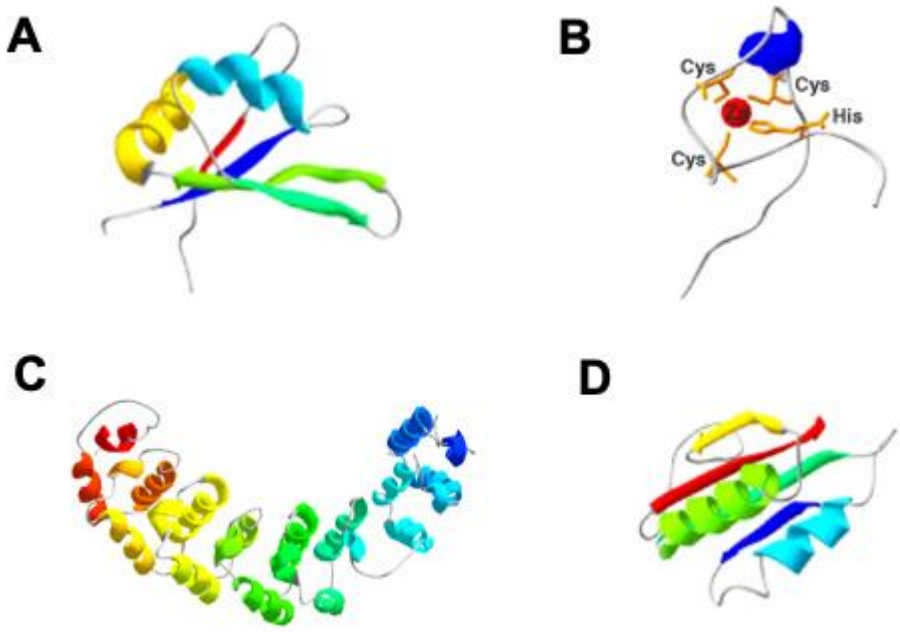


Figure 1.6: The major families of RNA-binding domains in kinetoplastid parasites. Representative structures of RNA-binding domains from *T. brucei* as predicted by the SWISS-MODEL homology modeling server. A) The RRM domain of RBP10. B) The CCCH zinc finger motif from ZFP1. C) The PUF9 Pumilio domain. D) The ALBA3 RNA-binding domain. Figure adapted from [Kolev, 2012].

Named for the *Drosophila* Pumilio and *Caenorhabditis elegans* Fem-3-binding factor proteins in which the domain was originally discovered, PUF proteins are known to promote mRNA decay and translational repression. The motif is characterized by a crescent-like structure composed of 8 consecutive alpha helical repeats, with each repeat directly interacting with a single nucleotide in the target RNA [Wang, 2018]. RNA elements bound by PUF proteins typically contain a core UGU motif at their 5' end; however, there are rare examples of PUF domains interacting with non-canonical sequences [Zhang, 2015]. There are 13 PUF proteins encoded in the genomes of kinetoplastid parasites [Clayton, 2007].

Finally, the Acetylation Lowers Binding Affinity (ALBA) motif is comprised of a four-stranded beta sheet and two alpha helices [Wardleworth, 2002]. The *T. brucei* and *Leishmania* genomes encodes four and two ALBA proteins, respectively [Subota, 2011; da Costa, 2017]. In both organisms, members of this protein family have been implicated in regulating developmentally expressed transcripts [Mani, 2011; Subota, 2011; Dupé, 2013].

RBPs in differentiation and adaptation

Most data linking RBPs to differentiation in kinetoplastid parasites are based on observations from *T. brucei*. Similar to *Leishmania*, *T. brucei* possess a multi-stage developmental program that requires passage through both a mammalian host and a hematophagous insect vector (i.e. the tsetse fly) [reviewed in Matthews, 2005]. In mammals, *T. brucei* replicate extracellularly in the blood, lymph, and interstitial fluids as bloodstream-form trypomastigotes (BSFs). As parasite numbers increase, BSFs transition

to a nondividing, intermediate stage known as the stumpy trypomastigote. Stumpy forms are taken up by the vector during blood-feeding, whereupon they differentiate into procyclic trypomastigotes and migrate in the anterior direction to colonize the tsetse proventriculus, the forward-most region of the midgut. Procyclic forms undergo an additional differentiation event to become epimastigotes before continuing on to invade the tsetse fly salivary glands. The cycle is completed by transformation to the mammalian-infective metacyclic trypomastigote stage, functionally analogous to the *Leishmania* metacyclic promastigote described in Section 1.2.

The first RBPs implicated in this process were three CCCH-family members known as ZFP1-3. ZFP1 is procyclic-specific while ZFP2 and 3 are constitutively expressed throughout the *T. brucei* lifecycle [Hendriks, 2001; Paterou, 2006]. All three proteins can be co-immunoprecipitated from procyclic cell extracts and genetic depletion of either ZFP1 or ZFP2 prevents procyclic differentiation [Paterou, 2006; Hendriks, 2005 and 2001]. In the case of ZFP3, ectopic overexpression in BSFs was sufficient to drive transformation to the procyclic stage [Paterou, 2006]. Global analysis of the ZFP3-bound transcriptome revealed a bias towards mRNAs that are specifically upregulated in stumpy-form trypomastigotes [Walrad, 2012]. The abundance of these ZFP3-selected messages was substantially increased by ZFP3 overexpression, indicative of a role in mRNA stability. On the other side of this equation is RBP10, an RRM domain-containing protein expressed exclusively in bloodstream-form *T. brucei*. RBP10 binds to procyclic-specific transcripts via a conserved $U(A)U_6$ RNA motif, targeting them for degradation [Mugo, 2017]. Forced RBP10 expression in BSF trypanosomes was found to prevent the

bloodstream-to-procyclic transformation, even in the presence of exogenous triggers, suggesting that this protein directly antagonizes procyclic development [Wurst, 2012].

Metacyclogenesis in *T. brucei* is mediated by a second RRM protein known as RBP6. RBP6 was originally identified based on its enrichment in trypanosomes occupying the tsetse fly proventriculus [Kolev, 2012]. Remarkably, ectopic expression of RBP6 in cultured procyclic forms was sufficient to induce differentiation, not only to the epimastigote stage but also to the transmissible metacyclic trypomastigote. Although the mRNA targets of RBP6 are not known, recombinant RBP6 was found to interact with an AU-rich motif *in vitro* [Clayton, 2019; Najafabadi, 2013]. The *T. brucei* ALBA proteins, ALBA1-4, are also implicated in vector-stage development. ALBA proteins are highly abundant, present at an estimated 10^4 copies per procyclic trypanosome [Mani, 2011; Clayton, 2013]. ALBA3 and 4 are specifically downregulated in the proventriculus and parasites depleted for ALBA3/4 display an elongated morphology, characteristic of the epimastigote stage [Subota, 2011]. All four ALBA proteins co-purify with one another and with components of the translation initiation machinery, suggestive of a role in translational control [Mani, 2011]. Still, a detailed mechanism of regulation remains to be clearly elucidated.

Only a handful of RBPs have been ascribed roles in the context of specific stress-response pathways. The best characterized example is the CCCH zinc finger protein ZC3H11, which is essential for *T. brucei* recovery from heat shock. ZC3H11 binds to a conserved AUU recognition motif in the 3'-UTRs of transcripts encoding *Hsp70* and other protein-folding chaperones [Droll, 2013]. Whereas most mRNAs aggregate in cytoplasmic stress granules under heat shock, ZC3H11-bound messages are preferentially

retained on polysomes, resulting in their continued translation [Kramer, 2008; Minia, 2016]. During the *T. cruzi* starvation response, ZC3H39 displays the opposite behavior, sequestering its mRNA targets to repress expression [Alves, 2014]. In this case, co-regulated messages share a common $A_3(C)A_2$ motif. Binding is differentially controlled such that, under nutrient stress, the ZC3H39-bound transcriptome is specifically enriched in targets with annotated functions in energy metabolism and translation. However, whether ZC3H39 is essential for survival during starvation is unclear.

Relative to *T. brucei*, next to nothing is known of the protein factors driving *Leishmania* stress responses and differentiation. Homologs of the RBPs described above likely exist in *Leishmania,* though few have been characterized to date. Among these are ALBA20 and ALBA13 from *L. infantum,* orthologs of *T. brucei* ALBA3 and ALBA1, respectively. Like their *T. brucei* counterparts, the *Leishmania* ALBA proteins are highly abundant and interact with one another *in vivo* [Dupé, 2013]. However, whereas *T. brucei* ALBA3 is implicated in differentiation within the vector, ALBA20 was found to stabilize transcripts encoding a number of amastigote-specific virulence factors during the intracellular stage of the *Leishmania* lifecycle [Dupé, 2013]. Thus, homologous RBPs do not necessarily fill equivalent roles in these organisms, at least with respect to development. Still, stress responses and differentiation are inextricably linked in the lifecycles of all three tritryp parasites, suggesting that these pathways may be regulated by common elements. Identifying the RBPs that govern either essential process may reveal key insights into the other.

The *Leishmania* purine stress response

As a major component of RNA, DNA, and other biomolecules within the cell, purines are central to a variety of cellular and metabolic processes. Like all parasitic protozoa, *Leishmania* lack the capacity for *de novo* purine synthesis and instead rely on salvage from the extracellular environment to satisfy their need for these essential nutrients [Marr, 1978]. As such, *Leishmania* have evolved a complex uptake and interconversion pathway to incorporate virtually any purine occurring naturally within the host [Figure 1.7; reviewed in Boitz, 2012].

In *L. donovani*, purine import is accomplished by four membrane permeases, collectively termed the *L. donovani* Nucleoside/Nucleobase Transporters, or LdNTs. LdNTs vary in transport activity and substrate specificity, with LdNT1 mediating uptake of pyrimidine and adenosine nucleosides [Vasudevan, 1998], LdNT2 transporting the nucleosides inosine, guanosine, and xanthosine [Carter, 2000], and LdNT3 and LdNT4 together responsible for the purine nucleobases adenine, hypoxanthine, guanine, and xanthine [Sanchez, 2004]. The LdNTs belong to the Equilibrative Nucleoside Transporter (ENT) family of proteins, which are conserved in mammals and other eukaryotes. While passive transport is the standard among most ENTs, the homologs found in *Leishmania* and other protozoan parasites function as proton symporters, utilizing the proton motive force to concentrate purines within the cell [Young, 2013; Stein, 2003; Landfear, 2004]. Additionally, the *Leishmania* plasma membrane is decorated with a variety of nucleases/nucelotidases and acid phosphatases, which process host-derived nucleic acids into suitable substrates for transport [Boitz, 2012].

Figure 1.7

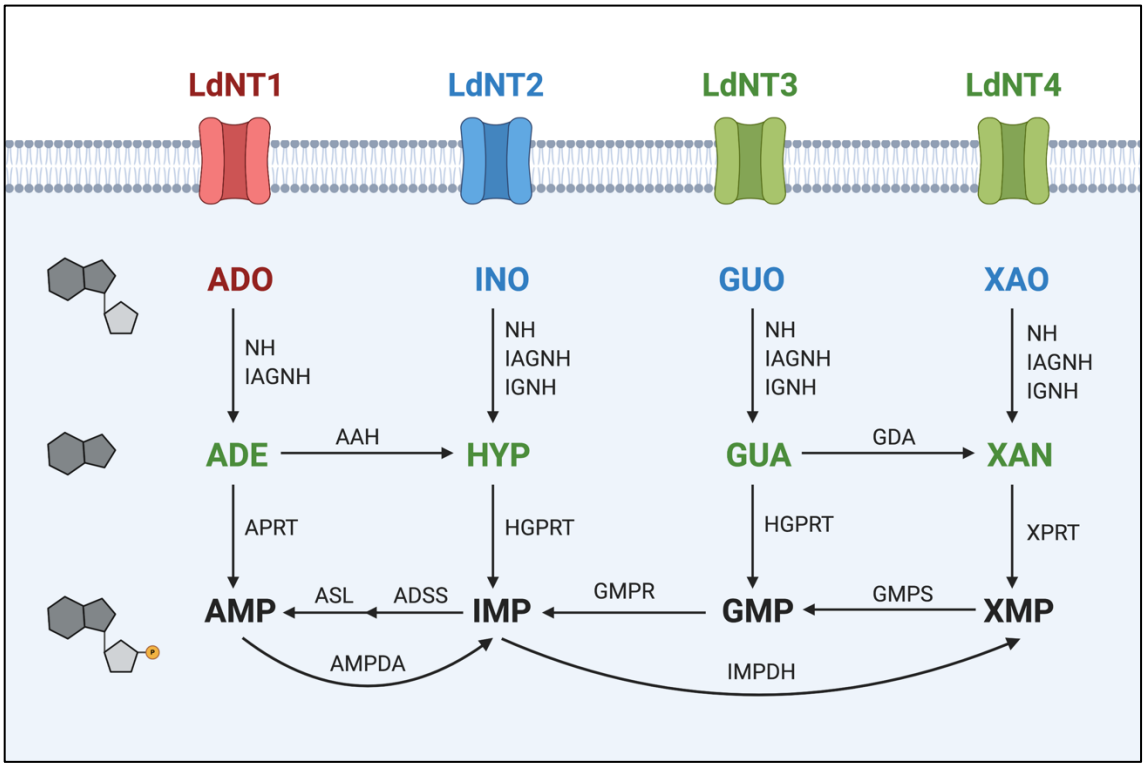


Figure 1.7: The *Leishmania* purine salvage and interconversion pathway. Purine salvage is meditated by four nucleoside (LdNT1 and 2) and nucleobase (LdNT3 and 4) transporters. Extracellular nucleic acids and nucleotides are converted to suitable substrates for uptake by surface 3' nucelotidases/nucleases and membrane acid phosphatases (not pictured). In purine-starved *L. donovani* promastigotes, LdNTs 1-3 are upregulated; LdNT4 is not. Transporters and their respective substrates are color coded. For ease of interpretation, greyscale diagrams (left side) indicate the nature of the purine metabolites listed in each row (i.e. nucleoside, nucleobase, or nucleotide). Abbreviations: ADO, adenosine; INO, inosine; GUO, guanosine; XAO, xanthosine; ADE, adenine; HYP, hypoxanthine; GUA, guanosine; XAN, xanthine; AMP, adenosine monophosphate; IMP, inosine monophosphate; GMP, guanosine monophosphate; XMP, xanthosine monophosphate; NH, nucleoside hydrolase; IAGNH, purine-specific nucleoside hydrolase; IGNH, inosine-guanosine nucleoside hydrolase; APRT, adenosine phosphoribosyltransferase; HGPRT, hypoxanthine-guanine phosphoribosyltransferase; XPRT, xanthine phosphoribosyltransferase; ASL, adenylosuccinate lyase; ADSS, adenylosuccinate synthetase; GMPR, GMP reductase; GMPS, GMP synthase; AMPDA, AMP deaminase; IMPDH, inosine monophosphate dehydrogenase. Adapted from [Martin, 2014] and [Boitz, 2012].

Following uptake by the parasite, purines are funneled into a complex metabolic network termed the purine salvage pathway [reviewed in Boitz, 2012]. While the reactions of the purine salvage pathway are interconnected and many are functionally redundant, it can be generally viewed as having two distinct branches dedicated to the metabolism of either adenylate or guanylate nucleotides. As shown in Figure 1.7, purines are readily converted between these two branches via the activity of IMP dehydrogenase (IMPDH) and GMP reductase (GMPR). By genetically uncoupling adenylate and guanylate metabolism (i.e. through deletion of both IMPDH and GMPR), it was previously shown that rather than directly sensing extracellular purines, *Leishmania* monitor the abundance of purines in the environment through surveillance of their intracellular nucleotide pools [Martin, 2016].

Despite their requirement for exogenous purines, *Leishmania* can survive for prolonged periods in the complete absence of these nutrients, suggesting that purine stress is a normal part of the *Leishmania* lifecycle [Carter, 2010; Martin, 2014]. Indeed, purine restriction is among a short list of other stressors inextricably woven into the developmental program of the parasite, required for efficient metacyclogenesis within the sand fly vector [Serafim, 2012]. Moreover, GMPR/IMPDH-deficient parasites incapable of sensing purine stress fail to survive long-term starvation *in vitro* and are significantly compromised in their ability to sustain a robust macrophage infection [Martin, 2016; Boitz, 2017]. Thus, the purine stress response is an important, parasite-specific aspect of *Leishmania* biology. Understanding its regulation may provide insight into mechanisms of development, adaptation and persistence in these and related kinetoplastid pathogens.

Global changes in purine-starved Leishmania

Our group and others have characterized features of the *Leishmania* purine stress response *in vitro*. In the absence of an exogenous purine source, cultured *L. donovani* promastigotes rapidly exit the cell cycle and adopt a characteristic elongated morphology [Carter, 2010]. Interestingly, these cells closely resemble the needle-like nectomonad stages that develop within the sand fly towards the end of bloodmeal digestion [Carter, 2010; Rogers, 2002]. Despite being growth-arrested, purine-starved *Leishmania* remain metabolically active and can resume replication at any point up to 90 days post-starvation by reintroduction of purines to the growth medium [Martin, 2014; Martin, 2016].

Using shotgun proteomics, it was previously demonstrated that purine starvation invokes a global remodeling of the *L. donovani* proteome, with changes segregating into two temporal phases [Martin, 2014]. In the first 6-24 hours, proteins involved in nucleic acid metabolism as well as ribosomal components are significantly downregulated. This is consistent with a global decrease in new mRNA and protein synthesis as would be expected of parasites exiting the cell cycle. At the same time, components of the purine salvage pathway are upregulated, presumably reflecting an attempt by *Leishmania* to increase uptake of the dwindling purine pool [Martin, 2014; Carter, 2010]. Later stages of adaptation (48 hr) are characterized by a more dramatic restructuring of cellular metabolism as parasites transition towards a long-term persistence phenotype. In this case, proteins involved in various catabolic processes, carbohydrate metabolism, and redox homeostasis are among those affected. Notably, protein-level changes do not track well with changes in the abundance of the corresponding mRNA, suggesting that

adaptation to purine stress is mediated by a variety of mechanisms operating at multiple post-transcriptional levels.

How is adaptation to purine starvation controlled?

Previous global studies were informative about *what* changes are involved in resisting purine stress, but they told us relatively little about *how* those changes are established. As described above, *Leishmania* and related kinetoplastids are incapable of regulating transcription on an individual gene basis. As a result, post-transcriptional control points (i.e. mRNA stability, translation, protein stability, etc.) are likely major determinants of protein abundance in these organisms. With the obvious exception of post-translational modification, RBPs are involved in virtually all aspects of post-transcriptional control. Thus, identifying purine-responsive RBPs and the *cis*-acting sequences they bind is the first step toward understanding regulation of the *Leishmania* purine stress response in molecular detail.

Among the earliest and most significantly upregulated genes in purine-starved *Leishmania* are three of the four membrane purine transporters, namely LdNTs 1-3 [Martin, 2014]. While LdNT1 and LdNT3 reflect changes at the levels of both mRNA stability and translation, the abundance of the LdNT2 transcript is unaffected by purine stress. This suggests that the purine transporters are not all controlled via the same post-transcriptional mechanisms. In addition, previous studies demonstrated that the 5' and 3' mRNA UTRs of these genes are sufficient to confer purine-responsive expression to a luciferase reporter, indicating that these regions encode distinct *cis*-acting regulatory elements [Soysa, 2014; Martin, 2014].

In my thesis studies, I approached the question of what factors regulate the purine stress response from the angle of mRNA. The LdNTs provided a tractable model for understanding purine-responsive gene expression in *Leishmania*. Thus, using a series of integrating reporter constructs, I performed deletional analysis of the *LdNT* mRNA UTRs to identify the *cis*-acting regulatory elements encoded within. In the experiments described in Chapter 2 of this thesis, I established that LdNT3 regulation depends on a repressive stem-loop in the mRNA 3'-UTR. I explored the evolutionary conservation of this element across other members of the order Kinetoplastida and, informed by these analyses, performed a careful molecular dissection of the stem-loop to identify critical residues required for its repressive activity. In Chapter 3, I turned my attention to nucleoside transporters LdNT1 and 2, both of which were found to harbor long pyrimidine-rich elements in their 3'-UTRs. I found that LdNT2 regulation is dependent upon this region alone; however, in the case of LdNT1, purine-responsive translation depends on an additional, preceding sequence feature. For all elements identified in this work, the mechanism of regulation was characterized and, in the case of LdNT3 and LdNT2, preliminary efforts were made to characterize the corresponding RBPs involved.

Chapter 2.

Purine-responsive expression of the *Leishmania donovani* NT3 purine nucleobase transporter is mediated by a conserved RNA stem-loop

Introduction

Sensing and responding to changes in the host environment is essential for the kinetoplastid parasite *Leishmania*. Purine stress in particular appears to factor regularly into the *Leishmania* lifecycle, as these organisms have evolved a robust stress response to cope with long-term purine deprivation. Previously, we characterized the global proteomic changes associated with purine stress [Martin, 2016; Martin, 2014]. These studies revealed a dramatic restructuring of cellular metabolism, geared toward reducing energy expenditure and enhancing general stress tolerance. The global transcriptomes of purine-starved cells were also significantly different; however, changes in mRNA abundance often tracked poorly with those manifested at the protein level, implicating both translational and post-translational mechanisms. Still, precisely *how* this signal is translated into an adaptive response remained ill-defined.

LdNT3 is one of the earliest and most substantially upregulated proteins in purine-starved *L. donovani* [Martin, 2014]. Past work from our laboratory established that LdNT3 upregulation is mediated at the levels of both mRNA abundance and translational efficiency, though protein stability was not examined [Martin, 2014; Soysa, 2014].

Additionally, it is published that a luciferase reporter flanked by the *LdNT3* mRNA UTRs is endowed with purine-responsive expression, indicating that either one or both of these regions encode distinct, *cis*-acting regulatory sequences. A short, regulatory element identified in the orthologous transporter from *T. brucei* was previously shown to control expression in response to both growth-stage and purine availability [Fernandez-Moya, 2014]. However, no such sequence was identified in *Leishmania.*

In the studies described in this chapter, we looked deeper at the molecular mechanisms underlying the leishmanial purine stress response through the lens of LdNT3. Using a novel application of a dual-luciferase reporter system developed in our laboratory [Soysa, 2014], we demonstrated that LdNT3 protein stability is not altered under purine stress, establishing the dominance of mRNA-centric control points in purine-responsive LdNT3 regulation. Based on homology to the *T. brucei* purine-response element, we identified a 33 nucleotide (nt) predicted stem-loop sequence in the *LdNT3* 3'-UTR (referred to as the LdNT3 stem-loop) that serves to repress expression when extracellular purines are abundant. We found that the *LdNT3* stem-loop is sufficient to confer purine-responsive regulation in heterologous sequence contexts but that, despite sharing significant sequence identity, the orthologous regulatory element from *T. brucei* does not function in *Leishmania*. Based on these data, we conducted a thorough mutational analysis to identify the functionally important regions required for repressor activity in purine-replete *L. donovani*. Lastly, we found that the *LdNT3* stem-loop is sufficient to confer purine-responsiveness to a high-abundance transcript, suggesting that the cognate RBP responsible for binding this element *in vivo* is present within the cell in substantial excess of what is required to regulate *LdNT3* expression alone.

Results

LdNT3 protein stability is not regulated in response to purine stress

To elucidate the molecular mechanisms that coordinate purine-responsive LdNT3 expression, we first determined the levels at which they operate. It is published that changes in both mRNA stability and translational efficiency contribute to LdNT3 upregulation under purine starvation [Martin, 2014; Soysa, 2014]. However, these earlier studies ignored the potential contribution of post-translational stabilization. In our experience, epitope-tagged LdNT3 is refractory to direct detection by western blotting at the low levels expressed in purine-replete cultures. Therefore, we modified a dual-luciferase system established in our laboratory to indirectly measure changes in LdNT3 protein stability via enzymatic reporter assay.

In the simplest iteration of the dual-luciferase system (described at length in Soysa, 2014), the firefly luciferase gene (*Fluc*) is fused in-frame with a selectable drug resistance marker and integrated in place of one allelic copy of the gene of interest. Reporter integration fully replaces the CDS of the targeted gene while preserving the endogenous intergenic regions (IGRs), which contain the requisite signals for trans-splicing and polyadenylation. Importantly, as the 5' and 3' mRNA UTRs are derived from these up- and downstream IGRs, respectively, their contributions to regulation are reflected in luciferase activity. To normalize Fluc activity between replicates, *Renilla* luciferase (*Rluc*) is similarly integrated into the locus of a gene for which expression does not change under the conditions of the experiment. In probing the purine stress response, we used UMP synthase (*UMPS*) as an unresponsive control, since neither UMPS mRNA nor protein abundance are affected by purine starvation [Martin, 2014; Soysa, 2014].

To adapt this system for the study of post-translational stability, we generated cell lines in which the *LdNT3* CDS was fused via its N-terminus to a Fluc reporter and integrated into either its endogenous locus or that of a purine-unresponsive control, *LdNT4* (Figure 2.1A). The multicistronic constructs used for integration contained a blasticidin resistance gene (*BSD*) to facilitate mutant selection and a 2A peptide from the *Thosea asigna* virus (*2A*) that is co-translationally cleaved, liberating the BSD-2A polypeptide to minimize the size of appended tag on LdNT3 [Szymczak, 2004; de Felipe, 2003]. In this configuration, the post-translational fate of the reporter is coupled to that of the transporter while mRNA stability and translation are governed by the native *LdNT3* UTRs and/or CDS such that changes in luciferase activity reflect the cumulative effect of mechanisms operating at all post-transcriptional levels. In contrast, the purine nucleobase transporter 4 (LdNT4), while homologous to LdNT3, is not differentially regulated with respect to purine availability and Fluc expression from this locus consistently reflects an absence of purine sensitivity [Martin, 2014]. Thus, expression of Fluc-LdNT3 integrated into this locus reflects only regulation conferred by elements contained within the *LdNT3* CDS itself or directly affecting stability of the protein. As anticipated, endogenously tagged Fluc-LdNT3 was significantly upregulated by 24 hours of purine starvation, consistent with previous experiments that implicated the UTRs in LdNT3 regulation. In cells expressing *Fluc-LdNT3* flanked by neutral *LdNT4* UTRs, however, luciferase activity was not affected by purine stress (Figure 2.1B). These data indicate that the *LdNT3* CDS does not encode additional *cis*-acting purine response elements nor does protein stability contribute to LdNT3 upregulation in purine-starved parasites. Thus, in

Figure 2.1

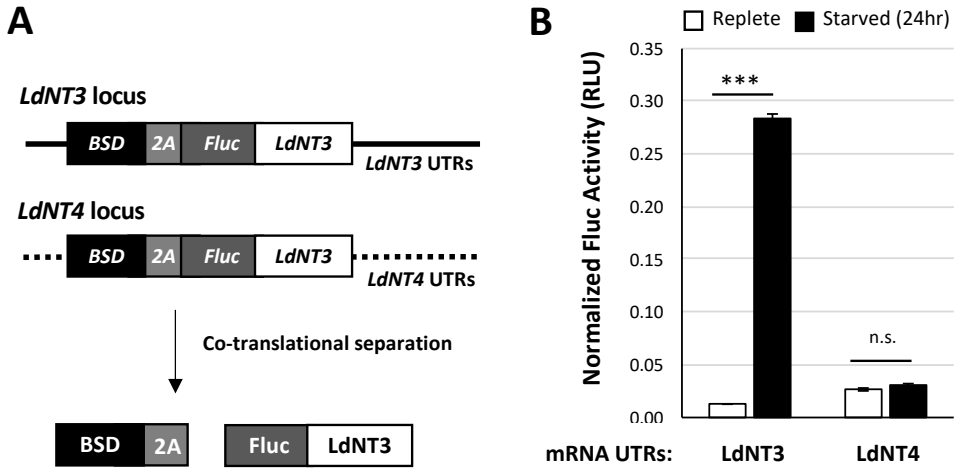


Figure 2.1: Changes in protein stability do not contribute to LdNT3 upregulation in purine-starved *Leishmania donovani*. A) Schematic of the genetic approach used to measure purine-responsive changes in LdNT3 protein-stability. Reporter constructs encoding Fluc-tagged LdNT3 were integrated into either the *LdNT3* or *LdNT4* locus as pictured. When flanked by native *LdNT3* mRNA UTRs (top), Fluc activity reflects the cumulative effect of regulation conferred at all post-transcriptional levels (i.e. mRNA stability, translational efficiency, and protein half-life). When flanked by purine-unresponsive *LdNT4* mRNA UTRs (bottom), changes in Fluc activity reflect only regulation conferred in *cis* by the *LdNT3* CDS itself or at the level of LdNT3 protein stabilization. Solid and dashed lines indicate purine-responsive and -unresponsive mRNA UTRs, respectively. Not pictured: In this and all subsequent experiments, a *Renilla* luciferase-puromycin resistance gene fusion (*Rluc-PAC)* expressed from the *UMPS* locus serves as an internal normalization control. B) Luciferase activity from cell lines depicted in A, measured after 24 hours of culture in the presence (replete) or absence (starved) of purines. Figure shows the mean and standard deviation of experiments performed in biological and technical duplicate. Asterisks (*) indicate significance: single-factor ANOVA calculated with Excel Descriptive Statistics Toolpak; n.s., not significant, $P \geq 0.05$; ***$P \leq 0.001$.

the absence of post-translational control, transcript stability and translational efficiency serve as primary control points mediating purine-responsive changes in LdNT3 abundance.

A 33 nt stem loop in the LdNT3 mRNA 3'-UTR represses expression under purine-replete conditions

The *LdNT3* 5'- and 3'-UTRs are together sufficient to confer purine-responsiveness to a reporter [Martin, 2014; Soysa, 2014], strongly implicating the presence of *cis*-acting regulatory sites within one or both of these regions. The 3'-UTR of the orthologous *Trypanosoma brucei* purine nucleobase transporter, *TbNT8.1*, encodes a predicted stem-loop that both is necessary and sufficient to repress TbNT8.1 expression when extracellular purines are abundant [Fernández-Moya, 2014]. Based on homology to this region, we identified a similar 33 nt predicted stem-loop in the *LdNT3* 3'-UTR, approximately 2.73 kb downstream of the stop codon (Figures 2.2A and B). While the exact secondary structure of this element *in vivo* has not been formally demonstrated, we will refer to these sequences as stem-loops throughout for convenience. Though absent from the UTRs of other purine-responsive genes in *L. donovani* (data not shown), this element was identified in the UTRs of orthologous purine transporters from a variety of kinetoplastids, suggesting a strong evolutionary pressure for conservation (Figures 2.2A and 2.S1). To test whether this region also confers purine-responsiveness in *L. donovani*, we generated cell lines in which a *Fluc-BSD* transgene was expressed from the endogenous *LdNT3* locus under the control of either wildtype UTRs or a modified 3'-UTR lacking the putative stem-loop (Figure 2.2B) Parasites expressing *Fluc-BSD* flanked by wildtype UTRs demonstrated an approximate 27-fold increase in luciferase activity in response to purine stress. The magnitude of this effect was diminished by nearly ninety-

Figure 2.2

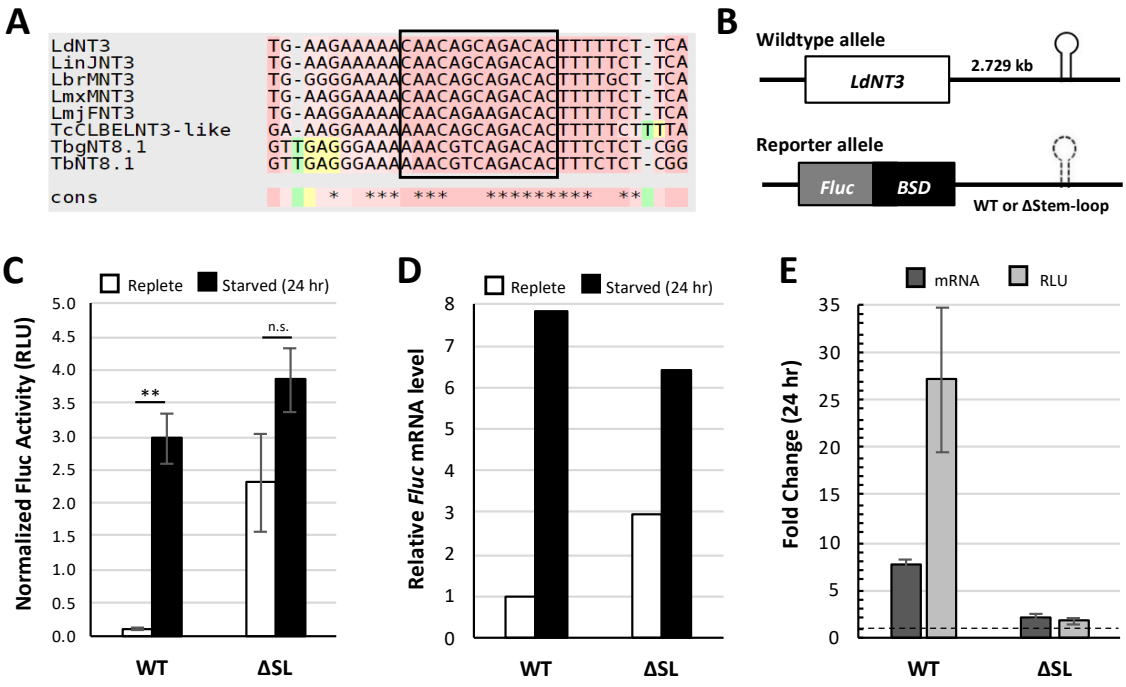


Figure 2.2: A 33 nt stem-loop in the *LdNT3* mRNA 3'-UTR mediates transcript instability and translational repression under purine-replete conditions. A) Alignment of putative purine-response elements from *Leishmania donovani* (LdNT3; LdBPK_131110.1), *L. infantum* (LinJNT3; LINF_130017100), *L. braziliensis* (LbrMNT3; LbrM.13.0990), *L. mexicana* (LmxMNT3; LmxM.13.1210), *L. major* (LmjFNT3; LmjF.13.1210), *Trypanosoma cruzi* (TcCLBELNT3-like; TcCLB.511051.30), and *Trypanosoma brucei* subspecies, *T. b. gambiense* (TbgNT8.1; Tbg972.11.4110) and *T. b. brucei* (TbNT8.1; TB927.11.3610). Sequences were aligned using the T-Coffee web server with default parameters [Di Tommaso, 2011]. Loop region, as predicted by the mFold web server [Zuker, 2003], is indicated with a box B-E) Investigating the effects of the *LdNT3* stem-loop on mRNA stability and/or translation by paired dual-luciferase and RT-qPCR analyses. Where indicated, bars and error bars represent the mean and standard deviation of assays performed with 2 independent clones. B) Wildtype and reporter alleles at the endogenous *LdNT3* locus. Individual cell lines express *Fluc-BSD* under the control of either native *LdNT3* UTRs (WT) or a mutant 3'-UTR lacking the *LdNT3* stem-loop (ΔSL). C) Luciferase activity from cell lines depicted in B after 24 hours of culture in the presence or absence of purine. D) Quantitation of *Fluc-BSD* transcripts from cultures assayed in C. Relative message level was determined by the comparative CT method using *UMPS* as an endogenous control gene and 'WT replete' as a reference sample. For ΔCT values and analysis details, refer to Table A.1. E) Fold change in *Fluc-BSD* mRNA level and luciferase activity after 24 hours of purine starvation. The comparative CT method was used to determine mRNA fold change for individual clones as shown in Table A.2. Single-factor ANOVA was calculated with Excel Descriptive Statistics Toolpak: ** $P \leq 0.01$; ***$P \leq 0.001$.

four percent in stem-loop-deficient mutants, wherein Fluc activity was increased only ~1.7-fold by starvation (Figures 2.2C and 2.2E). Specifically, deletion of the stem-loop resulted in a ~20.6-fold increase in basal Fluc expression under purine-replete conditions, consistent with the sequence functioning as a negative regulator (Figure 2.2C, white bars). Together, these data implicate the *LdNT3* stem-loop as a major regulator of purine-responsive LdNT3 expression.

We considered two potential mechanisms for how repressive control by the *LdNT3* stem-loop is achieved. In one scenario, repression is alleviated under purine stress via reduced binding and/or post-translational modification of a repressive RBP that associates with the stem-loop. In the alternative scenario, purine stress causes alternative polyadenylation site selection, resulting in a truncated *LdNT3* mRNA that does not include the repressive element. To distinguish between these two possibilities, the presence of the stem-loop in *LdNT3* mRNAs from purine-starved and -replete cultures was assessed via RT-PCR. An oligo-dT primer was used for first-strand cDNA synthesis to ensure that only mature polyadenylated mRNAs, but not unprocessed nuclear transcripts, were included in the analysis. As shown in Figure 2.S2, the repressor stem-loop was readily amplified from mature transcripts irrespective of purine availability, excluding the possibility that alternative polyadenylation contributes to the mechanism of stem-loop-mediated repressive control.

It is known that increases in both mRNA stability and translational efficiency contribute to LdNT3 upregulation under purine stress [Martin, 2014; Soysa, 2014]. To test whether the *LdNT3* stem-loop affects either or both of these processes, total RNA was isolated from the cultures analyzed in Figure 2C and *Fluc-BSD* transcripts were

quantified by RT-qPCR (Figure 2.2D). For each cell line, the fold-change in *Fluc-BSD* message level was compared to the corresponding change in luciferase activity to discern the relative contributions of mRNA stability and translation to reporter regulation. In parasites expressing *Fluc-BSD* flanked by wildtype *LdNT3* UTRs, luciferase activity increased ~27-fold in response to purine starvation while *Fluc-BSD* message levels increased just ~7.8-fold (Figure 2.2E). The disparity in these changes intimates that in addition to regulating mRNA stability, the *LdNT3* UTRs mediate an approximate 3.5-fold increase in translation, consistent with our previous data on LdNT3 regulation [Martin, 2014; Soysa, 2014]. In the case of *Fluc-BSD* transcripts lacking a stem-loop, the magnitude of the change in mRNA abundance was reduced to just over 2-fold. This was essentially proportional to the change in luciferase activity measured from the same cultures (Figure 2.2E). Hence, stem-loop deletion completely abolished the translational component to regulation bestowed by the *LdNT3* UTRs, strongly implicating this element in purine-responsive translational control. Additionally, the steady-state *Fluc-BSD* message level was ~3-fold higher among stem-loop-deficient mutants than when flanked by wildtype UTRs (Figure 2.2D, white bars). These data suggest that the *LdNT3* stem-loop also functions, in part, through destabilization of the transcript under purine-replete conditions.

We next asked whether this RNA element was sufficient to confer regulation to a reporter expressed from the *LdNT4* locus, which is normally not affected by purine stress. We integrated constructs encoding the Fluc-BSD reporter flanked by either wildtype *LdNT4* UTRs or an *LdNT4* 3'-UTR harboring the *LdNT3* stem-loop at one of two different positions (Figure 2.3A) that were predicted via Mfold to preserve stem-loop

folding [Zuker, 2003]. Based on publicly available data for the polyadenylation site of *L. major NT4* (TriTrypDB.org), the predicted length of the *LdNT4* 3'-UTR is ~1.43 kb, which would include the sites in which the *LdNT3* stem-loop was inserted (242 and 419 nt downstream of the *LdNT4* translation stop). Nonetheless, because the 3' terminus of the *LdNT4* transcript has not been definitively mapped, we verified via RT-PCR that both stem-loop insertion sites were included in the 3'-UTRs of mature, polyadenylated *Fluc-BSD* transcripts, regardless of purine availability (Figure 2.S2). As shown in Figure 2.3B, the effects of stem-loop insertion were position-dependent. Placing the stem-loop 242 bases into the *LdNT4* 3'-UTR led to a significant decrease in basal Fluc-BSD expression, which translated to a ~6-fold increase in luciferase activity under purine-restricted conditions. In contrast, when inserted at position +419 the stem-loop did not confer differential expression, possibly reflecting an inability of the element to fold properly in this genetic context. Thus, the *LdNT3* stem-loop is sufficient for purine-responsive expression but is sensitive to sequence context.

Figure 2.3

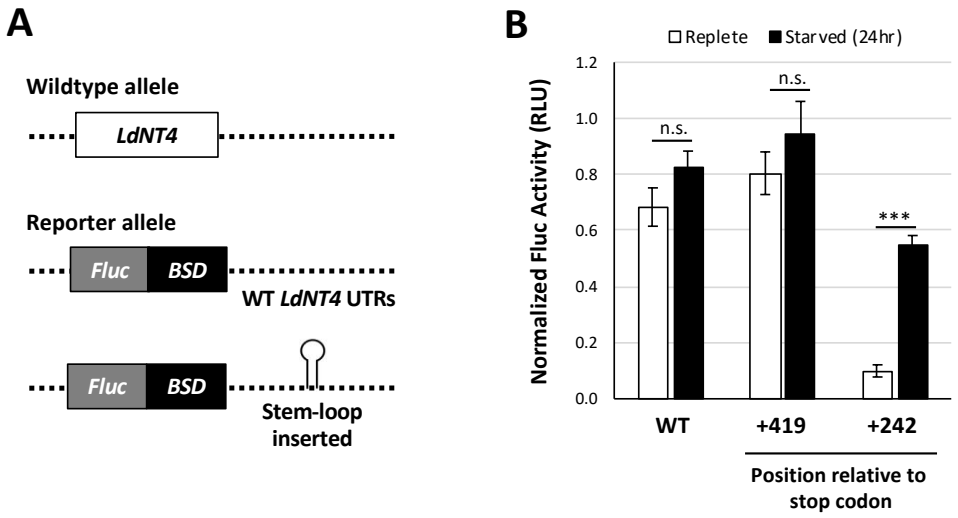


Figure 2.3: The *LdNT3* stem-loop is sufficient for purine-responsive regulation. A) Schematic of wildtype and reporter alleles at the endogenous *LdNT4* locus. Dashed lines indicate purine-unresponsive mRNA UTRs. In experimental cell lines, the *LdNT3* stem-loop was inserted into the *LdNT4* 3'-UTR, either 242- or 419-nt downstream of the stop codon. B) Normalized Fluc activity from cell lines depicted in A, after 24 hours of culture in the presence or absence of purines. Figure shows the mean and standard deviation of experiments performed in biological and technical duplicate. Single-factor ANOVA was calculated with Excel Descriptive Statistics Toolpak: n.s., not significant, $P \geq 0.05$; ***$P \leq 0.001$.

Regulation by the stem-loop is species-specific and depends upon conserved residues in the loop

The purine-response elements from *TbNT8.1* and *LdNT3* share a 33 nt core with 80% identity (Figures 2.2A and 2.4A). To determine if the orthologous *TbNT8.1* stem-loop was functional in *L. donovani*, we modified the reporter construct depicted in Figure 2.3A to insert the minimal *TbNT8.1* stem-loop at position +242 in the *LdNT4* 3'-UTR. While this element was sufficient for regulation in *T. brucei* [Martin, 2016], it was unable to confer purine-responsive expression to the *Fluc-BSD* reporter in *L. donovani* (Figure 2.4B), suggesting that the RBPs that associate with these elements in *L. donovani* and *T. brucei* have different binding specificities.

To determine if the inability of the *TbNT8.1* element to function in *Leishmania* is due to differences in the sequence of the stem, loop, or both regions, we generated chimeras in which the *TbNT8.1* loop sequence was appended to the stem of the *LdNT3* ortholog, and vice versa. These chimeric stem-loops were inserted at position +242 in the *LdNT4 Fluc-BSD* constructs and integrated into the *LdNT4* locus of a dual-luciferase compatible cell line. Parasites encoding a *TbNT8.1* loop on an *LdNT3* stem demonstrated a partial reduction in luciferase activity under purine replete conditions compared to the *TbNT8.1* stem-loop control, but there was no increase in expression upon purine starvation (Figure 2.4C). In contrast, Fluc-BSD expression was significantly repressed by the reciprocal *LdNT3* loop-*TbNT8.1* stem mutant in the presence of exogenous purine. This effect was reversed by 24 hours of purine stress, resulting in a similar level of luciferase induction to that conferred by the wildtype element (5.4- and 5.9-fold, respectively). These data suggest that the sequence of the *LdNT3* loop, but not stem, is essential for purine-responsive repressor activity in *L. donovani*.

Figure 2.4

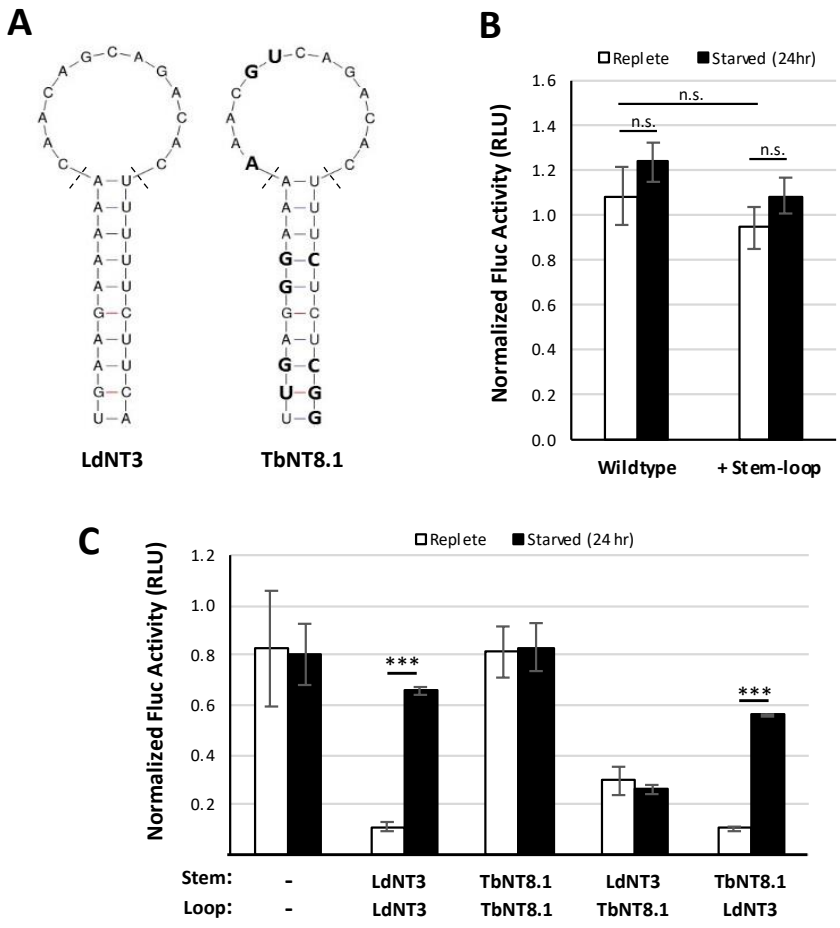


Figure 2.4: The sequence of the *LdNT3* loop, but not stem, is required for its function as a purine-response element in *L. donovani.* A) Secondary structures of the *LdNT3* and *TbNT8.1* stem-loops as predicted by mFold [Zuker, 2003]. Variant positions are highlighted on the *TbNT8.1* diagram in bold. Dashed lines delineate 'stem' vs 'loop' regions. B) Testing the regulatory capacity of the *TbNT8.1* purine-response element in *L. donovani.* C) Purine-responsive repression is conferred by a chimeric stem-loop encoding the *LdNT3* loop on a *TbNT8.1* stem, but not by the reciprocal swap. For both B and C, luciferase reporter lines were generated similarly to those depicted in Figure 2.4A, with either the *TbNT8.1* stem-loop (B) or *LdNT3-TbNT8.1* stem-loop chimeras (C) inserted at position +242 in the *LdNT4* 3'-UTR. A cell line expressing *Fluc-BSD* under the control of native *LdNT4* UTRs serves as a 'wildtype' control. Graphs display normalized luciferase activity measured after 24 hours of culture in the presence or absence of purines. Bars represent the mean and standard deviation of experiments performed in biological and technical duplicate. Single-factor ANOVA was calculated with Excel Descriptive Statistics Toolpak: n.s., not significant, $P \geq 0$.

We noted two distinct blocks of conservation within the loop regions of *LdNT3* stem-loop orthologs (labeled A and B in Figure 2.5A). We generated *LdNT3* stem-loop variants in which blocks A and B were mutated independently and tested their activity at the *LdNT4* locus. Disruption of either block resulted in a complete loss of regulation (Figure 2.5B), indicating that these conserved regions are important for the repressor function of the *LdNT3* stem-loop.

The *LdNT3* loop differs from that of *TbNT8.1* at just three positions (Figure 2.6A). To determine if the inactivity of the *TbNT8.1* stem-loop in *L. donovani* could be attributed to any one of these divergent bases, we generated a series of *TbNT8.1* stem-loop mutants in which each variant position was changed to the corresponding base from the leishmanial ortholog. A *TbNT8.1* stem with a wildtype *LdNT3* loop (labeled as 1-3 in Figure 2.6B) served as a positive control for purine-responsive induction. Like the wildtype *TbNT8.1* element, replacement of each of the three variant residues alone had no effect on luciferase activity, suggesting that species restriction is defined by multiple bases in the loop rather than any one individually. Similarly, simultaneous conversion of positions 1 and 2 to the corresponding *LdNT3* bases failed to restore purine-responsive regulation (1,2 in Fig. 2.6B). Two paired-position mutants (1,3 and 2,3) conferred varying degrees of repression that translated to a respective ~2.1-fold and ~3.7-fold increase in luciferase activity under purine stress; however, neither fully recapitulated the robust ~7.7-fold induction observed in control cells where all three bases were changed to their *LdNT3* counterparts. Thus, each of the three nonconserved bases is important for species specificity of the repressor stem-loops, and full repressor function in *L. donovani* depends on the sequence at all three positions. Interestingly, the sequence of the 2,3 loop

Figure 2.5

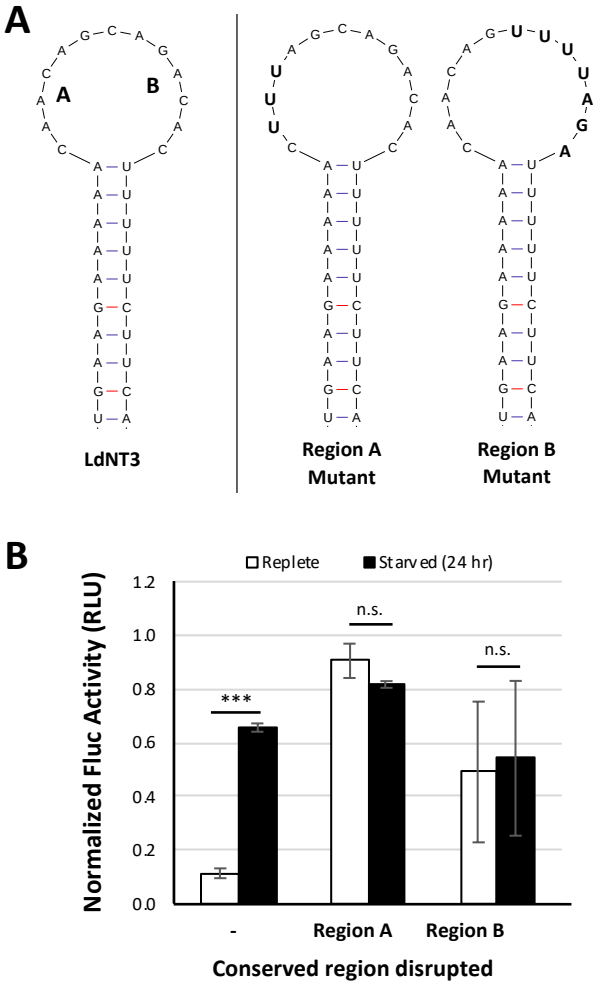


Figure 2.5: Evolutionarily conserved residues within the *LdNT3* loop are functionally important for purine-responsive gene expression. A) Evolutionarily conserved residues (Regions A and B) are highlighted in grey and their corresponding mutants are shown in boldface type. B) Reporter lines were generated similarly to those depicted in Figure 2.4A, with Region A and B stem-loop mutants inserted at position +242 in the *LdNT4* 3'-UTR. Parasites expressing *Fluc-BSD* under the control of an *LdNT4* 3'-UTR harboring a wildtype *LdNT3* stem-loop serve as a positive control for regulation. Graph displays normalized Fluc activity measured after 24 hours of culture in the presence or absence of purines. Bars and error bars represent the mean and standard deviation of experiments performed in biological and technical duplicate. Single-factor ANOVA was calculated with Excel Descriptive Statistics Toolpak: n.s., not significant, $P \geq 0.05$; ***$P \leq 0.001$.

Figure 2.6

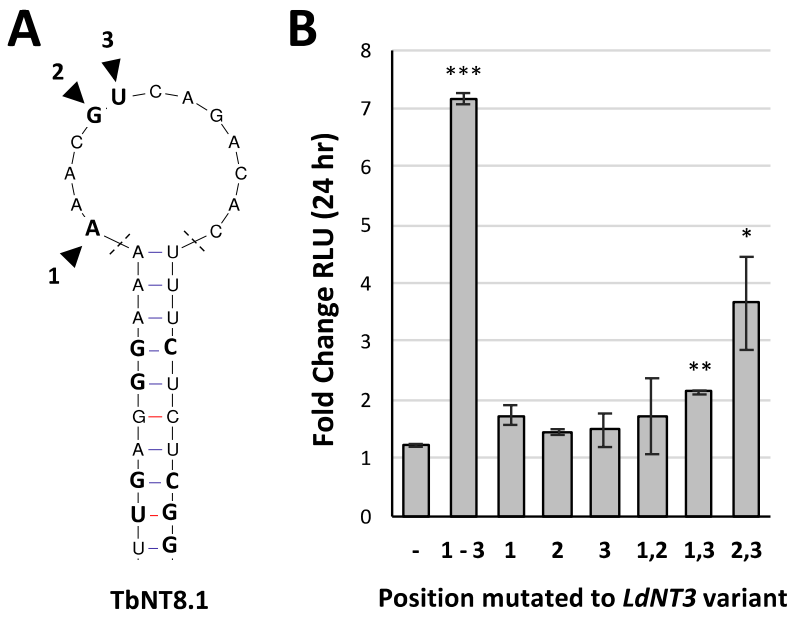


Figure 2.6: Species specificity of the purine-response element is defined by three bases within the loop. A) Three bases differentiate the activity of the *LdNT3* vs *TbNT8.1* loop in *L. donovani*. The positions at which the *LdNT3* and *TbNT8.1* stem-loops vary are shown in bold typeface; the loop residues subjected to mutational analysis in B are numbered. B) Regulation by single and paired-position *TbNT8.1* stem-loop mutants. Variant loop residues were mutated, either individually or in pairs, to the corresponding base encoded by the *LdNT3* sequence. Stem-loop mutants were inserted into the 3'-UTR of *LdNT4 Fluc-BSD* reporter constructs as shown in Figure 2.4A, and their regulatory capacity assessed via dual-luciferase assay. Cell lines harboring a wildtype *TbNT8.1* stem-loop (-) or a chimeric *LdNT3* loop-*TbNT8.1*-stem element (1-3) serve as negative and positive controls, respectively. Bars represent the mean and standard deviation of experiments performed in biological and technical duplicate. Asterisks (*) indicate significance. Single-factor ANOVA was calculated using Excel Descriptive Statistics Toolpak and post-test Bonferroni-corrected *P*-values are as follows: $*P \geq 0.007143$; $**P \leq 0.001428$; $***P \leq 0.0001428$.

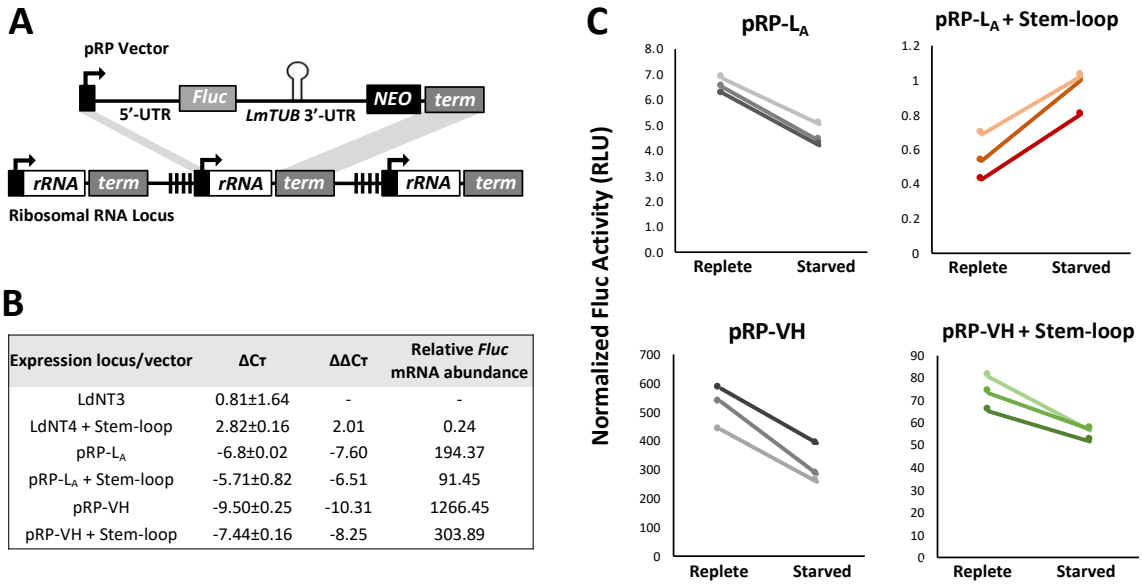
variant (Figure 2.6) corresponds to that from the orthologous *T. cruzi* transporter (Figure 2.2A), suggesting that the binding specificities of the cognate RBPs that associate with these stem-loops in *L. donovani* and *T. cruzi* have likely also diverged.

Regulation conferred by the LdNT3 stem-loop is likely mediated by a highly abundant trans-acting factor

The steady-state mRNA level of the *Fluc-BSD* transgene flanked by *LdNT4* UTRs and harboring an *LdNT3* stem-loop is approximately 75% lower than that of the same transgene expressed from the endogenous *LdNT3* locus (Figure 2.7B). Having demonstrated that the *LdNT3* stem-loop is sufficient to confer purine-responsiveness to this lower-copy message, we next asked if the sequence could also mediate regulation of more abundant transcripts. The pRP vectors are a set of integrating rRNA promoter vectors generated in our laboratory that offer a range of incrementally different expression profiles in *L. donovani*. In all configurations, Pol-I drives robust transgene transcription from the rRNA array. Graded expression is achieved using different combinations of UTRs that vary in their respective abilities to either promote or attenuate mRNA processing and stability [Soysa, 2015]. As depicted in Figure 2.7A, we introduced the *LdNT3* stem-loop into pRP-L_A and pRP-VH, low- and high-expressing vectors from the pRP suite. These constructs differ in the sequences of their 5'-UTRs but share an identical 3'-UTR derived from the *L. major* α-tubulin (*LmTUB*) intergenic region, facilitating the comparative study of the *LdNT3* stem-loop at two steady-state transcript levels while eliminating the confounding variable of local genetic context. A *Fluc* reporter gene was inserted into the multiple cloning sites of these vectors and linearized constructs were integrated into a dual-luciferase compatible cell line. As expected, the

relative abundance of the *Fluc* transcript was substantially higher in these cells than when expressed from the endogenous *LdNT3* locus, to over 90- and 300-fold in the case of stem-loop-modified pRP-L_A and pRP-VH, respectively. Relative to control cell lines harboring an unmodified 3'-UTR, the presence of a stem-loop in either vector substantially reduced *Fluc* mRNA levels, consistent with our previous observation that this sequence negatively affects mRNA stability (Figure 2.7B). To test whether the *LdNT3* stem-loop could mediate purine-responsive regulation of these higher-copy transcripts, cells were cultured for 48 hours in the presence or absence of purines and subjected to dual-luciferase analysis. Fluc activity was substantially reduced in both pRP-L_A and pRP-VH control cells by 48 hours of purine stress, likely reflecting a general decrease in Pol I-mediated transcription of the rRNA array (Figure 2.7C, left). This is supported by previous observation that Pol I protein levels are significantly downregulated in purine-starved *L. donovani* [Martin, 2014]. Interestingly, insertion of the *LdNT3* stem-loop into the 3'-UTR of pRP-L_A, but not pRP-VH, resulted in a moderate upregulation of luciferase activity upon starvation (Figure 2.7C, right), which overcame the general starvation-induced reduction in expression from the pRP vectors. The fact that the *LdNT3* stem-loop conferred purine responsive regulation to an mRNA ~90-fold more abundant than *LdNT3* suggests that the *trans*-acting factor(s) that binds this element is in considerable excess of what is required solely for *LdNT3* regulation. Given that the two vectors encode identical 3'-UTRs with the *LdNT3* stem-loop inserted at the same position, the failure of the stem-loop to confer upregulation in the context of the pRP-VH construct is consistent with the possibility that the high level of mRNA expressed from this construct exceeded the availability of the cognate binding protein.

Figure 2.7



Expression locus/vector	ΔCT	ΔΔCT	Relative *Fluc* mRNA abundance
LdNT3	0.81±1.64	-	-
LdNT4 + Stem-loop	2.82±0.16	2.01	0.24
pRP-L_A	-6.8±0.02	-7.60	194.37
pRP-L_A + Stem-loop	-5.71±0.82	-6.51	91.45
pRP-VH	-9.50±0.25	-10.31	1266.45
pRP-VH + Stem-loop	-7.44±0.16	-8.25	303.89

Figure 2.7: Regulation via the *LdNT3* stem-loop is likely mediated by a highly abundant *trans*-acting factor. A) Differential high-level Fluc expression is achieved using integrating rRNA promoter vectors. pRP-L_A and pRP-VH 5'-UTRs are derived from the upstream IGRs of the *T. brucei* procyclic acidic repetitive protein and *Crithidia fasciculata* phosphoglycerate kinase B gene, respectively. A 3'-UTR derived from the *L. major* α-tubulin (*LmTUB*) intergenic region is common to both vectors. For additional details on pRP vectors, refer to [Soysa, 2015]. Abbreviations and symbols: term = putative rRNA terminator sequence; NEO = neomycin resistance gene; black boxes = rRNA promoter; vertical hash marks = 64 base repeats. Black arrows indicate the direction of transcription by RNA polymerase I. Grey shading indicates homology for targeted integration. B) Relative abundance of *Fluc* mRNA expressed from the *LdNT4* locus or from pRP-L_A and pRP-VH vectors, as calculated via the comparative CT method. For each cell line, *Fluc* transcript level was normalized to *UMPS* to generate ΔCT, which was then used to calculate ΔΔCT and the fold-difference in *Fluc* abundance ($2^{-\Delta\Delta CT}$), relative to the same transgene expressed from the *LdNT3* locus. In pRP-L_A and pRP-VH constructs labeled "+ Stem-loop", the *LdNT3* stem-loop was inserted into the 3'-UTR 65 nt downstream of the *Fluc* translation stop. The mean and standard deviation from two biological replicates are shown for each analysis. C) Fluc activity from the pRP-L_A and pRP-VH vectors after 48 hours of culture in the presence or absence of purines. End points represent the mean of assays performed in technical duplicate for three independent clones of each cell line.

Discussion

Responding to the nutritional environment of the host is critical for successful parasitism. Previous studies from our group demonstrated that purine starvation invokes a robust nutrient stress response in *Leishmania donovani* that is characterized by a marked remodeling of the cellular proteome. It was established that purine nucleoside and nucleobase transporters, including LdNT3, are highly upregulated by purine stress, and that this regulation is mediated, at least in part, via regulatory elements encoded in 5'- and/or 3'-UTRs of their mRNAs [Carter, 2010; Martin, 2014; Soysa, 2014]. In the studies described in this chapter, we have built on these observations by performing a detailed examination of the post-transcriptional regulation of the LdNT3 purine nucleobase transporter in response to purine starvation.

Based on homology to a known purine-response element from *T. brucei* [Fernández-Moya, 2014], we identified a 33 nt stem-loop in the 3'-UTR of *LdNT3* that serves to repress expression when purines are abundant via both transcript destabilization and translational downregulation (Figure 2.2). This is the first report of a defined nutrient stress-response element in *Leishmania.* Deletion of this element almost entirely ablated purine-responsive control by the *LdNT3* UTRs. Whereas previous work established that both *LdNT3* mRNA stability and translation are increased by purine starvation [Martin, 2014; Soysa, 2014], we found that post-translational stability of LdNT3 is not affected by extracellular purine level. Thus, in the absence of a post-translational contribution, this singular element appears to be the primary factor mediating purine-responsive changes in LdNT3 abundance.

The putative stem-loop described in this work is conserved across a variety of dixenous and monoxenous kinetoplastids including multiple *Leishmania* species, *T. brucei* subspecies, and *Trypanosoma cruzi* (highlighted in Figure 2.2A) as well as *Crithidia fasciculata*, *Leptomonas pyrrhocoris*, and *Blechomonas ayalai* (Figure 2.S1). Conservation is particularly striking given that the relative position of the stem-loop varies substantially with respect to the stop codon and the 3'-UTRs of the orthologous transporters are otherwise poorly conserved. Thus, despite substantial expansion or contraction of the UTRs, this sequence has been maintained throughout evolutionary history, suggesting a strong selective pressure to maintain purine-responsive regulation. Interestingly, the stem-loop is absent from the orthologous purine transporter gene of the free-living kinetoplastid *Bodo saltans,* which also lacks the capacity for *de novo* purine synthesis [Opperdoes, 2016]. This observation may indicate that a robust purine stress response is uniquely important to the parasitic lifestyle.

As was shown for the *TbNT8.1* stem-loop in *T. brucei* [Fernández-Moya, 2014], the *LdNT3* stem-loop is sufficient to confer regulation to an otherwise purine-unresponsive reporter in *L. donovani*. However, sequence context appears to be important, since only one of two positions into which the *LdNT3* stem-loop was inserted supported repressor activity (Figure 2.3). This likely reflects differences in the propensities of the sequences surrounding the insertion points to negatively impact folding and/or accessibility of the stem-loop. The sequences of the *TbNT8.1* and *LdNT3* stem-loops differ by only 20%. We were therefore surprised to find that the two elements are not functionally equivalent in *L. donovani*. Species-restriction was attributed to just three variant bases encoded within the loop. One plausible explanation for this restriction

may be that that orthologous RBPs that bind to these repressor elements have different specificities. However, this hypothesis is not currently testable as the relevant *trans*-acting factor in *T. brucei* was not identified and we have yet to isolate a candidate RBP for the *LdNT3* stem-loop, despite numerous attempts using a range of biochemical approaches.

Several genes are differentially expressed in response to purine stress, yet the *LdNT3* purine-response element was not identified elsewhere in the *L. donovani* genome via bioinformatic analysis (unpublished observation). However, this does not preclude the possibility that *LdNT3* is part of a purine-responsive regulon consisting of multiple genes under the control of a common RBP, as the ability of individual RBPs to interact with several disparate binding sites is well-documented [Jolma, 2020; Dominguez, 2018]. Indeed, the observation that the stem-loop is sufficient to confer purine responsiveness to a transcript that is over 90-fold more abundant than *LdNT3* (Figure 2.7) strongly suggests that its binding partner is present in substantial excess of what is required for *LdNT3* regulation, and may therefore play a role in regulating other genes within or outside of the purine stress response pathway. Despite this possibility, we suspect that the *L. donovani* purine stress response is likely mediated by multiple independent but intersecting pathways. For instance, each of the membrane purine transporters appears to be regulated by a unique combination of post-transcriptional regulatory mechanisms; LdNT2 is regulated solely at the translational level, while both mRNA abundance and translation are altered for LdNT1 and LdNT3 during purine stress [Martin, 2014]. Future efforts to identify the proteins that associate with these *cis*-acting elements will help to unravel the complexities of the purine stress response.

Figure 2.S1

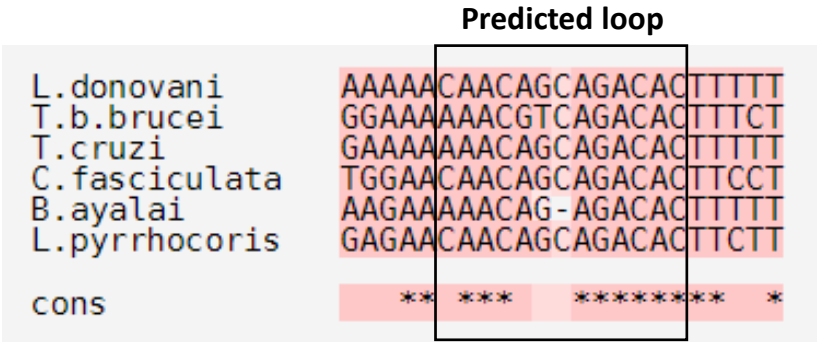


Figure 2.S1: The *LdNT3* stem-loop is conserved in the 3'-UTRs of orthologous purine transporters from a variety of kinetoplastid parasites. Alignment of the *L. donovani* (LdBPK_131110.1) and *T. b. brucei* (TB927.11.3610) stem-loops with putative response elements from the related *T. cruzi* (TcCLB.511051.30), *C. fasciculata* (CFAC1_220041000), *B. ayalai* (Baya_012_0780), and *L. pyrrhocoris* (LpyrH10_20_0660). The alignment was generated with the T-Coffee web server using default parameters (1).

Figure 2.S2

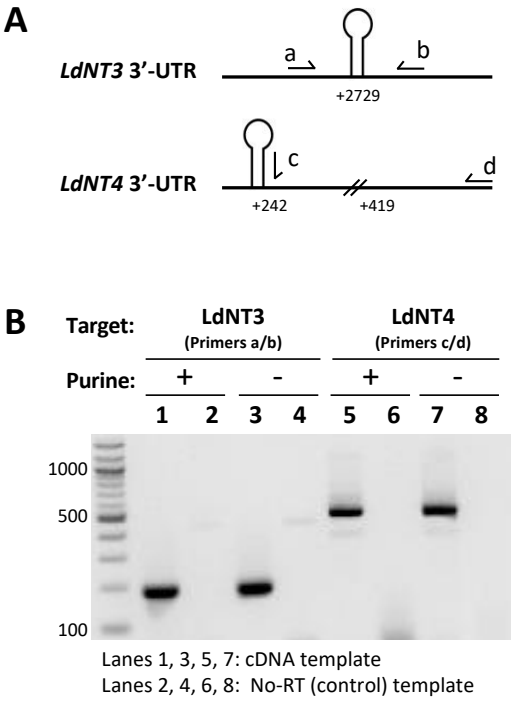


Figure 2.S2. The endogenous *LdNT3* stem-loop, as well as the sites of *LdNT3* stem-loop insertion in the *LdNT4* 3'-UTR, are present in their respective mRNAs regardless of purine availability. A) Schematic of PCR primers designed to detect regions of the *LdNT3* and *LdNT4* mRNA 3'-UTRs in mature messages. Top: Primers used to detect the *LdNT3* stem-loop in *LdNT3* mRNAs. Bottom: Primers used to i) detect the *LdNT3* stem-loop when inserted into the *LdNT4* 3'-UTR and ii) verify that position +419 of the *LdNT4* 3'-UTR is included in *Fluc* reporter mRNAs expressed from the *LdNT4* locus. Anticipated product sizes for PCR performed with primer sets a/b and c/d are 0.187 and 0.493 kb, respectively. Primer sequences are listed in Table A.7. B) PCR products generated using the primers shown in A. Parasites expressing *Fluc-BSD* flanked by *LdNT3* UTRs (lanes 1-4) or *LdNT4* UTRs harboring the *LdNT3* stem-loop at position +242 (lanes 5-8) were cultured in the presence or absence of purines for 24 hours. Total mRNA was isolated and reverse transcription was performed with oligo(dT) priming to generate first strand cDNAs from mature, poly(A)+ transcripts. Equal amounts of cDNA from each sample were used as template in subsequent PCR analyses. Products were resolved on a 2% gel with a 100 bp ladder (New England Biolabs). Lanes 1 and 3 show that the *LdNT3* stem-loop is present in *LdNT3* transcripts under both purine-starved and -replete conditions. Hence, it is unlikely that removal of the stem-loop via alternative polyadenylation accounts for the alleviation of repression under purine stress. Similarly, lanes 5 and 7 demonstrate that the *LdNT3* stem-loop inserted at position +242 in the *LdNT4* 3'-UTR is present in *Fluc* reporter mRNAs, irrespective of purine availability. Furthermore, the PCR product encompasses the +419 insertion site, indicating that failure of the *LdNT3* stem-loop to confer purine-responsiveness at this position is not due do its exclusion from the mRNA via polyadenylation site selection.

Chapter 3.

Distinct *cis*-acting elements govern purine-responsive regulation of the *Leishmania donovani* nucleoside transporters, NT1 and NT2

Introduction

Adapting to exogenous purine stress is an important aspect of *Leishmania* biology, yet the underlying mechanisms are ill-defined. As *Leishmania* do not control transcription of individual genes, they rely exclusively on post-transcriptional processes such as mRNA localization, translation and decay to enact changes in the proteome. At all of these stages, RNA-binding proteins (RBPs) serve as key factors. Characterizing the RBP-RNA interactions involved in the *Leishmania* purine stress response is therefore tantamount to understanding its regulation.

In *L. donovani*, purines are translocated across the plasma membrane by four nucleoside (LdNT1 and LdNT2) and nucleobase (LdNT3 and LdNT4) transporters [Landfear, 2004]. As early as 6 hours post-induction of purine stress, *L. donovani* promastigotes upregulate transport activity by increasing LdNT protein abundance [Carter, 2010; Martin, 2014]. This response is specific to LdNTs 1-3; LdNT4 expression is not sensitive to purines. In addition, the protein-level change is substantial, ranging from ~7.5 to 16-fold after 48 hours [Martin, 2014]. Thus, to investigate mechanisms of purine-responsive expression, we have used these genes as a model.

In previous work, we compared the RNA-centric processes affecting LdNT regulation [Martin, 2014; Soysa, 2014]. LdNTs 1-3 display robust translational enhancement under purine stress, despite a global reduction in protein synthesis [Martin, 2014; Shrivastava, 2019]. At the transcript level, both *LdNT1* and *LdNT3* are increased, whereas *LdNT2* is either unchanged or modestly reduced. Interestingly, though all four of the LdNTs are encoded by relatively high-copy messages, *LdNT2* is exceptional in its steady-state abundance, ranking in the 99.8th percentile in log-stage *L. donovani* promastigotes [Martin, 2014]. Together, these differences suggest that the purine transporters are regulated by independent, though possibly intersecting, post-transcriptional mechanisms.

In Chapter 2, I reported that the *LdNT3* mRNA 3'-UTR encodes a purine-responsive repressor element. In this chapter, I describe additional efforts to characterize the *cis*-acting sequences controlling LdNT1 and LdNT2. By systematic deletion mutagenesis, we identified a 76 nt-long polypyrimidine tract in the *LdNT2* mRNA 3'-UTR. Loss of this region led to a drastic reduction in transcript abundance and prevented translational enhancement under purine stress. Additionally, transcripts containing the *LdNT2* polypyrimidine tract localized to discrete cytoplasmic foci in purine-replete cells, suggesting that the high-copy *LdNT2* message is stored in RNA granules at steady-state. In the case of LdNT1, we found that purine-responsiveness is conferred by a polypyrimidine tract and additional upstream element, termed UE1. We established that both features are independently required for regulation, with the polypyrimidine tract and UE1 controlling mRNA abundance and translation, respectively. Finally, we found that

the *LdNT1* polypyrimidine tract can substitute for that of *LdNT2* to confer regulation in the context of the *LdNT2* 3'-UTR.

Results

Deletional mutagenesis of the LdNT2 UTRs reveals that purine-responsive expression is mediated by a 76 nt-long polypyrimidine tract

To better understand regulation of the nucleoside transporters in *Leishmania*, we began with a molecular dissection of *LdNT2*. The *LdNT2* 5'- and 3'-UTRs are together sufficient to confer purine-responsiveness to a luciferase reporter, strongly implicating the presence of distinct regulatory sites within one or both of these regions [Martin, 2014]. However, neither UTR was tested independently. As most *cis*-acting RNA elements have been identified in 3'-UTRs [Clayton, 2019], we suspected the *LdNT2* 5'-UTR is dispensable for purine-responsive expression. To test this, we used the approach depicted in 3.1A. A NanoLuciferase reporter construct (heretofore referred to as *LdNT2/NLuc*) was expressed from the endogenous *LdNT2* locus under the control of either wildtype *LdNT2* UTRs or a 5'-UTR from a purine-unresponsive control gene. For this purpose, we employed LdNT4, which is not differentially regulated with respect to purines [Martin, 2014]. The 5'-UTRs were substituted so as to preserve the dominant *LdNT2* 5' splice-acceptor site, located 269 nt upstream of translation start [TriTrypDB.org]. As we previously established that the *LdNT2* CDS is itself important for mRNA stability under purine stress [Martin, 2014], this sequence was included in the *LdNT2/NLuc* reporter construct to maintain endogenous-like expression. In parasites expressing *LdNT2/NLuc* flanked by wildtype UTRs, NLuc activity was robustly (~9.5-

Figure 3.1

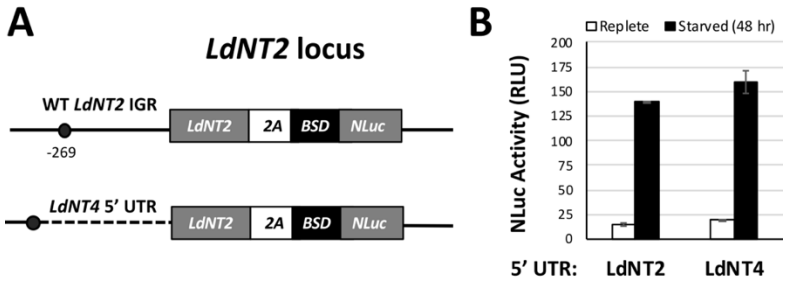


Figure 3.1: The *LdNT2* mRNA 5'-UTR is not required for robust purine-responsive regulation. A) Top: Schematic of multicistronic constructs (referred to as *LdNT2/NLuc*) used to study *cis*-regulation of the *LdNT2* message. *NLuc* was fused via its N-terminus to the blasticidin resistance gene (*BSD*) to enable mutant selection. The *LdNT2* CDS was included to maintain endogenous-like levels of expression, followed by the self-cleaving *Thosea asigna* virus 2A peptide (2A). For control cell lines, unmodified up- and downstream *LdNT2* intergenic regions (IGRs) were used to direct construct integration into the endogenous *LdNT2* locus. In this configuration, both stability and translation of the *NLuc* message are governed by the native *LdNT2* UTRs and/or CDS. Co-translational 2A cleavage separates the post-translational fates of LdNT2 from that of the downstream BSD-NLuc polypeptide [de Felipe, 2006]. Thus, NLuc activity does not reflect changes in LdNT2 protein stability, which could potentially mask any regulation conferred by the UTRs. Circle represents the dominant *LdNT2* splice acceptor site, 269 nt upstream of the translation start. Bottom: The *LdNT2* 5'-UTR was replaced with that of *LdNT4* to verify that this region is not required for regulation under purine stress. Solid and dashed lines indicate purine-responsive and -unresponsive mRNA UTRs, respectively. Not pictured: The second allelic copy of *LdNT2* was replaced with a phleomycin resistance gene (Phleo). A firefly luciferase-puromycin resistance gene fusion (*Fluc-PAC)* expressed from the *UMPS* locus was used as an internal control to normalize NLuc activity between replicates [Soysa, 2014]. B) Normalized NLuc activity from cell lines depicted in A, after 48 hours of culture in the presence or absence of purines. Figure shows the mean and standard deviation of experiments performed in biological duplicate.

fold) upregulated after 48 hours of purine starvation (Figure 3.1B). As anticipated, cell lines harboring the 5' sequence from *LdNT4* displayed an equivalent magnitude of NLuc induction, indicating that the *LdNT2* 5'-UTR is not required for purine-responsive control.

We next performed serial deletions to identify the *cis*-acting elements encoded within the *LdNT2* 3'-UTR. As determined by 3' Rapid Amplification of cDNA Ends (3' RACE), the preferred *LdNT2* 3' polyadenylation site lies 1.186 kb downstream of translation stop (Figure 3.2A). An *LdNT2/NLuc* reporter construct harboring wildtype *LdNT2* UTRs was therefore modified to tile this region with overlapping ~50-200 nt deletions and the effect on purine-responsive expression was examined. Deletions spanning the first 571 (Δ1 – Δ6) and last 511 (Δ7 – Δ11) bases had little to no effect on NLuc activity (Figure 3.2B). However, regulation was completely eliminated by deletion of 79 nts near the UTR midpoint. As this region was found to lie almost completely coincident with a 76 nt-long polypyrimidine tract, it is referred to as ΔCU.

Polypyrimidine tracts are ubiquitous in eukaryotic RNA. Through association with a variety of polypyrimidine tract binding proteins, including polypyrimidine tract binding protein 1 (PTB1), these features govern multiple stages of mRNA metabolism, including splicing, polyadenylation, nuclear export, and mRNA stability [reviewed in Romanelli, 2013]. The *T. brucei* PTB1 homolog, DRBD3, is an essential 37-kDa protein that binds mRNA at a conserved TTCCCCTCT motif [Das, 2015]. We observed that the *LdNT2* polypyrimidine tract encodes two overlapping, DRBD3-like binding sites (Figure 3.3A). Neither is perfectly identical to the published *T. brucei* DRBD3 consensus; however, in each, the identities of the divergent bases are among those tolerated for

Figure 3.2

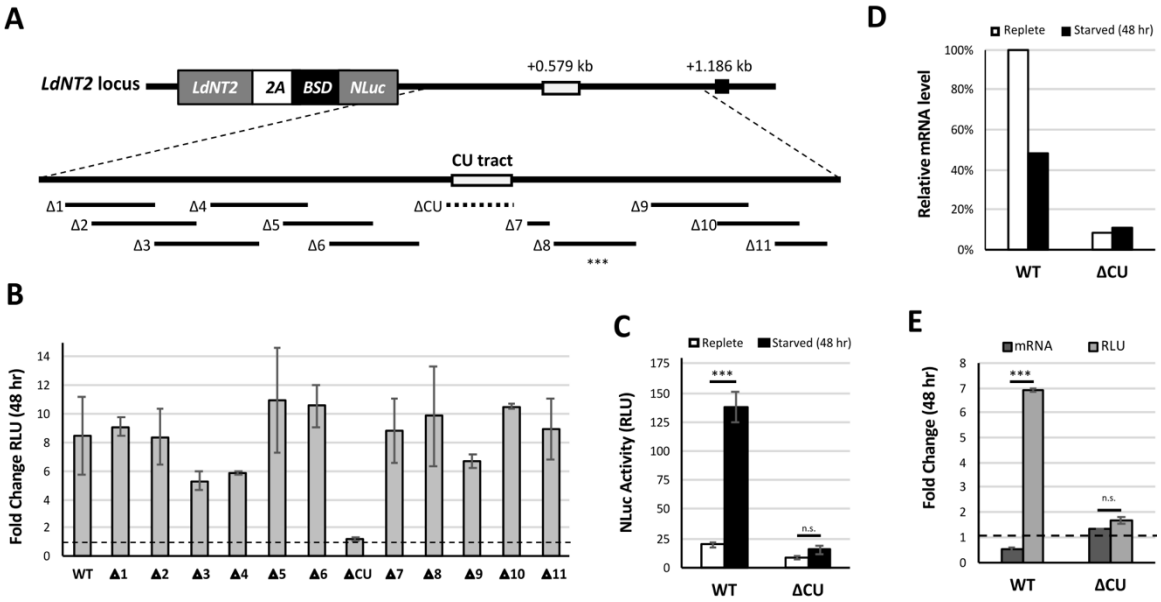


Figure 3.2: Deletional mutagenesis of the *LdNT2* 3'-UTR reveals that purine-responsive regulation is governed by a 76 nt-long polypyrimidine tract. A) Overlapping ~50-200 nt deletions (Δ1 - Δ11) generated in the 3'-UTR of the *LdNT2/NLuc* reporter construct, starting immediately downstream of translation stop and ending before the preferred polyadenylation site (black square). Dashed line highlights the region required for regulation, containing the *LdNT2* polypyrimidine (CU) tract. Numbers refer to distance between the indicated feature and translation stop. Not pictured: The second allelic copy of *LdNT2* was replaced with a phleomycin resistance gene (Phleo). A firefly luciferase-puromycin resistance gene fusion (*Fluc-PAC)* expressed from the *UMPS* locus was used as an internal control to normalize NLuc activity between replicates [Soysa, 2014]. B) Fold change in normalized NLuc activity after 48 hours of purine starvation, measured from cell lines depicted in A. WT refers to parasites expressing the *LdNT2/NLuc* construct under the control of native *LdNT2* UTRs. C-E) Investigating the contribution of the *LdNT2* polypyrimidine tract to mRNA stability and/or translation by paired dual-luciferase and RT-qPCR analyses. Where indicated, bars and error bars represent the mean and standard deviation of assays performed with 2 independent clones. C) Normalized NLuc activity measured after 48 hours of culture in the presence or absence of purines. D) Quantitation of *LdNT2/NLuc* transcripts from cultures assayed in C. Relative message level was determined by the comparative CT method using *UMPS* as an endogenous control gene and 'WT replete' as a reference sample. For ΔCT values and analysis details, refer to Table A.3. E) Fold change in *LdNT2/NLuc* mRNA level and NLuc activity after 48 hours of purine starvation. The comparative CT method was used to determine mRNA fold change for individual clones as shown in Table A.4. Single-factor ANOVA was calculated with Excel Descriptive Statistics Toolpak: ***$P \leq 0.001$; n.s., $P>0.05$.

Figure 3.3

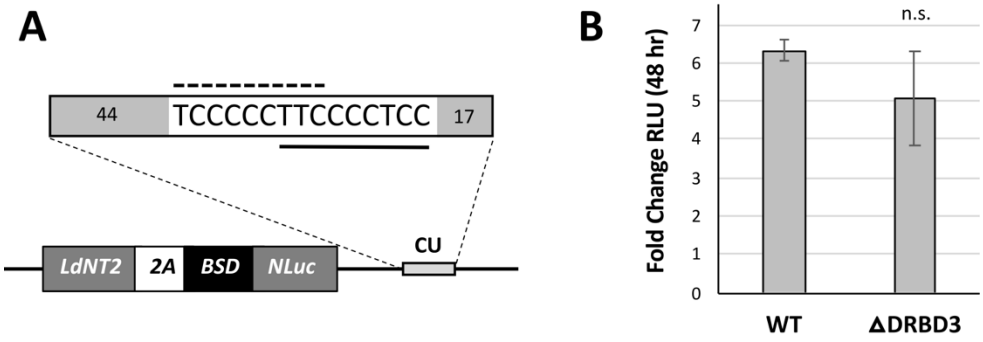


Figure 3.3: Binding sites for a known PTB protein in trypanosomes are encoded by the *LdNT2* polypyrimidine tract but are not required for regulation. A) Top: Sequence and relative position of two near-consensus DRBD3 binding sites in the *LdNT2* polypyrimidine tract (CU). Solid line indicates motif that differs from the published *T. brucei* DRBD3 consensus by a single residue; dashed line indicates a second, more degenerate site. In both cases, the divergent bases are tolerated for DRBD3 recognition [Das, 2015]. Numbers indicate the length of regions preceding and following putative DRBD3 motifs within the *LdNT2* polypyrimidine tract (76 nt in total). Diagram is not to scale. Bottom: *LdNT2/NLuc* reporter constructs were generated lacking putative DRBD3 motifs and integrated at the endogenous *LdNT2* locus. B) Regulation conferred by *LdNT2* mRNA UTRs, with and without DRBD3-like binding motifs. Data are normalized to *Fluc-PAC* expressed from the *UMPS* locus in the same cell line [Soysa, 2014]. Bars represent the mean and standard deviation of experiments performed in biological duplicate. Single-factor ANOVA was calculated with Excel Descriptive Statistics Toolpak. n.s., $P>0.05$.

recognition [Das, 2015]. In bloodstream-form trypanosomes, RNAi knockdown of DRBD3 destabilized many differentially expressed transcripts, including the *LdNT2* ortholog, *TbNT10* [Estevez, 2008; Stern, 2009]. These observations strongly suggested DRBD3 as a potential candidate for interaction with the *LdNT2* polypyrimidine tract. We generated *LdNT2/NLuc* constructs lacking both putative DRBD3 motifs (ΔDRBD3) to test whether these regions were specifically required for regulation by the *LdNT2* 3'-UTR. Surprisingly, after 48 hours of purine stress, ΔDRBD3 transgenic parasites displayed an equivalent magnitude of NLuc induction to those expressing the same construct flanked by WT *LdNT2* UTRs (Figure 3.3B). Thus, regulation conferred by the *LdNT2* polypyrimidine tract cannot be attributed to this particular region.

The polypyrimidine tract is a major determinant of LdNT2 transcript abundance and is required for translational enhancement under purine stress

LdNT2 is unique in that it falls in the top 99.8th percentile in mRNA abundance in log-stage *L. donovani* promastigotes [Martin, 2014]. At the same time, it ranks only in the 16th percentile in terms of ribosome occupancy [Bifeld, 2018]. It is also known that while translation of the *LdNT2* message is significantly upregulated in response to purine starvation, mRNA levels do not change [Martin, 2014]. Based on these data, we hypothesized that *LdNT2* translation is somehow restricted or repressed under purine-replete conditions. To test whether the *LdNT2* polypyrimidine tract affects either mRNA abundance or translation with respect to purines, total RNA was isolated from the cultures analyzed in Figure 3.2C and *LdNT2/NLuc* transcripts were quantified via RT-qPCR (Figure 3.2D). For either cell line, the impact of purine starvation on *LdNT2/NLuc* mRNA level was then compared against the corresponding fold-change in NLuc activity

to determine the translational contribution to regulation, reflected in the disparity between these two metrices (Figure 3.2E).

For parasites expressing *LdNT2/NLuc* flanked by wildtype UTRs, purine deprivation led to an approximate 50% decrease in transcript level. At the same time, NLuc activity was ~7-fold upregulated, suggesting that the *LdNT2* 3'-UTR and/or CDS together mediate a ~14-fold increase in translation under purine stress (Figure 3.2E, WT: dark vs light bars). This robust translational change is consistent with previous data for the endogenous *LdNT2* message [Martin, 2014]. Remarkably, *LdNT2/NLuc* mRNA abundance was reduced by 92% among parasites lacking the polypyrimidine tract (Figure 3.2D, white bars), implicating this sequence as an important determinant of message stability. At the protein level, the impact of polypyrimidine tract deletion was more modest, with ΔCU parasites demonstrating a 54% reduction in NLuc activity (Figure 3.2C, white bars). This disproportionate effect is in agreement with our hypothesis that a substantial portion of the *LdNT2/NLuc* mRNA pool is not translated under replete conditions. Furthermore, NLuc activity increased just ~1.7-fold in purine-starved ΔCU mutants, roughly equivalent to the change in mRNA abundance measured from the same cell line (Figure 3.2E, ΔCU: dark vs light bars). Thus, the translational component to regulation was completely eliminated by deletion of the *LdNT2* polypyrimidine tract. Taken together, the data support a model wherein the *LdNT2* message is poorly translated at steady-state but is maintained in high abundance by the polypyrimidine tract, such that it is available for translation under purine stress.

The LdNT2/NLuc transcript localizes to discrete cytoplasmic foci under purine-replete conditions

We next considered potential mechanisms that could account for both the exceptional abundance of *LdNT2* mRNA and its low translational efficiency. One possibility is that *LdNT2* messages are stored in ribonucleoprotein (RNP) granules under purine-replete conditions, where they are simultaneously protected from degradation and sequestered away from the translational machinery. In this scenario, purine starvation is expected to trigger release of sequestered messages, making them available for translation. In addition, this model implies that the polypyrimidine tract stabilizes *LdNT2* mRNA by contributing to its sequestration. To test these possibilities, the distribution of transcripts, with or without the polypyrimidine tract, was examined via RNA fluorescence in situ hybridization (RNA-FISH) under purine replete and depleted conditions. The 'WT' and ΔCU reporter lines described in Figure 3.2 were cultured for 48 hours in the presence or absence of purines and *LdNT2/NLuc* mRNAs were visualized using fluorescent probes specific to the *LdNT2* CDS.

In parasites expressing *LdNT2/NLuc* under the control of wildtype *LdNT2* UTRs (Figure 3.4A, WT), staining was enriched in discrete foci under purine-replete conditions. This observation supports our hypothesis that the transcript is sequestered in replete cells. Further, although a punctate RNA-FISH signal was also detected in these parasites after 48 hours of purine starvation, the fluorescence intensity per cell was ~32% lower (Figure 3.4D and G, blue plots). Considering that aggregated transcripts are readily detected by RNA-FISH but individual molecules are not, this reduction in signal in purine-starved parasites could possibly reflect a decrease in compartmentalization of the *LdNT2/NLuc* message, consistent with release from storage for translation. In the case of *LdNT2/NLuc*

Figure 3.4

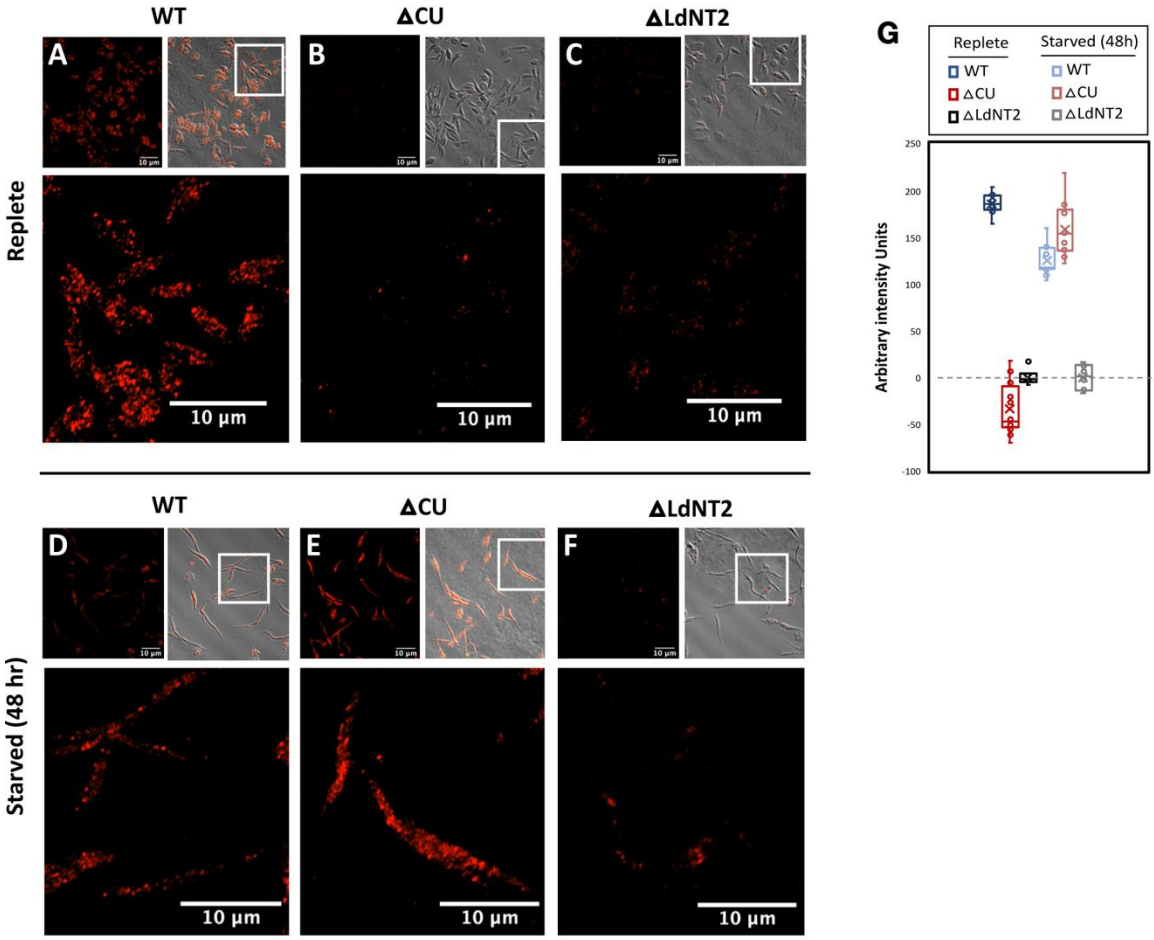


Figure 3.4: The *LdNT2* transcript localizes to discrete foci under purine-replete conditions. Promastigote parasites cultured for 48 hours in the presence (A-C) or absence (D-F) of purines were fixed on slides and processed for RNA-FISH. WT and ΔCU refer to the *LdNT2/NLuc* reporter lines of the same name depicted in Figure 3.2. *LdNT2/NLuc* transcripts were detected using fluorescently labeled probes specific for the *LdNT2* CDS. A homozygous *LdNT2* knockout line (ΔLdNT2) served as a control for nonspecific background. All fluorescent images were collected using the same settings; however, for qualitative assessment, representative images in A-C vs D-F were adjusted separately. [Author's note: the elongated morphology visible in panels D-F is characteristic of purine-starved *Leishmania.* This phenomenon is well-documented and discussed elsewhere]. G) Graph reflects the average fluorescent intensity per pixel per cell measured from several representative fields (n=12 in Wildtype and ΔCU, n=6 in ΔLdNT2). Under either condition, data were corrected by subtracting the average of the intensity values collected in the ΔLdNT2 control sample.

transcripts lacking the polypyrimidine tract (Figure 4.3E, ΔCU), staining was also localized to discrete foci under purine stress. We suspect that, in both WT and ΔCU lines, these puncta may represent *LdNT2/NLuc* messages aggregated in nutrient stress granules, the formation of which is well-documented in purine-starved *Leishmania* [Shrivastava, 2019]. Interestingly, however, the RNA-FISH signal did not exceed background levels in the replete ΔCU sample (Figure 3.4B and G; red plot). Thus, deletion of the *LdNT2* polypyrimidine tract appeared to disrupt compartmentalization of *LdNT2/NLuc* mRNA but specifically under purine-replete conditions. Further experiments will be required to more thoroughly assess transcript dynamics into and out of sequestration and to understand the nature of granules formed under purine-replete and starvation conditions (see Discussion).

Purine-responsive regulation of the LdNT1 nucleoside transporter requires cooperation between two distinct cis-acting elements

In *L. donovani,* NT1 is encoded by two tandem, closely related genes: *LdNT1.1* and *LdNT1.2.* Both are functional when injected into *Xenopus* oocytes; however, only *LdNT1.1* is expressed in promastigote parasites [Vasudevan, 1998]. For simplicity, all reference to *LdNT1* throughout pertains specifically to the gene expressed from the *LdNT1.1* locus.

By 3' RACE, we determined that the *LdNT1* message is polyadenylated at two positions, yielding 3'-UTRs of either 1.681 or 1.819 kb in length (Figure 3.5A). Approximately 1 kb downstream of translation stop, we noted a 48 nt polypyrimidine tract, reminiscent of the purine-response element described for *LdNT2*. We therefore performed a focused molecular dissection of the surrounding sequence to test if this

region also confers sensitivity to purines. A firefly luciferase-neomycin resistance gene fusion (*Fluc-NEO*) was expressed from the endogenous *LdNT1* locus under the control of either wildtype UTRs or a 3'-UTR harboring one of the ~50 nt deletions depicted in Figure 3.5A. Parasites were subjected to 48 hr of purine starvation and the impact on Fluc activity was evaluated. Luciferase induction was lost in all mutants lacking the *LdNT1* polypyrimidine tract (Δ6 - Δ7), consistent with the sequence functioning as a regulator of purine-responsive expression. However, several deletions preceding the polypyrimidine tract (Δ5, ΔUE1) also prevented regulation. To a resolution of 25 nt (i.e. the length of overlap between adjacent deletions), we determined that the 5' boundary of the *LdNT1* regulator lies 46 nt upstream of the polypyrimidine tract (position +0.910). The intervening sequence is heretofore referred to as the *LdNT1* upstream element, or UE1.

It is known that increases in both mRNA abundance and translational efficiency contribute to LdNT1 upregulation under purine stress [Martin, 2014]. We therefore asked if either of these processes are affected by the *LdNT1* polypyrimidine tract and/or UE1. The WT, ΔCU, and ΔUE1 reporter lines analyzed in Figure 3.5 (represented separately in Figure 3.6A for clarity) were cultured with and without purines for 48 hours. As described previously, the relative contributions of mRNA stability versus translation were then determined by comparing the purine-responsive change in *Fluc-NEO* mRNA abundance against the corresponding change in luciferase activity for each cell line. In the absence of purines, transgenic parasites harboring wildtype *LdNT1* UTRs demonstrated a 30% reduction in *Fluc-NEO* mRNA (Figure 3.6B, WT). This is in contrast to what has been reported for the endogenous *LdNT1* message, which is

Figure 3.5

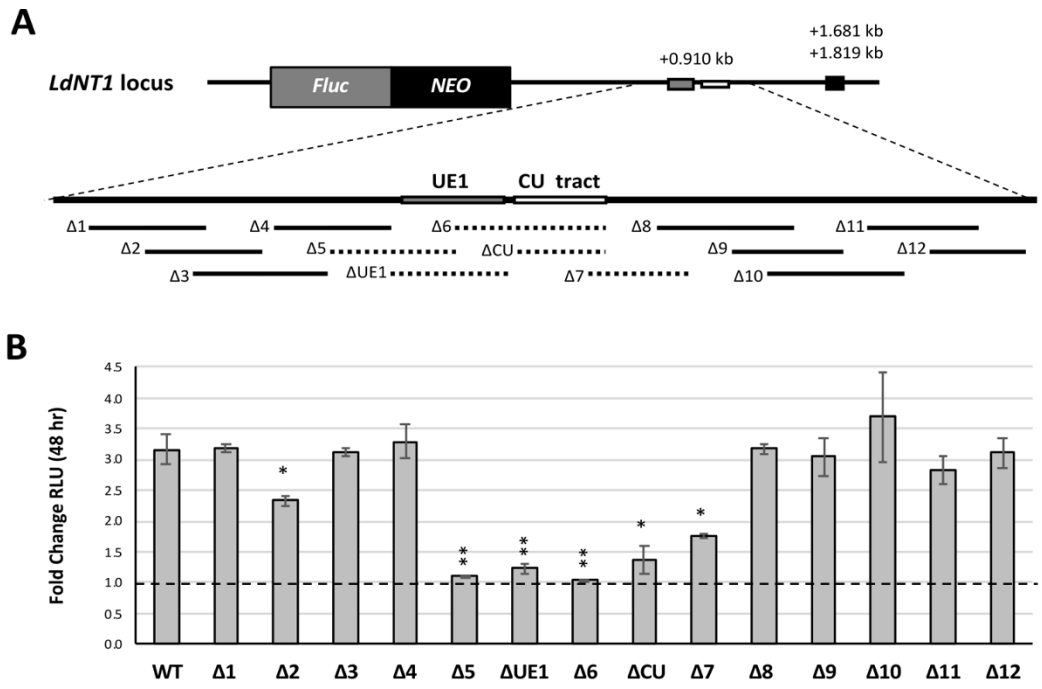


Figure 3.5: Deletional mutagenesis of the *LdNT1* 3'-UTR reveals a bipartite purine-response element. A) Firefly luciferase (*Fluc*) was fused in-frame to a selectable neomycin resistance marker (*NEO*) and flanked with the *LdNT1* IGRs to direct construct integration into the endogenous locus. Overlapping deletions (Δ1 - Δ12) encompass ~50 nt each, overlap by ~25 nt apiece, and span a total distance of 420 nt around the *LdNT1* polypyrimidine (CU) tract. Dashed lines indicate regions required for purine-responsive regulation, corresponding to the CU tract and upstream element (UE1). Black square represents the dominant *LdNT1* polyadenylation site(s) as determined by 3'-RACE. Numbers refer to distance between the indicated feature and translation stop. Not pictured: A *Renilla* luciferase-puromycin resistance gene fusion (*Rluc-PAC*) expressed from the *UMPS* locus serves as an internal normalization control. B) Fold change in normalized Fluc activity after 48 hours of purine starvation, measured from cell lines depicted in A. WT refers to parasites expressing *Fluc-NEO* under the control of native *LdNT1* UTRs. Bars represent the mean and standard deviation of experiments performed in biological duplicate. Single-factor ANOVA was calculated with Excel Descriptive Statistics Toolpak: $*P \leq 0.05$, $**P \leq 0.01$.

Figure 3.6

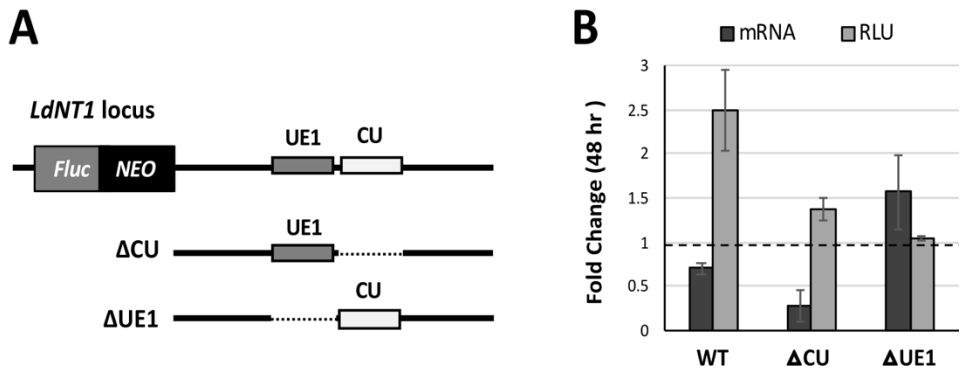


Figure 3.6: The *LdNT1* upstream element and polypyrimidine tract govern translation and mRNA stability independently. A) Investigating the contributions of the *LdNT1* upstream element (UE1) and polypyrimidine tract (CU) to mRNA stability and/or translation by paired dual-luciferase and RT-qPCR analyses. *Fluc-NEO* flanked by wildtype *LdNT1* UTRs serves as a control (referred to as WT in text). ΔCU and ΔUE1 cells lines are the same as those of the same name in Figure 3.5A but for ease of interpretation, their 3'-UTRs are represented again here. Not pictured: *Rluc-PAC* expressed from the *UMPS* locus was used to normalized between experiments. B) Fold change in *Fluc-NEO* mRNA level and Fluc activity after 48 hours of purine starvation, measured from the cell lines depicted in A. The mRNA fold change for individual clones was determined by the comparative CT method as shown in Table A.5. Bars represent the mean and standard deviation of experiments performed in biological duplicate.

modestly but significantly upregulated in purine-starved *L. donovani* promastigotes [Martin, 2014]. We interpret this to mean that the *LdNT1* CDS contributes to mRNA stability under purine starvation. Nonetheless, luciferase activity increased ~2.5-fold in the same cell line, pointing to a ~3.5-fold increase in translation mediated by the *LdNT1* UTRs (Figure 3.6B, WT: disparity between dark vs light bars). Interestingly, deletion of either the *LdNT1* polypyrimidine tract or UE1 had differing effects on abundance and translation of the reporter construct. In purine-starved parasites lacking the polypyrimidine tract, for instance, *Fluc-NEO* transcripts were even more substantially decreased than in the WT reporter line (73% vs 30%, respectively), suggesting that this region confers stability to the *LdNT1* message under purine stress. Yet these cells still displayed a robust 4.8-fold increase in translation (Figure 3.6B, ΔCU: dark vs light bars). In contrast, deletion of UE1 completely eliminated the translational contribution to regulation but did not negatively impact mRNA level (Figure 3.6B, ΔUE1). Taken together, these data suggest that the *LdNT1* polypyrimidine tract and UE1 function independently of each other to confer regulation at the levels of mRNA abundance and translation, respectively.

The LdNT1 polypyrimidine tract confers regulation in the context of the LdNT2 mRNA 3'-UTR, *in the absence of UE1*

For both *LdNT1* and *LdNT2*, purine-responsive expression is governed by a polypyrimidine tract in the mRNA 3'-UTR. However, in the latter case, regulation also requires cooperation with an adjacent translational enhancer (i.e. UE1). In a final experiment, we asked whether just the polypyrimidine tract from *LdNT1* could substitute for that of *LdNT2* to confer regulation outside of its endogenous genetic context. We

Figure 3.7

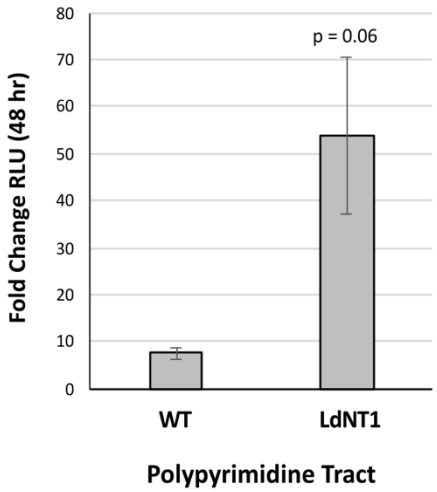


Figure 3.7: The *LdNT1* polypyrimidine tract confers purine-responsive regulation when substituted at the *LdNT2* locus. In the *LdNT2/NLuc* construct (depicted in Figure 3.1A, top), the *LdNT2* polypyrimidine tract was replaced with that of *LdNT1.* WT refers to *LdNT2/NLuc* flanked by wildtype *LdNT2* 5'- and 3'-UTRs. Graph displays fold change in NLuc activity after 48 hours of purine starvation. NLuc activity is normalized to Fluc expressed from the endogenous UMPS locus. Bars represent the mean and standard deviation of experiments performed in biological duplicate.

swapped these respective elements in the *LdNT2/NLuc* reporter construct and evaluated the impact on NLuc expression after 48 hours of purine stress. As depicted in Figure 3.7, *LdNT2/NLuc* flanked by wildtype *LdNT2* UTRs (WT) displayed a ~7.7-fold increase in luciferase activity. Remarkably, reporter lines harboring a polypyrimidine tract from *LdNT1* not only retained purine-responsive expression, but the magnitude of induction was ~7-fold greater than regulation conferred by the endogenous sequence (Figure 3.7, LdNT1). This suggests that, at least in the particular genetic context of the *LdNT2/NLuc* construct, the *LdNT1* polypyrimidine tract is sufficient to confer purine-sensitivity independently of UE1.

Discussion

We have used the purine transporters as a model to examine regulation of the *Leishmania* purine stress response. Previous studies suggested that these genes are governed by different post-transcriptional mechanisms and we recently described a repressive stem-loop in the 3'-UTR of *LdNT3* that serves to limit expression under purine-replete conditions [Licon, 2020]. Though conserved in orthologous genes from a variety of kinetoplastids, the *LdNT3* stem-loop was not found elsewhere in the *L. donovani* genome, including the other LdNTs. Thus, in the present work, we performed systematic deletion analysis of the *LdNT1* and *LdNT2* UTRs to identify the elements responsible for their control.

LdNT2 mRNA is exceptionally abundant yet poorly translated in purine-replete *L. donovani* promastigotes. We found that regulation depends on 76 nt-long polypyrimidine tract, encoded in the mRNA 3'-UTR. In the context of a reporter construct, we showed

that mRNA abundance is ~92% reduced by deletion of the *LdNT2* polypyrimidine tract (Figure 3.2D). Translational enhancement under purine stress was also eliminated (Figure 3.2E). Based on these observations, we suggest a model for *LdNT2* regulation wherein i) in replete cells, *LdNT2* messages are simultaneously stabilized and sequestered away from the translational machinery by storage in RNP granules, ii) translation is upregulated when stored transcripts are made accessible by trafficking out of granules in response to purine stress, and iii) *LdNT2* mRNA stability and sequestration are dependent on the polypyrimidine tract. Although the results of our RNA-FISH analysis are not entirely conclusive, they provide compelling preliminary evidence in support of this model.

Under purine-replete conditions, we found that transcripts harboring wildtype *LdNT2* UTRs localize to discrete, cytoplasmic foci (Figure 3.4A). This observation is generally in line with our model. However, it begs the question: what is the nature of the RNA-containing structures? Many types of RNP granules have been described in kinetoplastid parasites, with distinct protein markers identified. Broadly speaking, these are distinguished based on whether they form under favorable or stress conditions, and whether transcripts are stored or degraded within [reviewed Kramer, 2014]. For example, both heat shock and nutrient restriction induce formation of cytoplasmic stress granules, which act to store and protect mRNAs until stress is resolved [Kramer, 2008; Fritz, 2015]. Specialized sites of mRNA turnover known as processing bodies (P-bodies) have also been observed. As in other eukaryotes, P-bodies are constitutively present in kinetoplastids but increase in size and abundance in response to environmental insults [Holetz, 2007; Kramer, 2010; Fritz, 2015]. Recently, a paralog of

eukaryotic initiation factor eIF4E was found to concentrate in cytoplasmic granules in purine-starved *Leishmania* promastigotes. These structures also contained ribosomal subunits and mature mRNAs, suggestive of a role in translational repression [Shrivastava, 2019]. We suspect that the fluorescent puncta observed in purine-starved *LdNT2/NLuc* parasites (Figure 3.4D and E) represent one or more of these established granule types. However, to our knowledge, there have been no examples of translational repression by selective mRNA sequestration described in any kinetoplastid parasite to date, making the phenomenon we have observed in purine-replete cells potentially novel. In future studies, techniques such as fluorescence-activated particle sorting (FAPS) may be employed to more thoroughly characterize the RNA and protein composition of the *LdNT2*-containing granules [Hubstenberger, 2017].

In the same 'WT' cell line, intensity of the RNA-FISH signal was significantly reduced after exposure to purine stress (Figure 3.4D and G). As aggregated RNAs are more readily detected by RNA-FISH than are individual molecules, this could reflect a general shift in the *LdNT2/NLuc* mRNA pool from a sequestered to a free state. Such a phenomenon is consistent with our hypothesis that *LdNT2* transcripts are trafficked out of storage granules to be translated in response to purine starvation. However, because the overall *LdNT2/NLuc* message level is also reduced under stress (Figure 3.2D, WT), we are cautious in accepting this interpretation. As more conclusive evidence, single molecule RNA-FISH (smRNA-FISH) could provide insight into the dynamics of individual transcripts, into and/out of sequestration. It could also potentially be informative to compare the distribution of *LdNT2* to other messages for which purine-responsive localization has already been determined. *HSP83*, for example, yields a

diffuse cytoplasmic signal in replete *Leishmania* and accumulates in storage granules under purine restriction [Shrivastava, 2019]. In any case, further examination is required to verify the significance of this observation.

Our preferred model states that storage of the *LdNT2* message in RNP granules physically protects it from the degradation machinery in the cytosol. In this scenario, message stability is dependent upon sequestration. Working backward from our observation that the *LdNT2* polypyrimidine tract is a major regulator of mRNA abundance (Figure 3.2D), one would therefore predict that transcripts lacking this element are not recruited to granules under purine-replete conditions. Alternatively, it is also possible that *LdNT2* mRNA abundance is controlled independently of sequestration, with the polypyrimidine tract only influencing the former. In this case, *LdNT2* transcripts would be expected to accumulate in cytoplasmic foci in replete cells, irrespective of the polypyrimidine tract. In purine-replete ΔCU *LdNT2/NLuc* parasites, we did not readily detect fluorescent puncta and the RNA-FISH signal did not exceed background levels, as measured from the ΔLdNT2 control (Figures 3.4B and G). One interpretation of these data is that ΔCU *LdNT2/NLuc* transcripts are not sequestered in granules, in line with our preferred model of regulation. However, because abundance of the *LdNT2/NLuc* message is also dramatically reduced in this cell line (~92% lower than WT; Figure 3.2D), it is equally possible that an absence of signal in these samples merely reflects a transcript that has dropped below the limit of detection. As evidence in support of the former case, fluorescent puncta were clearly visible in ΔCU parasites exposed to purine stress (Figure 3.4E), despite there being no significant difference at the mRNA level (Figure 3.2D, ΔCU). This observation would seem to suggest that *LdNT2/NLuc* mRNAs in the ΔCU

cell line are not so low abundance as to be undetectable when aggregated. Thus, although the results of this experiment are not conclusive, we are encouraged that they genuinely reflect an inability of purine-replete *Leishmania* to compartmentalize transcripts lacking the *LdNT2* polypyrimidine tract. In future studies, smRNA-FISH will be conducted to distinguish between the two options outlined above.

As an important caveat, it should be noted that all of the observations described here are based on RNA-FISH experiments performed with a non-native RNA construct. *LdNT2/NLuc* encodes all components of the mature *LdNT2* message (i.e. 5' and 3' UTRs, CDS). In addition, several preliminary RNA-FISH experiments were conducted to detect the endogenous *LdNT2* transcript. Qualitatively, the subcellular distribution of this message appeared consistent with what we have reported for the 'WT' *LdNT2/NLuc* construct [data not shown]. However, we cannot exclude the possibility that aspects of regulation are affected. In future experiments, we will also verify the localizations described here with endogenous *LdNT2* mRNA.

We identified a second purine-responsive polypyrimidine tract in the *LdNT1* 3'-UTR. Like that of *LdNT2*, loss of this region had a negative impact on mRNA abundance. However, we found that translation of the *LdNT1* message is separately controlled by an adjacent sequence, termed UE1 (Figure 3.6). Both features are required for induction under purine stress, suggesting that they function cooperatively to enact changes in protein abundance. We were therefore surprised to find that, when substituted into the *LdNT2/NLuc* reporter construct, the polypyrimidine tract from *LdNT1*was sufficient for regulation, independent of UE1 (Figure 3.7). This observation speaks to the importance of local genetic context in *cis*-regulation, the subtleties of which are often overlooked.

Indeed, the effect of substitution robust, with the magnitude of reporter induction far exceeding that conferred by the wildtype *LdNT2* 3'-UTR under purine stress. The reason for this is unclear, particularly without knowing the impact of the substitution at the mRNA level. Does the *LdNT1* polypyrimidine tract confer greater message stability in this context, resulting in a larger pool of *LdNT2/NLuc* for translation under stress? By that same token, do *LdNT2/NLuc* messages harboring the *LdNT1* polypyrimidine tract also localize to cytoplasmic foci under replete conditions? The answers to these and other questions could provide further mechanistic insight into the respective functions of the two polypyrimidine tracts controlling nucleoside transport.

In combination with our previous work on LdNT3, we can now definitively state that each *L. donovani* purine transporter is controlled by distinct purine-response elements. However, the obvious question remains: What binds to these regions *in vivo?* We indirectly tested one candidate for *LdNT2* in the form of DRBD3. Despite strong evidence suggesting that this protein binds to and stabilizes transcripts encoding the *LdNT2* ortholog in *T. brucei* [Das, 2015; Estevez, 2008; Stern, 2009], deletion of two near-consensus DRBD3-binding sites in the *LdNT2* polypyrimidine tract had no significant effect on regulation. Admittedly, based purely on these results of this experiment, we cannot exclude the possibility of an interaction between *L. donovani* DRBD3 and *LdNT2*, occurring elsewhere in the *LdNT2* message or at the deleted sites. Further, as we did not examine the impact of this deletion at the transcript level, we cannot speculate as to a potential role for DRBD3 in regulating *LdNT2* mRNA abundance. We can merely state that the 15 nts deleted are not required for *LdNT2* induction in response to purines. In any case, given the mechanistic differences in

regulation conferred by each of the *LdNT* purine-response elements (i.e. positive versus negative control, mRNA stability versus translation), we suspect that the protein factors involved differ between the transporters.

Chapter 4.

Summary, conclusions, and future directions.

Summary and conclusions

In my dissertation work, I sought to expand our understanding of purine stress tolerance in *Leishmania donovani* by examining regulation of several representative, purine-responsive genes. To that end, I focused on the membrane nucleoside and nucleobase transporters, LdNTs 1-3. Using a series of luciferase reporter constructs, I established that the LdNTs are each controlled by one or more distinct *cis*-acting elements, encoded within the mRNA 3'-UTR. In order of their discovery, these are:

- A 33 nt predicted stem-loop in *LdNT3,* which operates at the levels of both mRNA abundance and translation to repress expression under purine-replete conditions.
- A 76 nt-long polypyrimidine tract in *LdNT2,* which serves to maintain *LdNT2* mRNA in a highly abundant and translationally repressed state under purine-replete conditions.
- A 48 nt-long polypyrimidine tract and adjacent 46 nt-long upstream element (UE1) in *LdNT1,* which function at the levels of mRNA abundance and translation, respectively, to promote expression under purine stress.

To our knowledge, the purine-response elements described in this thesis are the first to be identified in *Leishmania.* Their characterization highlights a remarkable degree of complexity in the regulation of the *Leishmania* purine stress response and sets the stage for future investigations to identify the larger network of RNA-protein and protein-protein interactions involved.

Complexity of regulation in the Leishmania purine stress response

Coordinating the expression of functionally related groups of genes allows cells to rapidly and efficiently respond to changes in the environment. In bacteria, this is accomplished by physically clustering genes into co-transcribed units, or operons. While mRNA synthesis undoubtably also contributes to regulation in nucleated cells (kinetoplastids excluded), the transcriptome does not always accurately predict protein level. To account for this disparity, in 2002, Keene and Tenenbaum proposed an amended model of eukaryotic gene expression wherein, following transcription, nascent mRNAs are further organized through interaction with *trans*-acting RNA-binding proteins (RBPs) [Keene, 2002]. These so-called RNA regulons are distinguished by the presence of specific RBP-recognition motifs and have since been implicated in a wide range of cellular processes, including growth, differentiation, and stress tolerance [reviewed in Keene, 2007]. In the absence of transcriptional control, RNA regulons are predicted to play a particularly important role in the coordination of kinetoplastid gene expression [Ouellette, 2009; Trenaman, 2019].

The LdNTs perform complementary and analogous roles in purine salvage. As part of the *L. donovani* purine stress response, LdNTs 1-3 are upregulated on a parallel

timescale (i.e. within the first 6 hours) post-induction of starvation [Martin, 2014]. Based on these observations alone, one might expect these genes to be subject to co-regulation by common *cis*-and *trans*-acting factors. However, sequence-level differences in the RNA-elements described here suggest otherwise. Indeed, the unique combinations of post-transcriptional mechanisms affecting each transporter (i.e. mRNA stability versus translation, positive versus negative control) indicate that they are controlled by different regulatory pathways entirely.

This complexity in purine-responsive regulation appears to extend well beyond the purine transporters in *Leishmania.* In previous studies, our group demonstrated that extracellular purines are sensed, not at the cell surface, but through surveillance of the intracellular adenylate and guanylate nucleotide pools [Martin, 2016]. Specifically, we reported that perturbation of the adenylate pool provides the signal required to withstand long-term purine deprivation. Parasites starved for adenylate nucleotides demonstrated changes in the purine salvage pathway similar to cells cultured in the complete absence of an exogenous purine source. Interestingly, starvation for guanylate nucleotides also provoked upregulation of several key purine salvage enzymes; however, the effect was not consistent across all genes examined. For instance, whereas regulation of the LdNTs was equivalent regardless of whether cells were starved for guanylate or adenylate nucleotides, a membrane 3'-nucleotidase/nuclease, which was robustly upregulated under adenylate nucleotide starvation, was completely unaffected by guanylate restriction. These differential effects suggest that, while only adenylate stress triggers the transition to a long-term persistence phenotype, components of the purine salvage pathway are regulated downstream of two separate purine-sensors/signaling pathways.

These observations beg the question: is there an evolutionary advantage to separating regulation of the purine salvage pathway? The LdNTs differ in substrate specificity. Are there circumstances under which *Leishmania* parasites specifically regulate just one or a subset of transporters, based on the balance of available purine metabolites in the host? The fact that these organisms do not directly detect individual purine species in the extracellular milieu is inconsistent with this idea. Perhaps a more likely explanation is that regulation of the LdNTs reflects an instance of convergent evolution, wherein individual genes operating within a common pathway have arrived at a similar pattern of expression because there was a selective advantage in doing so. Whatever the case, it should be noted that the data described in this thesis are not altogether inconsistent with a regulon-based model for orchestration of the purine stress response. Although the *LdNTs* are themselves not co-regulated, each may still fall within a different, purine-responsive cohort of coordinately expressed transcripts (i.e. a regulon). Indeed, our observation that the *LdNT3* stem-loop-binding protein is present in gross abundance (much more than strictly required to control *LdNT3* alone) could reflect a role in the regulation of other genes. Future efforts to identify proteins that bind to the purine-response elements identified in this thesis and define their full complement of RNA targets will be informative in this regard.

Regulation by granule sequestration

I have suggested a model for LdNT2 regulation wherein the highly abundant *LdNT2* transcript is translationally repressed under purine-replete conditions by storage in cytoplasmic RNA granules. Translational enhancement is then achieved by mRNA

release from granules under purine stress. Although the results of our RNA-FISH analysis are not conclusive, they provide compelling preliminary evidence in support of this model.

In terms of efficiency, this means of regulation is uniquely suited to kinetoplastid biology. In the nuclei of kinetoplastid parasites, the entire genome is constitutively transcribed such that the rate of synthesis of a given mRNA does not correlate with its steady-state abundance or translation. Compared to those of yeast and mammals, trypanosome mRNAs are relatively unstable, displaying a median half-life of ~13 minutes [Manful, 2011]. This intimates that the majority of transcripts are degraded shortly after synthesis. Thus, for a single-copy gene such as *LdNT2*, the energetic cost to amassing large quantities of mRNA is minimal, as it avoids the ATP-intensive degradation process. Moreover, having a reserve of mRNA available for translation at the onset of stress allows cells to respond quickly, without the need for new mRNA synthesis. This is particularly advantageous in context of purine stress, which triggers a global reduction in transcription initiation [Martin, 2014]. It is therefore all the more intriguing that, to our knowledge, there are no other examples of regulation by mRNA sequestration described in kinetoplastids. Whether this phenomenon is genuinely uncommon or simply understudied in these organisms is not clear. However, there are a number of other predicted nutrient transporters in *Leishmania* with high mRNA abundance and poor translational efficiency [Yates, P., unpublished observation] that, in light of our observations on *LdNT2*, could be interesting to investigate in future studies.

While our work has illustrated that *LdNT2* transcripts are compartmentalized under purine-replete conditions, further efforts are required to determine the nature of the

LdNT2-containing granules. In this regard, some clues may be taken from the distantly related apicomplexan parasite, *Plasmodium*. In vector-stage *Plasmodium*, it was reported that a cohort of transcripts, termed upregulated in infectious sporozoites (UIS), are simultaneously stabilized and translationally repressed by association with the Pumilio domain-containing protein, Puf2 [Gomes-Santos, 2011]. Though UIS transcripts were not detected directly, Puf2 localized to cytoplasmic RNP granules coincident with UIS translational repression. Puf2-containing granules did not colocalize with either exonuclease XRN1 or eIF2α, canonical markers of P-bodies are stress granules, respectively, leading to their designation as storage granules [Lindner, 2013]. I suspect that the *LdNT2*-containing granules described in this thesis may play a similar role in *Leishmania.* In future studies, it will be interesting to characterize the RNA and protein composition of the *LdNT2*-containing granules described here.

Future directions: Moving beyond mRNA

The work presented in this thesis highlights the utility of RNA-centric approaches for dissecting complex regulatory networks in kinetoplastids. Using simple genetic techniques, I have gleaned information pertaining to the mechanism, binding specificity, and/or abundance of several RBPs involved in regulation of the *Leishmania* purine stress response, all without knowledge of the RBPs themselves. Although the scope of my graduate work was limited to the purine transporters, investigations focused on other purine-responsive genes are likely to yield similar insight.

One particularly interesting result from these analyses centers on evolutionary conservation of the *LdNT3* repressor stem-loop. Specifically, I found that the orthologous

stem-loops from *T. brucei* and *L. donovani* are not functionally interchangeable for regulation across species, suggesting that the proteins that recognize these elements in either organism have diverged in binding specificity. This difference could potentially be exploited in future studies to identify the proteins in question. For instance, one could imagine a cross-species complementation screen in which a library of *T. brucei* open reading frames is heterologously expressed in *Leishmania.* In the same cell line, a negative selectable marker is then expressed under the control of the repressive *T. brucei* stem-loop. Only binding of the cognate *T. brucei* protein to its RNA target would repress marker expression and enable cell survival.

Regardless of the approach taken, the ultimate goal for the three *cis*-acting elements described here is to characterize their RBP partners *in vivo.* Once candidates have been identified, there will be a variety of questions to pursue. Immunoprecipitation of candidate RBPs will reveal the network of other RNA-protein and protein-protein interactions in which they are involved. In addition, it will be important to understand how the associations, localizations, and post-translational modifications of these proteins are affected by purine stress. Given the substantial evidence implicating purine starvation as a trigger for *Leishmania* metacyclogenesis, it will also be interesting to test the contributions of identified RBPs on differentiation and survival in the sandfly and mammalian hosts.

Chapter 5.

Materials and methods.

Leishmania donovani culture

All cell lines described herein were generated from the *L. donovani* 1S-2D clonal subline LdBob, originally obtained from Dr. Stephen Beverley [Goyard, 2003]. LdBob promastigotes were routinely maintained at 26 °C in 5% CO_2 and cultured in Dulbecco's Modified Eagle-Leishmania (DME-L) medium supplemented with 5% SerumPlus™ (SAFC BioSciences/Sigma Aldrich, St. Louis, MO; a purine-free alternative to standard FBS), 1mM L-glutamine, 1x RPMI vitamin mix, 10uM folic acid, 50 ug/ml hemin, and 100 uM hypoxanthine as a purine source. For general culture maintenance, blasticidin, puromycin, and phleomycin were used at 30 ug/ml, 25 ug/ml, and 50 ug/ml, respectively. To elicit purine starvation, logarithmically growing cells were pelleted via centrifugation (5000 x g for 5 min), washed once in DME-L lacking hypoxanthine but containing all other media supplements, and resuspended at a density of 2 x 10^6 cells/ml in either purine-replete or purine-free medium.

Luciferase reporter constructs and cloning

Basic construction

With the exception of plasmid A.1F (see below), all gene targeting vectors were generated using the multi-fragment ligation approach described in [Fulwiler, 2011] and

depicted in Figure A.1. For multicistronic constructs, 2A-containing transgenes were provided by donor plasmids pCRm-coBSD-2A-Fluc (Figures A.1B and A.1C) and pCRm-2A-coBSD-NLuc (Figures A.1E and A.1F) [Yates, P., manuscript in preparation]. Selectable *Fluc-DRG* fusions (Figure A.1D and A.G) were donated by pCRm-luc2-BSD and pCRm-luc2-NEO (Genbank Accession numbers KF035118.1 and KF035120.1, respectively). Targeting sequences used to direct integration were PCR amplified from genomic DNA with Phusion High Fidelity DNA polymerase (New England Biosciences, Ipswitch, MA) using the primers listed in Table A.8. All vector components were digested with SfiI (or AlwNI, where indicated), gel-purified, and assembled in a single ligation step.

Additional modifications: Chapter 2

To generate *LdNT3* stem-loop-deficient mutants, the *LdNT3*-targeting construct shown in Figure A.1D was modified with the QuikChange Site-directed Mutagenesis Kit (Stratagene, Lajolla, CA) using primers listed in Table A.9. To insert the *LdNT3* stem-loop and variations thereof into the *LdNT4* 3'-UTR, the *LdNT4* version of construct A.1D was subjected to whole-plasmid PCR amplification via Phusion polymerase using the primers listed in Table A.10. PCR products were DpnI-treated to eliminate template plasmid and circularized via the Gibson Assembly method using NEBuilder HiFi DNA Assembly Master Mix (New England Biosciences, Ipswitch, MA).

All primers used to modify pRP vectors are listed in Table A.11. To integrate the firefly luciferase gene into the rRNA array, *Fluc* was amplified from pCRm-luc2-BSD and cloned into the SfiI sites of pRP-L_A and pRP-VH [Soysa, 2015]. Subsequent

insertion of the *LdNT3* stem-loop into pRP-L_A and pRP-VH was accomplished in two stages. First, the vectors were subjected to whole-plasmid amplification to introduce two nonidentical BstXI sites into their shared 3'-UTRs (see primer sequences for detail). Second, a version of the *LdNT3* stem-loop was synthesized with flanking BstXI and PCR primer binding sites (Genscript, Piscataway, NJ) and inserted into the BstXI sites of the modified pRP vector.

Additional modifications: Chapter 3

The *LdNT2* 5'-UTR replacement construct depicted in Figure A.1F was assembled in a step-wise fashion. First, genomic DNA was isolated from transgenic parasites expressing *LdNT2/NLuc* flanked by endogenous *LdNT2* IGRs. Using the primers listed in Table A.12, the entire *LdNT2/NLuc* reporter locus (including the preceding 5'-IGR) was PCR amplified and cloned into the SwaI restriction sites of pBB. The resultant vector (not pictured) was then subjected to whole plasmid amplification to exclude the 269 nt immediately upstream of translation start (i.e. the *LdNT2* 5'-UTR). A length of 250 nt immediately preceding *LdNT4* translation start was amplified from genomic DNA and both components were assembled via Gibson Assembly method. For deletional mutagenesis of the *LdNT2* 3'-UTR, the primers listed in Table A.13 were used to perform whole plasmid amplification of vector A.1E, with each divergent primer pair excluding a ~50-200 nt-long region. PCR products were circularized via NotI restriction sites encoded in primer 3' ends. A similar approach was taken to delete putative DRBD3 binding motifs from the *LdNT2* polypyrimidine tract, using the primers listed in Table A.14. However, in this case, circularization was achieved by Gibson Assembly.

To introduce deletions into the *LdNT1* 3'-UTR, the downstream IGR was PCR amplified as two separate halves, with each separated by the intended deletion site. Individual fragments were assembled with a *Fluc-NEO* fusion transgene using the multi-fragment ligation scheme shown in Figure A.1G. Primers are listed in Table A.15

To substitute the *LdNT1* polypyrimidine tract for that of *LdNT2* in *LdNT2/NLuc* reporter constructs, XbaI restriction sites were introduced in place of the *LdNT2* polypyrimidine tract via whole plasmid amplification. The *LdNT1* polypyrimidine tract was then amplified from genomic DNA and inserted via directional XbaI cloning. All relevant primers are listed in Table A.16

Transfections

In all luciferase assays described throughout work, a compatible luciferase expressed under the control of purine-unresponsive UTRs from UMP synthase (UMPS) served as a control from normalization. Hence, all Fluc- and NLuc-based reporter constructs were transfected into LdBob derivatives expressing either *Rluc-PAC* or *Fluc-PAC,* respectively, from the endogenous *UMPS* locus [Soysa, 2014]. Recipients of *LdNT2/NLuc* constructs also harbored a heterozygous *LdNT2* deletion (*UMPS/ump*s*::Fluc-PAC; LdNT2/ldnt2::Phleo)*, such that, in the resultant cell lines, the only expressed copy of *LdNT2* was encoded by the reporter locus. For deletion mutagenesis of the *LdNT1* 3'-UTR, the *Fluc-NEO* reporter constructs depicted in Figure A.1G were delivered to a recipient line expressing *Fluc-BSD* from the endogenous *LdNT1* locus (*UMPS/ump*s*::Rluc-PAC; LdNT1/ump*s*::Fluc-BSD*) such that integration was directed by the *Fluc* reporter gene.

Transfections of mid-log stage promastigotes were performed with ~3ug SwaI-linearized plasmid DNA using the high-voltage electroporation protocol described by Robinson and Beverley [Robinson, 2003]. Immediately following electroporation, cells were transferred into 5 mL of complete DME-L and 200 ul was added to the first column of wells on a 96 well plate and subjected to 2-fold serial dilution to derive independent clones. Transfections were incubated overnight at 26 °C in 5% CO_2. Selection was initiated the following day by adding 100 ul of 2X blasticidin (60 ug/ml) or, for integration of *NEO*-containing pRP-L_A/VH and *LdNT1.1* deletion vectors, G418 (50 ug/ml) to each well. Proper integration of the constructs (and, where applicable, inclusion of deletions) was verified via PCR for all clones.

Dual-Luciferase analysis

Dual-luciferase assays were performed using either the Dual-Glo Luciferase Assay System (Fluc and Rluc) or the Nano-Glo Dual-Luciferase Reporter Assay (Nluc and Fluc) from Promega (Madison, WI). Analyses were performed using 35 ul of cell culture in white polystyrene 96-well half-area plates (Corning, Amsterdam) as described in the respective product technical manuals. For each incubation step, plates were protected from light and shaken for 10 minutes at room temperature (RT) on an orbital shaker. Luminescence was measured using a Veritas Microplate Luminometer (Turner BioSystems, Sunnyvale, CA).

RNA, cDNA, and RT-qPCR analysis

Total cellular RNA was isolated from 5 x 10^6 log-stage parasites using the Qiagen RNeasy Plus Micro Kit following the protocol for animal and human cells. Cell lysates were disrupted using a QIAshredder spin column. To eliminate contaminating genomic DNA, RNA samples were subjected to DNaseI digestion using the TURBO DNA-*free* kit (ThermoFisher). First-strand cDNA synthesis was subsequently performed using one of several reverse transcriptase systems. To generate cDNA for downstream application as template in RT-PCR analyses (Figures 2.S2), reverse transcription was performed with Invitrogen SuperScript III and oligo(dT) priming. Complimentary RNA was removed from cDNA products with RNase H prior to PCR (see Table A.6 for PCR primers). For RT-qPCR analyses, cDNA was generated using Promega ImProm-II reverse transcriptase with random hexamer priming (Figure 2.2) or the Applied Biosystems High Capacity cDNA Reverse Transcription Kit (Figures 2.7, 3.2, 3.6). The resultant cDNA samples were diluted so as to reduce the total input RNA to 6 ng per ul and dye-based RT-qPCR was performed with NEB Luna Universal qPCR Master Mix using 2 ul (12 ng) of diluted cDNA. Previously validated PCR primers [Martin, 2014] are listed in Table A.7. Reactions were run on an Applied Biosystems StepOnePlus instrument using the "Fast" ramp speed and the following thermocycling parameters: 95 °C for 60 seconds; 40 cycles of denaturing at 95 °C for 15 seconds followed by a 30 second extension at 60 °C. A final melt curve step was included to verify the specificity of amplification. The relative abundance of the *Fluc* (or *LdNT2,* in Figure 3.2) message expressed from various genetic loci was determined using the comparative CT (ΔΔCT) method as described [Martin, 2014].

RNA-FISH

To generate probes for RNA-FISH, the *LdNT2* CDS was amplified from genomic DNA using primers listed in TableA.8 and purified using the NEB Monarch PCR and DNA Cleanup kit. DNA probes were labeled for 4 hr via nick translation (Vysis; Abbott Laboratories, Abbott Park, Illinois) with Spectrum Orange-dUTP (Vysis) according to manufacturer instructions. Labeled probes were suspended in SLI/WCP hybridization buffer (Vysis) to a final concentration of 16.6 ng/ul.

For analysis of *L. donovani* promastigotes, parasites were pelleted by centrifugation and resuspended at a density of ~$2x10^7$ cell/ml in 3.7% formaldehyde. To minimize clumping, cell pellets were gently disrupted by running the collection tube 5x along a microtube rack prior to fixation. Fixed cells (500 ul) were distributed over polylysine-coated slides and allowed to settle for ~20 minutes at RT before the cell suspension was aspirated from the slide surface. Slides were then stored in 70% EtOH at -20ºC until use. For hybridization, labeled probes were denatured at 75 ºC for 10 min. Just prior to probe application, slides were dehydrated in an EtOH series (3 min each in 90% and 100% EtOH) and allowed to air dry at RT. Slides were then hybridized with 20 ul of denatured probes in a humid chamber for 14-16 hours at 37ºC. As an additional precaution against desiccation during hybridization, coverslip-mounted slides were also sealed with rubber cement. Post-hybridization washes consisted of i) one 3-minute wash in 2xSSC/50% formamide at 37ºC and ii) one 1-minute wash in 2xSSC/0.1% Triton X-100 at RT. Coverslips were then mounted with Prolong Gold DAPI antifade and imaged on a ZEISS LSM 980 in Airyscan SR mode with a 63x1.4 NA objective.

For individual images, average fluorescence intensity per cell was quantitated as the signal intensity per pixel scaled to the number of parasites per field (counted by DAPI-stained kinetoplasts). To account for differences in the background signal under either purine-treatment condition, the average intensity measured in the ΔLdNT2 control was subtracted from measurements collected under the same conditions in experimental cell lines. Thus, data are represented as delta values.

REFERENCES

Abdeladhim, M., Kamhawi, S., & Valenzuela, J. G. (2014). What's behind a sand fly bite? The profound effect of sand fly saliva on host hemostasis, inflammation and immunity. *Infect Genet Evol, 28*, 691–703.

Abhishek, K., Sardar, A. H., Das, S., Kumar, A., Ghosh, A. K., Singh, R., … Das, P. (2017). Phosphorylation of translation initiation factor 2-Alpha in *Leishmania donovani* under stress is necessary for parasite survival. *Mol Cell Biol*, *37*(1), e00344-16.

Afroz, T., Cienikova, Z., Cléry, A., & Allain, F. H. T. (2015). One, two, three, four! How multiple RRMs read the genome sequence. *Method Enzymol*, *558*, 235–278.

Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P., & Sereno, D. (2016). A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglect Trop D, 10*(3), e0004349.

Alemayehu, B., & Alemayehu, M. (2017). Leishmaniasis: A review on parasite, vector and reservoir host. *Health Sci J, 11*(4), 519.

Alvar, J., Vélez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., … de Boer, M. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE*, *7*(5), e35671.

Alves, L. R., Oliveira, C., Mörking, P. A., Kessler, R. L., Martins, S. T., Romagnoli, B. A. A., Marchini, F. K., & Goldenberg, S. (2014). The mRNAs associated to a zinc finger protein from Trypanosoma cruzi shift during stress conditions. *RNA Biol*, *11*(7), 921–933.

Anderson, B. A., Wong, I. L. K., Baugh, L., Ramasamy, G., Myler, P. J., & Beverley, S. M. (2013). Kinetoplastid-specific histone variant functions are conserved in Leishmania major. *Mol Biochem Parasitol*, *191*(2), 53–57.

Antoine, J., Prina, E., Jouanne, C., Bongrand, P., & Cedex, P. (1990). Parasitophorous vacuoles of *Leishmania amazonensis*-infected macrophages maintain an acidic pH. *Infect Immun, 58*(3), 779-787.

Antoine, J.-C., Prina, E., Lang, T., & Courret, N. (1998). The biogenesis and properties of the parasitophorous vacuoles that harbour *Leishmania* in murine macrophages. *Trends Microbiol, 7*(10), 392–401.

Arenas, R., Torres-Guerrero, E., Quintanilla-Cedillo, M. R., & Ruiz-Esmenjaud, J. (2017). Leishmaniasis: A review. *F1000Research*, 6, 750.

Azizi, H., Dumas, C., & Papadopoulou, B. (2017). The Pumilio-domain protein PUF6 contributes to SIDER2 retroposon-mediated mRNA decay in *Leishmania*. *RNA, 23*, 1874-1885.

Badis, G., Saveanu, C., Fromont-Racine, M., & Jacquier, A. (2004). Targeted mRNA degradation by deadenylation-independent decapping. *Mol Cell, 15*, 5–15.

Bai, Y., Salvadore, C., Chiang, Y.-C., Collart, M. A., Liu, H.-Y., & Denis, C. L. (1999). The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5. *Mol Cell Biol, 19*(10), 6642-6651.

Bangs, J. D., Crainll, P. F., Hashizumell, T., Mccloskeyllii, J. A., & Boothroyd, J. C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem, 267*(14), 9805–9815.

Bates, P.A., Tetley, L. (1993). *Leishmania mexicana*: induction of metacyclogenesis by cultivation of promastigotes at acidic pH. *Exp Parasitol, 76*(4), 412–423.

Bates, P. A. (2008). *Leishmania* sand fly interaction: progress and challenges. *Curr Opin Microbiol, 11*, 340-344.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., … El-Sayed, N. M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science, 309*, 416–422.

Besteiro, S., Williams, R. A. M., Morrison, L. S., Coombs, G. H., & Mottram, J. C. (2006). Endosome sorting and autophagy are essential for differentiation and virulence of *Leishmania major*. *J Biol Chem, 281*(16), 11384–11396.

Bifeld, E., Lorenzen, S., Bartsch, K., Vasquez, J.-J., Siegel, T. N., & Clos, J. (2018). Ribosome profiling reveals HSP90 inhibitor effects on stage-specific protein synthesis in *Leishmania donovani*. *MSystems, 3*(6).

Boeck, R., Tarun, S., Rieger, M., Deardorff, J. A., Mü ller-Auer, S., & Sachs, A. B. (1996). The yeast Pan2 protein Is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity, *J Biol Chem, 271*(1), 432-438.

Boitz, J. M., Ullman, B., Jardim, A., & Carter, N. S. (2012). Purine salvage in *Leishmania*: Complex or simple by design? *Trends Parasitol, 28*(8), 345-352.

Brengues, M., Teixeira, D., & Parker, R. (2005). Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science, 310*, 486–489.

Bringaud, F., Müller, M., Cerqueira, G. C., Smith, M., Rochette, A., El-Sayed, N. M. A., … Ghedin, E. (2007). Members of a large retroposon family are determinants of

post-transcriptional gene expression in *Leishmania*. *PLoS Pathog*, *3*(9), 1291–1307.

Britto, C., Ravel, C., Bastien, P., Blaineau, C., Pages, M., Dedet, J., & Wincker, P. (1998). Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene, 222*, 107–117.

Buchan, J. R., Parker, R., & Ross, J. (2009). Eukaryotic stress granules: The ins and out of translation. *Mol Cell*, *36*(6), 932.

Burza, S., Croft, S. L., & Boelaert, M. (2018). Leishmaniasis. *The Lancet, 392,* 951-970.

Camacho, E., González-de la Fuente, S., Rastrojo, A., Peiró-Pastor, R., Solana, J. C., Tabera, L., … Aguado, B. (2019). Complete assembly of the *Leishmania donovani* (HU3 strain) genome and transcriptome annotation. *Sci Rep*, *9*, 6127.

Carter, N. S., Yates, P. A., Gessford, S. K., Galagan, S. R., Landfear, S. M., & Ullman, B. (2010). Adaptive responses to purine starvation in *Leishmania donovani*. *Mol Microbiol*, *78*(1), 92–107.

Chakravarty, J., & Sundar, S. (2010). Drug resistance in leishmaniasis. *J Glob Infect Dis, 2*(2), 167–176.

Chappuis, F., Sundar, S., Hailu, A., Ghalib, H., Rijal, S., Peeling, R. W., … Boelaert, M. (2007). Visceral leishmaniasis: What are the needs for diagnosis, treatment and control. *Nat Rev Microbiol, 5,* 873-882.

Chicoine, J., Benoit, P., Gamberi, C., Paliouras, M., Simonelig, M., & Lasko, P. (2007). Bicaudal-C recruits CCR4-NOT deadenylase to target mRNAs and regulates oogenesis, cytoskeletal organization, and its own expression. *Dev Cell*, *13*(5), 691–704.

Clayton, C. (2012). 'mRNA turnover in trypanosomes' in *RNA Metabolism in Trypanosomes*. Springer Berlin Heidelberg: Springer. 79-97.

Clayton, C. (2016). Gene expression in kinetoplastids. *Curr Opin Microbiol, 32,* 46-51.

Clayton, C., & Shapira, M. (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol,156,* 93-101.

Clayton, C. (2019). Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol., 9*(6).

Collart, M. A., & Struhl, K. (1993). CDC39, an essential nuclear protein that negatively regulates transcription and differentially affects the constitutive and inducible HIS3 promoters. *EMBO J, 12*(1), 177-186.

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., … Vannini, A. (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*, *37*(1), 337–352.

Cunningham, M. L., Richard, G. T., Salvatore, J. T., & Beverley, S. M. (2001). Regulation of differentiation to the infective stage of the protozoan parasite *Leishmania major* by tetrahydrobiopterin. *Science*, *292*(5515), 285–287.

da Costa Lima, T. D., Moura, D. M. N., Reis, C. R. S., Vasconcelos, J. R. C., Ellis, L., Carrington, M., … de Melo Neto, O. P. (2010). Functional characterization of three *Leishmania* poly(A) binding protein homologues with distinct binding properties to RNA and protein partners. *Eukaryot Cell*, *9*(10), 1484–1494.

da Costa, K. S., Galúcio, J. M. P., Leonardo, E. S., Cardoso, G., Leal, É., Conde, G., & Lameira, J. (2017). Structural and evolutionary analyses of *Leishmania* Alba proteins. *Mol Biochem Parasitol, 217*, 23–31.

Das, A., Bellofatto, V., Rosenfeld, J., Carrington, M., Romero-Zaliz, R., Del Val, C., & Estévez, A. M. (2015). High throughput sequencing analysis of *Trypanosoma brucei* DRBD3/PTB1-bound mRNAs. *Mol Biochem Parasitol*, *199*(1–2), 1–4.

de Felipe, P., Luke, G. A., Hughes, L. E., Gani, D., Halpin, C., & Ryan, M. D. (2006). E unum pluribus: Multiple proteins from a self-processing polyprotein. *Trends in Biotechnol, 24*, 68-75.

de Gaudenzi, J., Frasch, A. C., & Clayton, C. (2005). RNA-binding domain proteins in kinetoplastids: A comparative analysis. *Eukaryot Cell*, *4*(12), 2106–2114.

Debrabant, A., Joshi, M. B., Pimenta, P. F. P., & Dwyer, D. M. (2004). Generation of *Leishmania donovani* axenic amastigotes: Their growth and biological characteristics. *Int J Parasitol*, *34*(2), 205–217.

Dhalia, R., Reis, C. R. S., Freire, E. R., Rocha, P. O., Katz, R., Muniz, J. R. C., … De Melo Neto, O. P. (2005). Translation initiation in *Leishmania major*: Characterisation of multiple eIF4F subunit homologues. *Mol Biochem Parasitol*, *140*(1), 23–41.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J., Taly, J., & Notredame, C. (2011) T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res, 39,* W13-W17.

Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., van Nostrand, E. L., Pratt, G. A., Yeo, G. W., Graveley, B. R., & Burge, C. B. (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Mol Cell, 70*, 854-867.

Dostálová, A., Käser, S., Cristodero, M., & Schimanski, B. (2013). The nuclear mRNA export receptor Mex67-Mtr2 of *Trypanosoma brucei* contains a unique and essential zinc finger motif. *Mol Microbiol 88*(4), 728–739.

Dostálová, A., & Volf, P. (2012). *Leishmania* development in sand flies: parasite-vector interactions overview. *Parasite Vector, 5*(276).

Droll, D., Minia, I., Fadda, A., Singh, A., Stewart, M., Queiroz, R., & Clayton, C. (2013). Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. *PLoS Path*, *9*(4), e1003286.

Dupé, A., Dumas, C., & Papadopoulou, B. (2014). An Alba-domain protein contributes to the stage-regulated stability of amastin transcripts in *Leishmania*. *Mol Microbiol*, *91*(3), 548–561.

Ekanayake, D. K., Minning, T., Weatherly, B., Gunasekera, K., Nilsson, D., Tarleton, R., … Sabatini, R. (2011). Epigenetic regulation of transcription and virulence in *Trypanosoma cruzi* by O-linked thymine glucosylation of DNA. *Mol Cell Biol*, *31*(8), 1690–1700.

Ekanayake, D. K., Cipriano, M. J., & Sabatini, R. (2007). Telomeric co-localization of the modified base J and contingency genes in the protozoan parasite *Trypanosoma cruzi. Nucleic Acids Res*, *35*(19), 6367–6377.

El-Sayed, N. M., Myler, P. J., & Blandin, G. (2005). Comparative genomics of trypanosomatid parasitic protozoa. *Science*, *309*(5733), 404–409.

Estévez, A. M. (2008). The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res*, *36*(14), 4573–4586.

Fadda, A., Färber, V., Droll, D., & Clayton, C. (2013). The roles of 3′-exoribonucleases and the exosome in trypanosome mRNA degradation. *RNA*, *19*(7), 937–947.

Fernández-Moya, S. M., Carrington, M., & Estévez, A. M. (2014). A short RNA stem-loop is necessary and sufficient for repression of gene expression during early logarithmic phase in trypanosomes. *Nucleic Acids Res, 42*, 7201–7209.

Fiebig, M., Kelly, S., & Gluenz, E. (2015). Comparative life cycle transcriptomics revises *Leishmania mexicana* genome annotation and links a chromosome duplication with parasitism of vertebrates. *PLoS Pathog*, *11*(10), e1005186.

Freire, E. R., Sturm, N. R., Campbell, D. A., & de Melo Neto, O. P. (2017). The role of cytoplasmic mRNA cap-binding protein complexes in *Trypanosoma brucei* and other trypanosomatids. *Pathogens, 6*(55).

Freire, E. R., Vashisht, A. A., Malvezzi, A. M., Zuberek, J., Langousis, G., Saada, E. A., … Campbell, D. A. (2014). eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in *Trypanosoma brucei*. *RNA, 20*(8), 1272–1286.

Fritz, M., Vanselow, J., Sauer, N., Lamer, S., Goos, C., Siegel, T. N., … Kramer, S. (2015). Novel insights into RNP granules by employing the trypanosome's microtubule skeleton as a molecular sieve. *Nucleic Acids Res, 43*(16), 8013–8032.

Fulwiler, A. L., Soysa, D. R., Ullman, B., & Yates, P. A. (2011). A rapid, efficient and economical method for generating leishmanial gene targeting constructs. *Mol and Biochem Parasitol, 175*(2), 209–212.

Galindo, I., & Ramirez Ochoa, J. L. (1989). Study of *Leishmania mexicana* electrokaryotype by clamped homogeneous electric field electrophoresis. *Mol Biochem Parasitol, 34,* 245-252.

Garneau, N. L., Wilusz, J., & Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nat Rev Mol, 8,* 113-126.

Gebauer, F., & Hentze, M. W. (2004). Molecular mechanisms of translational control. *Nat Rev Mol, 5,* 827-835.

Ghorbani, M., & Farhoudi, R. (2018). Leishmaniasis in humans: Drug or vaccine therapy? *Drug Des Devel Ther, 12,* 25-40.

Gilinger, G., & Bellofatto, V. (2001). Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res, 29*(7), 1556-1564.

Goldstrohm, A. C., Hook, B. A., Seay, D. J., & Wickens, M. (2006). PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol, 13*(6), 533–539.

Gomes-Santos, C. S. S., Braks, J., Prudêncio, M., Carret, C., Gomes, A. R., Pain, A., Feltwell, T., Khan, S., Waters, A., Janse, C., Mair, G. R., & Mota, M. M. (2011). Transition of *Plasmodium* sporozoites into liver stage-like forms is regulated by the RNA binding protein Pumilio. *PLoS Pathog, 7*(5).

Goos, C., Dejung, M., Wehman, A. M., M-Natus, E., Schmidt, J., Sunter, J., … Kramer, S. (2019). Trypanosomes can initiate nuclear export co-transcriptionally. *Nucleic Acids Res, 47*(1), 266–282.

Gossage, S. M., Rogers, M. E., & Bates, P. A. (2003). Two separate growth phases during the development of *Leishmania* in sand flies: Implications for understanding the life cycle. *Int J Parasitol*, *33*(10), 1027–1034.

Goyard, S., Segawa, H., Gordon, J., Showalter, M., Duncan, R., Turco, S. J., & Beverley, S. M. (2003). An *in vitro* system for developmental and genetic studies of *Leishmania donovani* phosphoglycans. *Mol Biochem Parasitol. 130*, 31–42.

Gruenheid, S., Pinner, E., Desjardins, M., & Gros, P. (1997). Natural resistance to infection with intracellular pathogens: The Nramp1 protein Is recruited to the membrane of the phagosome. *J Exp Med*, *185*(4), 717–730.

Haldar, A. K., Sen, P., & Roy, S. (2011). Use of antimony in the treatment of leishmaniasis: Current status and future directions. *Mol Biol Int*, 1–23.

Hendriks, E. F., & Matthews, K. R. (2005). Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. *Molecular Microbiology*, *57*(3), 706–716.

Hendriks, E. F., Robinson, D. R., Hinkins, M., & Matthews, K. R. (2001). A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *EMBO J*, *20*(23), 6700–6711.

Holetz, F. B., Correa, A., Ávila, A. R., Nakamura, C. V., Krieger, M. A., & Goldenberg, S. (2007). Evidence of P-body-like structures in *Trypanosoma cruzi*. *Biochem Biophys Res Co*, *356*(4), 1062–1067.

Holzer, T. R., McMaster, W. R., & Forney, J. D. (2006). Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes, and axenic amastigotes in *Leishmania mexicana*. *Mol and Biochem Parasitol*, *146*(2), 198–218.

Hook, B. A., Goldstrohm, A. C., Seay, D. J., & Wickens, M. (2007). Two yeast PUF proteins negatively regulate a single mRNA. *J Biol Chem*, *282*(21), 15430–15438.

Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J. B., Munier, A., Fradet, M., Daunesse, M., Bertrand, E., Pierron, G., Mozziconacci, J., Kress, M., & Weil, D. (2017). P-Body purification reveals the condensation of repressed mRNA regulons. *Mol Cell*, *68*(1), 144-157.

Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., & Wright, P. E. (2004). Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat Struct Mol Biol*, *11*(3), 257–264.

Huynh, C., Sacks, D. L., & Andrews, N. W. (2006). A *Leishmania amazonensis* ZIP family iron transporter is essential for parasite replication within macrophage phagolysosomes. *J Exp Med*, *203*(10), 2363–2375.

Ilg, T., Stierhof, Y.-D., Craik, D., Simpson, R., Handman, E., & Bacic, A. (1996). Purification and structural characterization of a filamentous, mucin-like proteophosphoglycan secreted by *Leishmania* parasites. *J Biol Chem*, *271*(35), 21583-21596.

Inbar, E., Hughitt, V. K., Dillon, L. A. L., Ghosh, K., El-Sayed, N. M., & Sacks, D. L. (2017). The transcriptome of *Leishmania major* developmental stages in their natural sand fly vector. *mBio*, *8*(2), e00029-17.

Ivens, A. C., Lewis, S. M., Bagherzadeh, A., Zhang, L., Chan, H. M., & Smith, D. F. (1998). A Physical Map of the *Leishmania major* Friedlin Genome. *Genome Resh*, *8*, 135–145.

Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.-A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchettin, G., Borzym, K., … Myler, P. J. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, *309*(5733), 436-442.

Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K., & Krzyzosiak, W. J. (2016). Identifying proteins that bind to specific RNAs - Focus on simple repeat expansion diseases. *Nucleic Acids Res*, *44*(19), 9050–9070.

Jin, C., & Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev*, *21*(12), 1519–1529.

Jolma, A., Zhang, J., Mondragón, E., Kivioja, T., Yin, Y., Morris, Q., Hughes, T. R., James Maher III, L., & Taipale, J. (2020). Binding specificities of human RNA binding proteins towards structured and linear 1 RNA sequences. *Genome Res*, *30*, 1-12.

Kamenska, A., Simpson, C., & Standart, N. (2014). EIF4E-binding proteins: New factors, new locations, new roles. *Biochem Soc Trans*, *42*, 1238–1245.

Kamhawi, S., Ramalho-Ortigao, M., Pham, V. M., Kumar, S., Lawyer, P. G., Turco, S. J., … Valenzuela, J. G. (2004). A role for insect galectins in parasite survival. *Cell*, *119*, 329-341.

Keene, J. D., & Tenenbaum, S. A. (2002). Eukaryotic mRNPs May represent posttranscriptional operons. *Mol Cell*, *9*(6), 1161–1167.

Keene, J. D. (2007). RNA regulons: Coordination of post-transcriptional events. *Nature Reviews Genetics*, 8(7), 533–543.

Killick-Kendrick, R., Molyneux, D. H., & Ashford, R. W. (1974). Ultrastructural observations on the attachment of *Leishmania* in the sandfly. *Trans R Soc Trip Med Hyg*, *68*(4), 269.

Killick-Kendrick, R., & Rioux, J. A. (2002). Mark-release-recapture of sand flies fed on leishmanial dogs: the natural life-cycle of *Leishmania infantum* in *Phlebotomus ariasi*. *Parassitologia*, *44*, 67–71.

Kim, J. H., & Richter, J. D. (2006). Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Mol Cell*, *24*(2), 173–183.

Kim, D. I., Birendra, K. C., Zhu, W., Motamedchaboki, K., Doye, V., & Roux, K. J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *PNAS, 111*(24).

Köhler, A., & Hurt, E. (2007, October). Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol, 8*, 761-773.

Kolev, N. G., Franklin, J. B., Carmi, S., Shi, H., Michaeli, S., & Tschudi, C. (2010). The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog*, *6*(9), e1001090.

Kolev, N. G., Ramey-Butler, K., Cross, G. A. M., Ullu, E., & Tschudi, C. (2012). Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science, 338*(6112), 1352–1353.

Kramer, S., Queiroz, R., Ellis, L., Webb, H., Hoheisel, J. D., Clayton, C., & Carrington, M. (2008). Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2α phosphorylation at Thr169. *J Cell Sci*, *121*(18), 3002–3014.

Kramer, S. (2012). Developmental regulation of gene expression in the absence of transcriptional control: The case of kinetoplastids. *Mol Biocheml Parasitol, 181*, 61-72.

Kramer, S. (2014). RNA in development: How ribonucleoprotein granules regulate the life cycles of pathogenic protozoa. *Wiley Interdiscip Rev: RNA*, *5*(2), 263–284.

Kramer, S. (2017). The ApaH-like phosphatase TbALPH1 is the major mRNA decapping enzyme of trypanosomes. *PLoS Pathog*, *13*(6), e1006456.

Kramer, S., Bannerman-Chukualim, B., Ellis, L., Boulden, E. A., Kelly, S., Field, M. C., & Carrington, M. (2013). Differential localization of the two *T. brucei* poly(A)

binding proteins to the nucleus and RNP granules suggests binding to distinct mRNA pools. *PLoS ONE*, *8*(1), e54004.

Kramer, S., Kimblin, N. C., & Carrington, M. (2010). Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*. *BMC Genomics*, *11*(1), 283.

Lahav, T., Sivam, D., Volpin, H., Ronen, M., Tsigankov, P., Green, A., … Myler, P. J. (2011). Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. *FASEB J*, *25*(2), 515–525.

Lamphear, B. J., Kirchweger, R., Skern, T., & Rhoads, R. E. (1995). Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases: Implications for cap-dependent and cap-independent translational initiation. *J Biol Chem*, *270*(37), 21975–21983.

Landfear, S. M., Ullman, B., Carter, N. S., & Sanchez, M. A. (2004). Nucleoside and nucleobase transporters in parasitic protozoa. *Eukaryot Cell*, *3*(2), 245-254.

Laufs, H., Müller, K., Fleischer, J., Reiling, N., Jahnke, N., Jensenius, J. C., … Laskay, T. (2002). Intracellular survival of *Leishmania major* in neutrophil granulocytes after uptake in the absence of heat-labile serum factors. *Infect Immun*, *70*(2), 826–835.

LeBowitz, J. H., Smith, H. Q., Rusche, L., & Beverley, S. M. (1993). Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev*, *7*(6), 996–1007.

Li, C. H., Irmer, H., Gudjonsdottir-Planck, D., Freese, S., Salm, H., Haile, S., … Clayton, C. (2006). Roles of a *Trypanosoma brucei* 5′/3′ exoribonuclease homolog in mRNA degradation. *RNA*, *12*(12), 2171–2186.

Liang, X. H., Haritan, A., Uliel, S., & Michaeli, S. (2003). *Trans*- and *cis*-splicing in trypanosomatids: Mechanism, factors, and regulation. *Eukaryot Cell*, *2*(5), 830-840.

Licon, M. H., & Yates, P. A. (2020). Purine-responsive expression of the *Leishmania donovani* NT3 purine nucleobase transporter is mediated by a conserved RNA stem-loop. *J Biol Chem*, *295*(25), 8449–8459.

Lighthall, G. K., & Giannini, S. H. (1992). The Chromosomes of *Leishmania*. *Parasitol Today*, *8*(6), 192-199.

Lin, C., & Miles, W. O. (2019). Survey and summary beyond ClIp: Advances and opportunities to measure RBP-RNA and RNA-RNA interactions. *Nucleic Acids Res*, *47*(11), 5490–5501.

Lindner, S. E., Mikolajczak, S. A., Vaughan, A. M., Moon, W., Joyce, B. R., Sullivan, W. J., & Kappe, S. H. I. (2013). Perturbations of *Plasmodium* Puf2 expression and RNA-seq of Puf2-deficient sporozoites reveal a critical role in maintaining RNA homeostasis and parasite transmissibility. *Cell Microbiol*, *15*(7), 1266–1283.

Lodge, R., & Descoteaux, A. (2006). Phagocytosis of *Leishmania donovani* amastigotes is Rac1 dependent and occurs in the absence of NADPH oxidase activation. *Eur J Immunol 36*(10), 2735–2744.

Lodge, R., Diallo, T. O., & Descoteaux, A. (2006). *Leishmania donovani* lipophosphoglycan blocks NADPH oxidase assembly at the phagosome membrane. *Cell Microbiol 8*(12), 1922–1931.

Lowell, J. E., Rudner, D. Z., & Sachs, A. B. (1992). 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease, *Genes Dev, 6,* 2088-2099.

Luchtan, M. (2004). TcruziDB: an integrated *Trypanosoma cruzi* genome resource. *Nucleic Acids Res, 32,* D344-D346.

Luo, Y., Na, Z., & Slavoff, S. A. (2018). P-Bodies: Composition, properties, and functions. *Biochem, 57,* 2424-2431.

Lykke-Andersen, J., & Wagner, E. (2005). Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1. *Genes Dev*, *19*(3), 351–361.

Lypaczewski, P., Hoshizaki, J., Zhang, W. W., McCall, L. I., Torcivia-Rodriguez, J., Simonyan, V., … Matlashewski, G. (2018). A complete *Leishmania donovani* reference genome identifies novel genetic variations associated with virulence. *Sci Rep*, *8*(1).

Machado, G. U., Prates, F. V., & Machado, P. R. L. (2019). Disseminated leishmaniasis: Clinical, pathogenic, and therapeutic aspects. *An Bras Dermatol, 94*(1), 9–16.

Maillet, L., Tu, C., Hong, Y. K., Shuster, E. O., & Collart, M. A. (2000). The essential function of Not1 lies within the Ccr4-Not complex. *J Mol Biol*, *303*(2), 131–143.

Manful, T., Fadda, A., & Clayton, C. (2011). The role of the 5′-3′ exoribonuclease XRNA in transcriptome-wide mRNA degradation. *RNA*, *17*(11), 2039–2047.

Mani, J., Güttinger, A., Schimanski, B., Heller, M., Acosta-Serrano, A., Pescher, P., … Roditi, I. (2011). Alba-domain proteins of trypanosoma brucei are cytoplasmic RNA-Binding proteins that interact with the translation machinery. *PLoS ONE*, *6*(7), e33463.

Marr, J. J., Berens, R. L., & Nelson, D. J. (1978). Purine metabolism in *Leishmania donovani* and *Leishmania braziliensis. Biochimica et Biophysica Acta, 554,* 360-371.

Martin, J. L., Yates, P. A., Boitz, J. M., Koop, D. R., Fulwiler, A. L., Cassera, M. B., … Carter, N. S. (2016). A role for adenine nucleotides in the sensing mechanism to purine starvation in *Leishmania donovani*. *Mol Microbiol*, *101*(2), 299–313.

Martin, J. L., Yates, P. A., Soysa, R., Alfaro, J. F., Yang, F., Burnum-Johnson, K. E., … Carter, N. S. (2014). Metabolic reprogramming during purine stress in the protozoan pathogen *Leishmania donovani*. *PLoS Pathog*, *10*(2), e1003938.

Martínez-Calvillo, S., Vizuet-De-Rueda, J. C., Florencio-Martínez, L. E., Manning-Cela, R. G., & Figueroa-Angulo, E. E. (2010). Gene expression in trypanosomatid parasites. *J Biomed Biotechnol.*

Masina, S., Zangger, H., Rivier, D., & Fasel, N. (2007). Histone H1 regulates chromatin condensation in *Leishmania* parasites. *Exp Parasitol*, *116*(1), 83–87.

Matthews, K. R. (2005). The developmental cell biology of *Trypanosoma brucei*. *J Cell Sci*, *118*(2), 283–290.

Matthews, K. R., Tschudi, C., & Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev, 8,* 491–501.

Méndez, S., Fernández-Pérez, F. J., de la Fuente, C., Cuquerella, M., Gómez-Muñoz, M. T., & Alunda, J. M. (1999). Partial anaerobiosis induces infectivity of *Leishmania infantum* promastigotes. *Parasitol Res*, *85*, 507–509.

Mikolajczak, S. A., Silva-Rivera, H., Peng, X., Tarun, A. S., Camargo, N., Jacobs-Lorena, V., Daly, T. M., Bergman, L. W., de la Vega, P., Williams, J., Aly, A. S. I., & Kappe, S. H. I. (2008). Distinct malaria parasite sporozoites reveal transcriptional changes that cause differential tissue infection competence in the mosquito vector and mammalian host. *Mol Cell Biol*, *28*(20), 6196–6207.

Minia, I., Merce, C., Terrao, M., & Clayton, C. (2016). Translation regulation and RNA granule formation after heat shock of procyclic form *Trypanosoma brucei*: Many heat-induced mRNAs are also increased during differentiation to mammalian-infective forms. *PLoS Neglect Trop D, 10*(9), e0004982.

Mishra, M., Singh, M. P., Choudhury, D., Singh, V. P., & Khan, A. B. (1992). Ampothericin versus pentamidine in antimony-unresponsive kala-azar. *The Lancet, 340,* 1256-1257.

Mittra, B., Cortez, M., Haydock, A., Ramasamy, G., Myler, P. J., & Andrews, N. W. (2013). Iron uptake controls the generation of *Leishmania* infective forms through regulation of ROS levels. *J Exp Med, 210*(2), 401–416.

Moradin, N., & Descoteaux, A. (2012). *Leishmania* promastigotes: building a safe niche within macrophages. *Front Cell Infect Microbiol, 2,* 1-7.

Morales, L., Mateos-Gomez, P. A., Enjuanes, L. and Sola, I. (2013). RNA-Affinity chromatography. *Bio-protocol, 3*(13): e808.

Muhlrad, D., & Parker, R. (2005). The yeast EDC1 mRNA undergoes deadenylation-independent decapping stimulated by Not2p, Not4p, and Not5p. *EMBO J, 24*(5), 1033–1045.

Mugo, E., & Clayton, C. (2017). Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in *Trypanosoma brucei*. *PLoS Pathog, 13*(8), e1006560.

Müller, M., Padmanabhan, P. K., Rochette, A., Mukherjee, D., Smith, M., Dumas, C., & Papadopoulou, B. (2010). Rapid decay of unstable Leishmania mRNAs bearing a conserved retroposon signature 3'-UTR motif is initiated by a site-specific endonucleolytic cleavage without prior deadenylation. *Nucleic Acids Res, 38*(17), 5867–5883.

Murphy, W. J., Watkins, K. F., & Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: Evidence for *trans* splicing. *Cell, 47,* 517-525.

Myler, P. J., Audleman, L., Devos, T., Hixson, G., Kiser, P., Lemley, C., … Stuart, K. (1999). *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci, 96,* 2902-2906.

Myler, P. J., Beverley, S. M., Cruz, A. K., Dobson, D. E., Ivens, A. C., McDonagh, P. D., … Stuart, K. D. (2001). The *Leishmania* genome project: New insights into gene organization and function. *Med Microbiol Immun, 190*(1–2), 9–12.

Najafabadi, H. S., Lu, Z., MacPherson, C., Mehta, V., Adoue, V., Pastinen, T., & Salavati, R. (2013). Global identification of conserved post-transcriptional regulatory programs in trypanosomatids. *Nucleic Acids Res, 41*(18), 8591–8600.

Noé, G., De Gaudenzi, J. G., & Frasch, A. C. (2008). Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes. *BMC Molec Biol, 9*(107).

Opperdoes, F., Butenko, A., Flegontov, P., Yurchenko, V., Lukes, J. (2016). Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J Eukaryot Microbiol, 63*, 657-678.

Ouellette, M., & Papadopoulou, B. (2009). Coordinated gene expression by post-transcriptional regulons in African trypanosomes. *J Biol*, *8*(100).

Pages, M., Bastien, P., Vc, F., Rossi, rie, Bellis, M., Wincker, P., … Roizes, rard. (1986). Chromosome size and number polymorphisms in *Leishmania infantum* suggest amplification/deletion and possible genetic exchange. *Mol Biochem Parasitol*, *36*, 161–168.

Paterou, A., Walrad, P., Craddy, P., Fenn, K., & Matthews, K. (2006). Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development. *J Biol Chem*, *281*(51), 39002–39013.

Perry, K. L., Watkins, K. P., & Agabian, N. (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. *Proc Natl Acad Sci, 84*, 8190-8194.

Peters, N. C., Egen, J. G., Secundino, N., Debrabant, A., Kimblin, N., Kamhawi, S., … Sacks, D. (2008). *In vivo* imaging reveals an essential role for neutrophils in leishmaniasis transmitted by sand flies. *Science*, *321*(5891), 970–974.

Phillipson, M., & Kubes, P. (2019). The healing power of neutrophils. *Trends Immunol, 40*(7), 635–647.

Pimenta, P. F. P., Modi, G. B., Pereira, S. T., Shahabuddin, M., & Sacks, D. L. (1997). A novel role for the peritrophic matrix in protecting *Leishmania* from the hydrolytic activities of the sand fly midgut. *Parasitology*, *115*(4), 359–369.

Pimenta, P. F. P., Saraiva, E. M. B., Rowtont, E., Modi, G. B., Garrawayt, L. A., Beverley, S. M., … Sacks, D. L. (1994). Evidence that the vectorial competence of phlebotomine sand flies for different species of *Leishmania* is controlled by structural polymorphisms in the surface lipophosphoglycan. *Proc Natl Acad Sci, 91*, 9155-9159.

Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., Roth, J. G., Gai, H., Kovalski, J. R., Siprashvili, Z., Palmer, T. D., Carette, J. E., & Khavari, P. A. (2018). RNA-protein interaction detection in living cells. *Nat Methods*, *15*(3), 207–212.

Ramanathan, M., Porter, D. F., & Khavari, P. A. (2019). Methods to study RNA–protein interactions. *Nat Methods*, *16*(3), 225–234.

Reverdatto, S. V., Dutko, J. A., Chekanova, J. A., Hamilton, D. A., & Belostotsky, D. A. (2004). mRNA deadenylation by PARN is essential for embryogenesis in higher plants. *RNA*, *10*(8), 1200–1214.

Reynolds, D., Cliffe, L., Förstner, K. U., Hon, C. C., Siegel, T. N., & Sabatini, R. (2014). Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania major* and *Trypanosoma brucei*. *Nucleic Acids Res*, *42*(15), 9717–9729.

Robinson, K. A., & Beverley, S. M. (2003). Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasitol*, *128*, 217–228.

Rogers, M. E., Chance, M. L., & Bates, P. A. (2002). The role of promastigote secretory gel in the origin and transmission of the infective stage of *Leishmania mexicana* by the sandfly *Lutzomyia longipalpis*. *Parasitology*, *124*(5), 495–507.

Rogers, M. E., & Bates, P. A. (2007). *Leishmania* manipulation of sand fly feeding behavior results in enhanced transmission. *PLoS Pathog*, *3*(6), 818–825.

Rogers, M. E., Ilg, T., Nikolaev, A. V., Ferguson, M. A. J., & Bates, P. A. (2004). Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG. *Nature*, *430*(6998), 463-467.

Rogers, M., Kropf, P., Choi, B. S., Dillon, R., Podinovskaia, M., Bates, P., & Müller, I. (2009). Proteophosophoglycans regurgitated by *Leishmania*-infected sand flies target the L-arginine metabolism of host macrophages to promote parasite survival. *PLoS Pathog*, *5*(8), e1000555.

Romanelli, M. G., Diani, E., & Lievens, P. M. J. (2013). New insights into functional roles of the polypyrimidine tract-binding protein. *Int J Mol Sci*, *14*(11), 22906–22932.

Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J of Cell Bio*, *196*(6), 801–810.

Russell, D. V., Xu, S., & Chakraborty, P. (1992). Intracellular trafficking and the parasitophorous vacuole of *Leishmania mexicana*-infected macrophages. *J Cell Sci*, *103*, 1193–1210.

Saavedra, C., Tung, K.-S., Amberg, D. C., Hopper, A. K., & Cole, C. N. (1996). Regulation of mRNA export in response to stress in *Saccharomyces cerevisiae*. *Genes Dev*, *10*, 1606–1620.

Sachs,', A. B., Davis, R. W., & Kornberg1, R. D. (1987). A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol Cell Biol*, *7*(9), 3268-3279.

Sacks, D. L., Modi, G., Rowton, E., Spä Th §, G., Epstein, L., Turco, S. J., & Beverley, S. M. (2000). The role of phosphoglycans in *Leishmania*-sand fly interactions. *PNAS*, *97*(1), 406–411.

Sacks, D. L., & Perkins, P. V. (1984). Identification of an infective stage of *Leishmania* promastigotes. *Science*, *223*(4643), 1417-1419.

Saini, S., Kumar Ghosh, A., Singh, R., Das, S., Abhishek, K., Kumar, A., … Das, P. (2016). Glucose deprivation induced upregulation of phosphoenolpyruvate carboxykinase modulates virulence in *Leishmania donovani*. *Mol Microbiol*, *102*(6), 1020–1042.

Saito, R. M., Elgort, M. G., & Campbell, D. A. (1994). A conserved upstream element is essential for transcription of the *Leishmania tarentolae* mini-exon gene. *EMBO J*, *13*(22), 5460–5469.

Samaras, N., & Spithill, T. W. (1987). Molecular karyotype of five species of *Leishmania* and analysis of gene locations and chromosomal rearrangements. *Mol Biochem Parasitol*, *25*, 279–291.

Sanchez, M. A., Tryon, R., Pierce, S., Vasudevan, G., & Landfear, S. M. (2004). Functional expression and characterization of a purine nucleobase transporter gene from *Leishmania major*. *Mol Membr Biol*, *21*(1), 11–18.

Santos, V. C., Araujo, R. N., Machado, L. A. D., Pereira, M. H., & Gontijo, N. F. (2008). The physiology of the midgut of *Lutzomyia longipalpis* (Lutz and Neiva 1912): PH in different physiological conditions and mechanisms involved in its control. *J Exp Biol*, *211*(17), 2792–2798.

Scholler, J. K., Reed, S. G., & Stuart, K. (1986). Molecular karyotype of species and subspecies of *Leishmania*. *Mol Biochem Parasitol*, *20*, 279-293.

Schulz, D., Zaringhalam, M., Papavasiliou, F. N., & Kim, H. S. (2016). Base J and H3.V regulate transcriptional termination in *Trypanosoma brucei*. *PLoS Genet*, *12*(1), e1005762.

Schwede, A., Ellis, L., Luther, J., Carrington, M., Stoecklin, G., & Clayton, C. (2008). A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. *Nucleic Acids Res*, *36*(10), 3374–3388.

Schwede, A., Manful, T., Jha, B. A., Helbig, C., Bercovich, N., Stewart, M., & Clayton, C. (2009). The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. *Nucleic Acids Res*, *37*(16), 5511–5528.

Secundino, N. F. C., Eger-Mangrich, I., Braga, E. M., Santoro, M. M., & Pimenta, P. F. P. (2005). *Lutzomyia longipalpis* peritrophic matrix: Formation, structure, and chemical composition. *J. Med. Entomol*, *42*(6), 928-938.

Serafim, T. D., Coutinho-Abreu, I. V., Oliveira, F., Meneses, C., Kamhawi, S., & Valenzuela, J. G. (2018). Sequential blood meals promote *Leishmania* replication and reverse metacyclogenesis augmenting vector infectivity. *Nat Microbiol*, *3*(5), 548–555.

Serafim, T. D., Figueiredo, A. B., Costa, P. A. C., Marques-da-Silva, E. A., Gonçalves, R., de Moura, S. A. L., … Afonso, L. C. C. (2012). *Leishmania* metacyclogenesis is promoted in the absence of purines. *PLoS Neglect Trop D*, *6*(9), e1833.

Serpeloni, M., Moraes, C. B., MunizJoã, J. R. C., Motta, M. C. M., Ramos, A. S. P., Kessler, R. L., … ÁvilaAndré, A. R. (2011). An essential nuclear protein in trypanosomes is a component of mRNA transcription/export pathway. *PLoS ONE*, *6*(6), e20730.

Serpeloni, M., Vidal, N. M., Goldenberg, S., Ávila, A. R., & Hoffmann, F. G. (2011). Comparative genomics of proteins involved in RNA nucleocytoplasmic export. *BMC Evol Biol*, *11*(7).

Shrivastava, R., Drory-Retwitzer, M., & Shapira, M. (2018). Nutritional stress targets LeishIF4E-3 to storage granules that contain RNA and ribosome components in *Leishmania. PLoS Negl Trop Dis.*, *13*(3), e0007237.

Siegel, T. N., Hekstra, D. R., Kemp, L. E., Figueiredo, L. M., Lowell, J. E., Fenyo, D., … Cross, G. A. M. (2009). Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev*, *23*(9), 1063–1076.

Siegel, T. N., Tan, K. S. W., & Cross, G. A. M. (2005). Systematic study of sequence motifs for RNA *trans*-splicing in *Trypanosoma brucei*. *Mol Cell Biochem*, *25*(21), 9586–9594.

Somanna, A., Mundodi, V., & Gedamu, L. (2002). In vitro cultivation and characterization of *Leishmania chagasi* amastigote-like forms. *Acta Trop, 83*, 37-42.

Soysa, R., Carter, N. S., & Yates, P. A. (2014). A dual luciferase system for analysis of post-transcriptional regulation of gene expression in *Leishmania*. *Mol Biochem Parasitol*, *195*, 1–5.

Soysa, R., Tran, K. D., Ullman, B., & Yates, P. A. (2015). Integrating ribosomal promoter vectors that offer a choice of constitutive expression profiles in *Leishmania donovani*. *Mol Biochem Parasitol*, *204*, 89–92.

Spithill, T. W., & Samaras, N. (1987). Genomic organization, chromosomal location and transcription of dispersed and repeated tubulin genes in *Leishmania major*. *Mol Biochem Parasitol*, *24*, 23-37.

Stein, A., Vaseduvan, G., Carter, N. S., Ullman, B., Landfear, S. M., & Kavanaugh, M. P. (2003). Equilibrative nucleoside transporter family members from *Leishmania donovani* are electrogenic proton symporters. *J Biol Chem*, *278*(37), 35127–35134.

Stern, M. Z., Gupta, S. K., Salmon-Divon, M., Haham, T., Barda, O., Levi, S., Wachtel, C., Nilsen, T. W., & Michaeli, S. (2009). Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism. *RNA*, *15*(4), 648–665.

Steverding, D. (2017). The history of leishmaniasis. *Parasit Vectors*, *10*(82).

Stone, N. R. H., Bicanic, T., Salim, R., & Hope, W. (2016, March 1). Liposomal Amphotericin B (AmBisome®): A review of the pharmacokinetics, pharmacodynamics, clinical experience and future directions. *Drugs*, *76*(4), 485-500.

Sturgill-Koszycki, S., Schlesinger, P. H., Chakraborty, P., Haddix, P. L., Collins, H. L., Fok, A. K., … Russell, D. G. (1994). Lack of acidification in *Mycobacterium* phagosomes produced by exclusion of the vesicular proton-ATPase. *Science, 263*(5147), 678-681.

Subota, I., Rotureau, B., Blisnick, T., Ngwabyt, S., Durand-Dubief, M., Engstler, M., & Bastin, P. (2011). ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. *Mol Biol Cell*, *22*(22), 4205–4219.

Sundar, S., Makharia, A., More, D. K., Agrawal, G., Voss, A., Fischer, C., … Murray, H. W. (31AD). Short-course of oral miltefosine for treatment of visceral leishmaniasis. *Clin Infect Dis*, *4*(2000), 1110–1113.

Sunter, J., & Gull, K. (2017, September 1). Shape, form, function and *Leishmania* pathogenicity: from textbook descriptions to biological understanding. *Open Biol*, 7, 170165.

Sunyoto, T., Potet, J., & Boelaert, M. (2018). Why miltefosine - A life-saving drug for leishmaniasis-is unavailable to people who need it the most. *BMJ Glob Health*, *3*(3), e000709.

Svárovská, A., Ant, T. H., Seblová, V., Jecná, L., Beverley, S. M., & Volf, P. (2010). *Leishmania major* glycosylation mutants require phosphoglycans (lpg2 -) but not lipophosphoglycan (lpg1-) for survival in permissive sand fly vectors. *PLoS Neglect Trop D*, *4*(1), e580.

Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F., & Vignali, D. A. A. (2004). Correction of multi-gene deficiency in vivo using a single "self-cleaving" 2A peptide-based retroviral vector. *Nat Biotechnol*, *22*, 589–594.

Terrao, M., Marucha, K. K., Mugo, E., Droll, D., Minia, I., Egler, F., … Clayton, C. (2018). The suppressive cap-binding complex factor 4EIP is required for normal differentiation. *Nucleic Acids Res*, *46*(17), 8993–9010.

Thomas, S., Green, A., Sturm, N. R., Campbell, D. A., & Myler, P. J. (2009). Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics*, *10*(152).

Tonelli, R. R., da Silva Augusto, L., Castilho, B. A., & Schenkman, S. (2011). Protein synthesis attenuation by phosphorylation of eIF2α is required for the differentiation of *Trypanosoma cruzi* into infective forms. *PLoS ONE*, *6*(11), e27904.

Tosato, V., Ciarloni, L., Ivens, A. C., Rajandream, M. A., Barrell, B. G., & Bruschi, C. V. (2001). Secondary DNA structure analysis of the coding strand switch regions of five *Leishmania major* Friedlin chromosomes. *Curr Genet*, *40*(3), 186–194.

Trenaman, A., Glover, L., Hutchinson, S., & Horn, D. (2019). A post-transcriptional respiratome regulon in trypanosomes. *Nucleic Acids Res*, *47*(13), 7063–7077.

Tupperwar, N., Meleppattu, S., Shrivastava, R., Baron, N., Gilad, A., Wagner, G., … Shapira, M. (2020). A newly identified *Leishmania* IF4E-interacting protein, Leish4E-IP2, modulates the activity of cap-binding protein paralogs. *Nucleic Acids Res*, *48*(8), 4405–4417.

Uliana, S. R. B., Ruiz, J. C., & Cruz, A. K. (2006). 'Chapter B02. *Leishmania* genomics: Where do we stand? A short history of the *Leishmania* Genome Project' in *Bioinformatics in Tropical Disease Research: A Practical and Case-Study Approach [Internet]*. Bethesda (MD): National Center for Biotechnology Information.

Ullu, E., Tschudi, C., & Chakraborty, T. (2004). RNA interference in protozoan parasites. *Cell Microbiol*, *6*(6), 509-519.

Utter, C. J., Garcia, S. A., Milone, J., & Bellofatto, V. (2011). Poly(A)-specific ribonuclease (PARN-1) function in stage-specific mRNA turnover in *Trypanosoma brucei*. *Eukaryot Cell, 10*(9), 1230–1240.

Van Leeuwen, F., Wijsman, E. R., Kieft, R., Van Der Marel, G. A., Van Boom, J. H., & Borst, P. (1997). Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*, *Genes Dev, 11,* 3232-3241.

Van Luenen, H. G. A. M., Farris, C., Jan, S., Genest, P. A., Tripathi, P., Velds, A., … Borst, P. (2012). Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell*, *150*(5), 909–921.

Van Zandbergen, G., Hermann, N., Laufs, H., Solbach, W., & Laskay, T. (2002). *Leishmania* promastigotes release a granulocyte chemotactic factor and induce interleukin-8 release but inhibit gamma interferon-inducible protein 10 production by neutrophil granulocytes. *Infect Immun*, *70*(8), 4177–4184.

Vasudevan, G., Carter, N. S., Drew, M. E., Beverley, S. M., Sanchez, M. A., Seyfang, A., … Landfear, S. M. (1998). Cloning of *Leishmania* nucleoside transporter genes by rescue of a transport-deficient mutant. *Proc Natl Acad Sci, 95*, 9873-9878.

Vector, S., Schlein, Y., Jacobson, R. L., & Shlomai, J. (1991). Chitinase secreted by *Leishmania* functions in the sand fly vector, *Proc: Bio Sci, 245*(1313), 121-126.

Walker, D. M., Oghumu, S., Gupta, G., McGwire, B. S., Drew, M. E., & Satoskar, A. R. (2013). Mechanisms of cellular invasion by intracellular parasites. *Cell Mol Life Sci*, *71*(7), 1245–1263.

Walrad, P. B., Capewell, P., Fenn, K., & Matthews, K. R. (2012). The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in *Trypanosoma brucei*. *Nucleic Acids Res*, *40*(7), 2869–2883.

Walrad, P., Paterou, A., Acosta-Serrano, A., & Matthews, K. R. (2009). Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA *cis*-elements. *PLoS Pathog*, *5*(2), e1000317.

Walters, L. L., Modi, G. B., Chaplin, G. L., & Tesh, R. B. (1989). Ultrastructural development of *Leishmania chagasi* in its vector, *Lutzomyia longipalpis (Diptera: Psychodidae), Am. J. Trop.Med. Hyg*, *41*(3), 295-317.

Wang, M., Ogé, L., Perez-Garcia, M. D., Hamama, L., & Sakr, S. (2018, February 1). The PUF protein family: Overview on PUF RNA targets, biological functions, and post transcriptional regulation. *Int J Mol Sci*, *19*(2), 410.

Wardleworth, B. N., Russell, R. J. M., Bell, S. D., Taylor, G. L., & White, M. F. (2002). Structure of Alba: an archaeal chromatin protein modulated by acetylation. *EMBO J, 21*(17), 4654–4662.

Wells, S. E., Hillner, P. E., Vale, R. D., & Sachs, A. B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell, 2,* 135–140.

W.H.O. Control of the leishmaniases: Report of a meeting of the WHO Expert Committee on the Control of the Leishmaniases, Geneva, 22-26 March 2010. WHO technical reporter series: no 949.

Wincker, P., Ravel, C., Blaineau, C., Pages, M., Jauffret, Y., Dedet, J.-P., & Bastien, P. (1996). The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species, *Nucleic Acids Res, 24*(9), 1688-1694.

Wortmann, G., Zapor, M., Ressner, R., Fraser, S., Hartzell, J., Pierson, J., … Magill, A. (2010). Lipsosomal amphotericin B for treatment of cutaneous leishmaniasis. *Am J Trop Med Hyg, 83*(5), 1028–1033.

Wright, J. R., Siegel, T. N., & Cross, G. A. M. (2010). Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. *Mol Biochem Parasitol 172*(2), 141–144.

Wurst, M., Seliger, B., Jha, B. A., Klein, C., Queiroz, R., & Clayton, C. (2012). Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in *Trypanosoma brucei*. *Mol Microbiol, 83*(5), 1048–1063.

Yamashita, A., Chang, T. C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C. Y. A., & Shyu, A. Bin. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat Struct Mol Biol*, *12*(12), 1054–1063.

Yang, G., Choi, G., & No, J. H. (2016). Antileishmanial mechanism of diamidines involves targeting kinetoplasts. *Antimicrob Agents Chemother, 60*(11), 6828–6836.

Yoffe, Y., Zuberek, J., Lerer, A., Lewdorowicz, M., Stepinski, J., Altmann, M., … Shapira, M. (2006). Binding specificities and potential roles of isoforms of eukaryotic initiation factor 4E in *Leishmania*. *Eukaryot Cell, 5*(12), 1969–1979.

Young, J. D., Yao, S. Y. M., Baldwin, J. M., Cass, C. E., & Baldwin, S. A. (2013). The human concentrative and equilibrative nucleoside transporter families, SLC28 and SLC29. *Mol Aspects Med*, 34, 529-547.

Zhang, C., & Muench, D. G. (2015). A nucleolar PUF RNA-binding protein with specificity for a unique RNA sequence. *J Biol Chem, 290*(50), 30108–30118.

Zilberstein, D., & Shapira, M. (1994). The role of pH and temperature in the development of *Leishmania* parasites. *Annu Rev Microbiol*, 48, 449-470.

Zinoviev, A., Léger, M., Wagner, G., & Shapira, M. (2011). A novel 4E-interacting protein in *Leishmania* is involved in stage-specific translation pathways. *Nucleic Acids Res, 39*(19), 8404–8415.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415

APPENDICES

RBP-BioID: Efforts to identify the *LdNT3* stem-loop-binding protein

Throughout my graduate studies, I made several attempts to identify proteins that bind to the *LdNT3* stem-loop. Specifically, I worked on a novel, proximity biotinylation-based strategy (termed RNA-binding protein biotin identification, or RBP-BioID) for *in vivo* RBP capture. Although these efforts proved unsuccessful, a similar approach was recently used to uncover RNA-protein interactions in human embryonic kidney cells, serving as a proof of concept for this technique [Ramanathan, 2018]. Here I provide a brief overview of RBP-BioID and suggestions for future optimization in the context of *LdNT3*.

RBP-BioID was conceived based on the popular biotin identification (BioID) system. Originally designed to study protein-protein interactions, BioID relies on the properties of a promiscuous biotin ligase, which catalyzes the formation and release of highly reactive biotin-AMP [Roux, 2012]. When fused to a protein of interest, the biotin ligase covalently labels exposed lysine residues on proteins within a 10-30 nm radius, irrespective of whether they associate directly or indirectly with the fusion protein [Kim, 2016]. Biotinylated proteins are then captured by biotin-streptavidin affinity purification for identification by mass spectrometry.

To adapt this system for the study of RBPs (Figure A.1), I leveraged a well-characterized RNA-protein interaction from the *Escherichia coli* MS2 bacteriophage. Two copies of an RNA stem-loop structure from the MS2 genome (i.e. the MS2 stem-loop) were inserted into the *LdNT3* mRNA 3'-UTR, 30 nt up- and downstream of the

purine-responsive *LdNT3* stem-loop. In the same cell line, a promiscuous biotin ligase (BioID2) was then fused in-frame to the MS2 coat protein (MS2) and stably expressed in *trans*. MS2 binds the MS2 stem-loop with nanomolar affinity [Peabody, 1993]. Thus, flanking the *LdNT3* stem-loop with these sequences was expected to tether MS2-BioID2 immediately adjacent to the RNA element of interest, facilitating biotinylation of the associated proteins.

In theory, this approach offered several advantages over existing techniques for studying the RNA-binding proteome [reviewed in Jazurek, 2016 and Ramanathan, 2019]. The stringent wash conditions afforded by biotin-streptavidin purification reduce the likelihood of capturing background or nonspecifically associated proteins post-lysis. In addition, just as BioID did for protein-protein interactions, RBP-BioID was designed to capture weak, transient, and indirect RNA-protein associations that are often missed by other methods [Roux, 2012]. However, the major disadvantage of RBP-BioID is its dependence on the MS2 stem-loop to anchor BioID2; inserting this sequence into an RNA target has the potential to disrupt folding, accessibility, and/or the interactions in which said transcript is involved.

Unfortunately, in the case of the *LdNT3*, I found that introduction of the MS2 stem-loops almost completely eliminated purine-responsive regulation by the 3'-UTR (data not shown). This suggests that modification of the UTR somehow prevented binding of the *LdNT3* stem-loop to its cognate *trans*-acting regulatory protein. For this reason, I did not continue with development of the RBP-BioID system. However, it may be possible to address this issue in future studies by changing the length and/or sequence of the spacer regions separating the *LdNT3* stem-loop from the flanking MS2 binding

sites. Alternatively, adding additional G-C base pairs to the base of the *LdNT3* stem-loop could help stabilize the structure in the context of the modified 3'-UTR.

Figure A.1

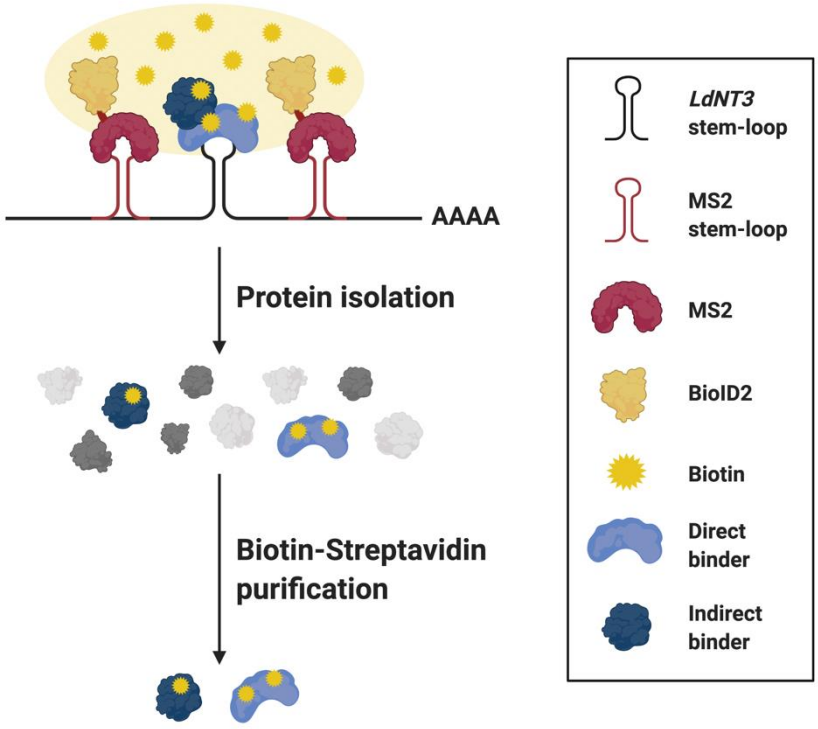
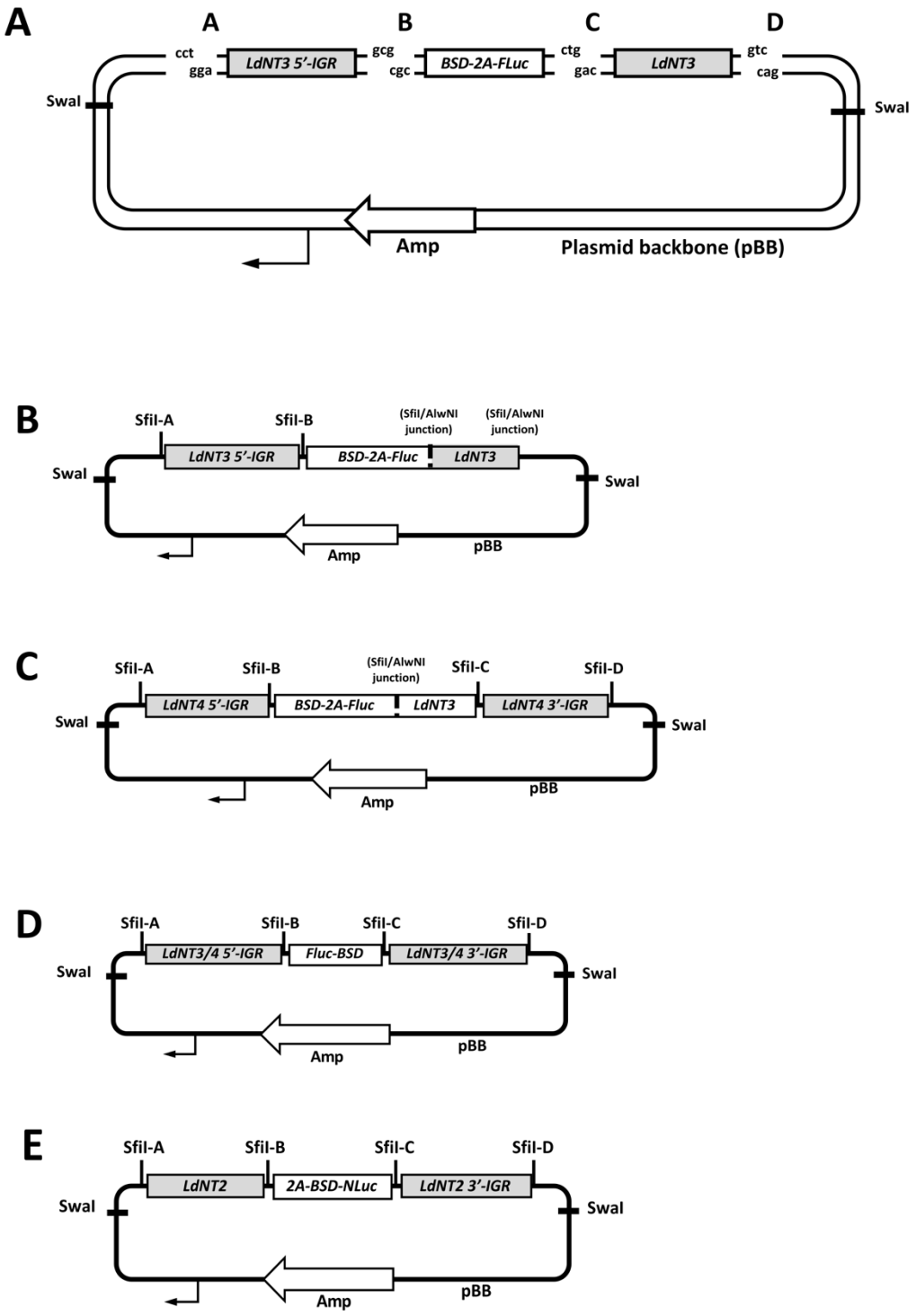


Figure A.1. Schematic of the RBP-BioID strategy for proximity biotinylation and RBP capture. In the endogenous *LdNT3* locus, MS2 stem-loops were inserted up- and downstream of the *LdNT3* stem-loop such that the resultant transcripts harbor the mRNA 3'-UTR depicted above. An MS2-BioID2 fusion transgene was stably expressed in *trans* from the ribosomal RNA locus (not pictured). Binding of MS2 to its RNA target anchors BioID2 adjacent to the *LdNT3* stem-loop and its associated proteins. Upon supplementation with exogenous biotin, BioID2 generates a localized cloud of reactive biotin-AMP (pale yellow), chemically tagging proximal proteins, irrespective of whether they associate directly or indirectly with the *LdNT3* stem-loop. Proteins are then isolated from cell lysate and biotinylated targets are captured by stringent biotin-streptavidin affinity purification for analysis by mass spectrometry.

Figure A.2


A
A
B
C
D
cct
gga
LdNT3 5'-IGR
gcg
cgc
BSD-2A-FLuc
ctg
gac
LdNT3
gtc
cag
SwaI
SwaI
Amp
Plasmid backbone (pBB)
B
SfiI-A
SfiI-B
(SfiI/AlwNI junction)
(SfiI/AlwNI junction)
LdNT3 5'-IGR
BSD-2A-Fluc
LdNT3
SwaI
SwaI
Amp
pBB
C
SfiI-A
SfiI-B
(SfiI/AlwNI junction)
SfiI-C
SfiI-D
LdNT4 5'-IGR
BSD-2A-Fluc
LdNT3
LdNT4 3'-IGR
SwaI
SwaI
Amp
pBB
D
SfiI-A
SfiI-B
SfiI-C
SfiI-D
LdNT3/4 5'-IGR
Fluc-BSD
LdNT3/4 3'-IGR
SwaI
SwaI
Amp
pBB
E
SfiI-A
SfiI-B
SfiI-C
SfiI-D
LdNT2
2A-BSD-NLuc
LdNT2 3'-IGR
SwaI
SwaI
Amp
pBB

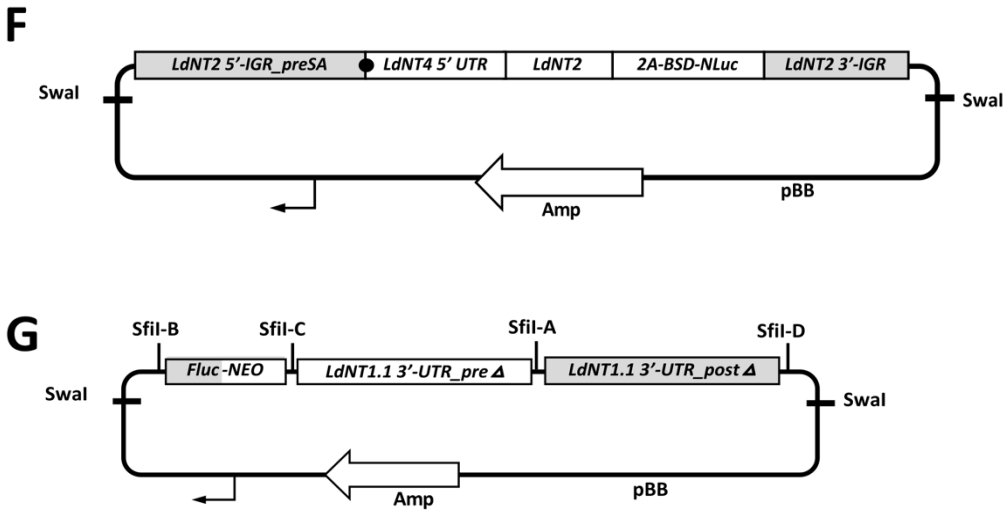


Figure A.2. Gene targeting vector assembly. Constructs utilized in Chapters 2 and 3 are depicted in A-D and D-G, respectively. For simplicity, only the strategy for in-frame genetic fusion of the firefly luciferase (*Fluc*) gene to *LdNT3* is shown here; however, with the exception of F, all other constructs were assembled similarly. Sequences used to direct targeted integration are shaded grey. A) Individual vector components are flanked by sites for restriction endonuclease SfiI or, in the case of the *LdNT3* CDS in B, AlwNI. Restriction sites are designed to generate nonidentical 3'-overhangs upon digestion (designated as junctions A through D), facilitating ordered and directional plasmid assembly in a single ligation reaction. The minimal plasmid backbone (pBB) encodes an origin of replication (black arrow) and an ampicillin resistance gene (AMP). B) Final targeting vector assembled from the components depicted in A. Where indicated, ligation of SfiI- and AlwNI- generated overhangs results in the production of a noncleavable SfiI/AlwNI junction, facilitating subsequent SfiI-mediated cloning of the *BSD-2A-Fluc-LdNT3* fusion. C) Targeting vector for integration of *BSD-2A-Fluc-LdNT3* into the *LdNT4* locus. *BSD-2A-Fluc-LdNT3* was amplified out of the vector depicted in B using primers listed in Table S1. D) Vectors for targeted gene replacement of *LdNT3* or *LdNT4* with the *Fluc-BSD* reporter. E) Assembly of the *LdNT2/NLuc* reporter construct. F) *LdNT2/NLuc* 5'-UTR replacement construct. The complete *LdNT2/NLuc* reporter locus (flanked by both up- and downstream *LdNT2* IGRs) was amplified from transgenic parasites already harboring the plasmid depicted in E and subcloned into pBB for further manipulation. To ensure endogenous-like 5' processing, the preferred *LdNT2* SA site (black circle) was maintained and integration was directed by the remaining portion of 5'-IGR. G) Approach for deletional mutagenesis of the *LdNT1.1* 3'-UTR. The *LdNT1.1* 3' IGR was PCR amplified from genomic DNA as two separate segments, separated by the intended deletion site. Notably, constructs were transfected into a recipient cell line already expressing *Fluc-BSD* from the endogenous *LdNT1.1* locus, such that integration was directed by the *Fluc* reporter gene [Soysa, 2014].

Table A.1. Relative *Fluc-BSD* mRNA abundance (analysis shown in Figure 2.2D). △CT values of biological replicates were averaged prior to relative quantification of Fluc-BSD message level by the comparative CT method.

LdNT3 3'-UTR	Purine	Replicate	ΔCT	Avrg. ΔCT	ΔΔCT	$2^{-\Delta\Delta C_T}$
WT	+	Clone 1	-0.75 ± 0.21	-0.95 ± 0.28	0	1
		Clone 2	-1.15 ± 0.22			
WT	-	Clone 1	-3.67 ± 0.29	-3.92 ± 0.36	-2.97	7.81
		Clone 2	-4.17 ± 0.22			
ΔSL	+	Clone 1	-2.64 ± 0.38	-2.51 ± 0.19	-1.55	2.93
		Clone 2	-2.37 ± 0.12			
ΔSL	-	Clone 1	-3.90 ± 0.18	-3.63 ± 0.39	-2.68	6.39
		Clone 2	-3.35 ± 0.27			

Table A.2. Purine-responsive fold change in *Fluc-BSD* mRNA abundance (analysis shown in Figure 2.2E).

LdNT3 3'-UTR	Clone	Purine	ΔCT	ΔΔCT	$2^{-\Delta\Delta C_T}$	Avrg. $2^{-\Delta\Delta C_T}$
WT	Clone 1	+	-0.75 ± 0.21	-2.91	7.54	7.82 ± 0.39
		-	-3.67 ± 0.29			
WT	Clone 2	+	-1.15 ± 0.22	-3.02	8.09	
		-	-4.17 ± 0.22			
ΔSL	Clone 1	+	-2.64 ± 0.38	-1.27	2.4	2.19 ± 0.30
		-	-3.90 ± 0.18			
ΔSL	Clone 2	+	-2.37 ± 0.12	-0.98	1.97	
		-	-3.35 ± 0.27			

Table A.3. Relative *LdNT2/NLuc* mRNA abundance (analysis shown in Figure 3.2D). △CT values of biological replicates were averaged prior to relative quantification of *LdNT2/NLuc* message level by the comparative CT method.

LdNT2 3'-UTR	Purine	Clone	ΔCT	Avrg. ΔCT	ΔΔCT	$2^{-\Delta\Delta C_T}$
WT	+	Clone 1	-7.39 ± 0.13	-7.19 ± 0.92	0	1
		Clone 2	-6.09 ± 0.10			
WT	-	Clone 1	-6.99 ± 0.14	-6.14 ± 0.**57**	1.05	0.48
		Clone 2	-6.18 ± 0.08			
ΔCU	+	Clone 1	-3.51 ± 0.14	-3.59 ± 0.28	3.60	0.08
		Clone 2	-3.92 ± 0.12			
ΔCU	-	Clone 1	-3.66 ± 0.18	-4.00 ± 0.30	3.19	0.11
		Clone 2	-4.08 ± 0.15			

Table A.4. Purine-responsive fold change in *LdNT2/NLuc* mRNA abundance (analysis shown in Figure 3.2E).

LdNT2 3'-UTR	Clone	Purine	ΔCT	ΔΔCT	$2^{-\Delta\Delta C_T}$	Avrg. $2^{-\Delta\Delta C_T}$
WT	Clone 1	+	-7.39 ± 0.13	1.30	0.41	0.49 ± 0.11
		-	-6.09 ± 0.10			
WT	Clone 2	+	-6.99 ± 0.14	0.81	0.57	
		-	-6.18 ± 0.08			
ΔCU	Clone 1	+	-3.51 ± 0.14	-0.40	1.32	1.33 ± 0.02
		-	-3.92 ± 0.12			
ΔCU	Clone 2	+	-3.66 ± 0.18	-0.43	1.34	
		-	-4.08 ± 0.15			

Table A.5. Purine-responsive fold change in *Fluc-NEO* mRNA abundance (analysis shown in Figure 3.6).

LdNT2 3'-UTR	Clone	Purine	ΔCT	ΔΔCT	$2^{-\Delta\Delta C_T}$	Avrg. $2^{-\Delta\Delta C_T}$
WT	Clone 1	+	-3.91 ± 0.13	0.42	0.75	0.70 ± 0.06
		-	-3.49 ± 0.03			
WT	Clone 2	+	-4.64 ± 0.08	0.60	0.66	
		-	-4.04 ± 0.06			
ΔCU	Clone 1	+	-3.56 ± 0.21	1.27	0.42	0.29 ± 0.18
		-	-2.29 ± 0.09			
ΔCU	Clone 2	+	-3.75 ± 0.06	2.64	0.16	
		-	-1.11 ± 0.15			
ΔSL	Clone 1	+	-2.46 ± 0.10	-0.35	1.27	1.57 ± 0.41
		-	-2.80 ± 0.11			
ΔSL	Clone 2	+	-3.44 ± 0.04	-0.89	1.86	
		-	-4.33 ± 0.10			

Table A.6. Primers for detection of the LdNT3 stem-loop in transgene mRNAs.

Primer Name	Sequence
a	tacacacacggacacagcaagagtg
b	gcactcacgaaggcacgatc
c	gagatgaagaaaaacaacagcagacac
d	tcgcttaaagtgacagtatctcc

Table A.7. Primers for RT-qPCR analyses.

Primer Name	Target TriTrypDB Accession Number	Sequence
UMPS_qForward	LdBPK_160560.1	gattgagcagacgcacgagtacg
UMPS_qReverse		cggcacgaatcacttccgacag
Fluc_qForward	Not Applicable	ggtcgtgctcatgtaccgcttc
Fluc_qReverse		gcttaggtcgtacttgtcgatgagagt
LdNT2_qForward	LdBPK_362040.1	cgcactcttcatgtcgatcatccag
LdNT2_qReverse		ccgattccaatgccgaagtagatgc

Table A.8. Primers for the construction of endogenous targeting vectors. Primer sequences corresponding to the 5'- or 3'-targeting site are in plain text. SfiI/AlwNI sites are in bold with 3'-overhangs underlined.

Primer Name	Sequence	Product Description/Use	Construct
SfiI-A_LdNT3 5-IGR_Forward	ga **ggccacctaggcc** cgcatccctgaacgcatgtg	LdNT3 5'-targeting site	B,D
SfiI-B_LdNT3 5'-IGR_Reverse	ga **ggccacgcaggcc** cttgctgtactagggcgagga		
AlwNI-C_LdNT3 CDS_Forward	ga **cagctgctg** gcggcggtgggagcggtggcggttcgggtatcctcaacttctcctc	LdNT3 3'-targeting site	B
AlwNI-D_LdNT3 CDS_Reverse	ga **caggacctg** agagagtgaacgacagtgtg		
SfiI-C_LdNT3 3-IGR_Forward	ga **ggcctctgtggcc** cccttcctgccctaagcacaatg	LdNT3 3'-targeting site	D
SfiI-D_LdNT3 3-IGR_Reverse	ga **ggcctgactggcc** cgtcagcgatgcgaatcaacac		
SfiIA_LdNT4 5'-IGR_Forward	ga **ggccacctaggcc** gagggatgcacgacgtggcgatc	LdNT4 5'-targeting site	C, D
SfiIB_LdNT4 5'-IGR_Reverse	ga **ggccacgcaggcc** gctcgcactgcgccgtg		
SfiIC_LdNT4 3'-IGR_Forward	ga **ggcctctgtggcc** acttaagcgaaatggcgatacg	LdNT4 3' targeting site	C, D
SfiID_LdNT4 3'-IGR_Reverse	ga **ggcctgactggcc** tcgcttaaagtgacagtatctcc		
SiflA_LdNT2 CDS_Forward	ga **ggccacctaggcc** atgacgggccaatctgctg	LdNT2 5' targeting site	E
SiflB_LdNT2 CDS_Reverse	ga **ggccacgcaggcc** gtaggtcagggttatggcaagagac		
SiflC_LdNT2 3'-IGR_Forward	ga **ggcctctgtggcc** ttcccgcagtggagcttctg	LdNT2 3' targeting site	E
SiflD_LdNT2 3'-IGR_Reverse	ga **ggcctgactggcc** gaggttgtggagagatgcttgcag		
SfiIB_coBSD Forward	ga **ggcctgcgtggcc** atggccaagccgctcagccag	BSD-2A-Fluc-LdNT3 fusion	C
NT3 CDS_SfiIC Reverse	ga **ggccacaggaggcc** agagagtgaacgacagtgtg		
SfiID_SwaI_pBB_Forward	ga **ggccagtcaggcc** atttaaa tggaacaacactcaaccctatc	pBB vector	B - D
SfiI-A_SwaI_pBB_Reverse	ga **ggcctaggtggcc** atttaaat gaaacctgtcgtgccagctgc		

Table A.9. Primers for deletion of the *LdNT3* stem-loop by QuikChange mutagenesis. *LdNT3* version of Plasmid 3.1D used as template.

Primer Name	Sequence
QuikChange_LdNT3_Forward	gcctgcagggcaagaagaggtctcagagccttctcgtctcc
QuikChange_LdNT3_Reverse	gcctgcaggggagacgagaaggctctgagacctcttcttgc

Table A.10. Primers to introduce the *LdNT3* and *TbNT8.1* stem-loops (and mutants thereof) into the *LdNT4* 3'-UTR. *LdNT4* version of Plasmid 3.1D used as template.

Primer Name	Sequence	Stem-loop Sequence Introduced
Wildtype LdNT3_+419_Forawrd*	aaacaacagcagacacttttttcttcacctccc tcgcttatcaagcaaaaatctttccatc	gggagatgaagaaaaacaacagcagacacttttttcttcacctccc
Wildtype LdNT3_+419_Reverse*	aaaagtgtctgctgttgtttttcttcatctccc aactaagtagacaaaaaagacgccatc	
Wildtype LdNT3_Forward	tgaagaaaaacaacagcagacacttttttcttcacctccc caaagggctgtgaagaagacg	gggagatgaagaaaaacaacagcagacacttttttcttcacctccc
Wildtype LdNT3_Reverse	gtgaagaaaaagtgtctgctgttgtttttcttcatctccc caaagacgctgaaacaggag	
Wildtype TbNT8.1_Forward	gagagagggaaaaaacgtcagacactttctctccctccc caaagggctgtgaagaagacg	tgttgagggaaaaaacgtcagacactttctctcgg
Wildtype TbNT8.1_Reverse	ggagggagagaaagtgtctgacgtttttttccctctctccc caaagacgctgaaacaggag	
LdNT3 Loop_TbNT8.1 Stem_Forward	agggaaacaacagcagacactttctctcgg caaagggctgtgaagaagac	tgttgagggaaacaacagcagacactttctctcgg
LdNT3 Loop_TbNT8.1 Stem_Reverse	agagaaagtgtctgctgttgtttccctcaaca caaagacgctgaaacaggag	
TbNT8.1 Loop_LdNT3 Stem_Forward	agaaaaaaaacgtcagacactttttcttca caaagggctgtgaagaagac	tgaagaaaaaaaacgtcagacactttttcttca
TbNT8.1 Loop_LdNT3 Stem_Reverse	agaaaaagtgtctgacgtttttttttcttca caaagacgctgaaacaggag	
LdNT3[Region A mutant]_Forward	agaaaaactttagcagacactttttcttca caaagggctgtgaagaagac	tgaagaaaaactttagcagacactttttcttca
LdNT3[Region A mutant]_Reverse	agaaaaagtgtctgctaaagtttttcttca caaagacgctgaaacaggag	
LdNT3[Region B mutant]_Forward	agaaaaacaacagttttagatttttcttca caaagggctgtgaagaagac	tgaagaaaaacaacagttttagatttttcttca
LdNT3[Region B mutant]_Reverse	agaaaaatctaaaactgttgtttttcttca caaagacgctgaaacaggag	
TbNT8.1[A11C, G15A]_Forward	gaaacaacatcagacactttctctcgg caaagggctgtgaagaagac	ttgagggaaacaacatcagacactttctctcgg
TbNT8.1[A11C, G15A]_Reverse	gaaagtgtctgatgttgtttccctcaa caaagacgctgaaacaggag	
TbNT8.1[A11C, U16G]_Forward	gaaacaacggcagacactttctctcgg caaagggctgtgaagaagac	ttgagggaaacaacggcagacactttctctcgg
TbNT8.1[A11C, U16G]_Reverse	gaaagtgtctgccgttgtttccctcaa caaagacgctgaaacaggag	
TbNT8.1[G15A, U16G]_Forward	gaaaaaacagcagacactttctctcgg caaagggctgtgaagaagac	ttgagggaaaaaacagcagacactttctctcgg
TbNT8.1[G15A, U16G]_Reverse	gaaagtgtctgctgttttttccctcaa caaagacgctgaaacaggag	
TbNT8.1[A11C]_Forward	gaaacaacgtcagacactttctctcgg caaagggctgtgaagaagac	ttgagggaaacaacgtcagacactttctctcgg
TbNT8.1[A11C]_Reverse	gaaagtgtctgacgttgtttccctcaa caaagacgctgaaacaggag	
TbNT8.1[G15A]_Forward	gaaaaaacatcagacactttctctcgg caaagggctgtgaagaagac	ttgagggaaaaaacatcagacactttctctcgg
TbNT8.1[G15A]_Reverse	gaaagtgtctgatgttttttccctcaa caaagacgctgaaacaggag	
TbNT8.1[U16G]_Forward	gaaaaaacggcagacactttctctcgg caaagggctgtgaagaagac	ttgagggaaaaaacggcagacactttctctcgg
TbNT8.1[U16G]_Reverse	gaaagtgtctgccgttttttccctcaa caaagacgctgaaacaggag	

Table A.11. Primers for cloning into and manipulation of pRP-LA and pRP-VH.

Primer Name	Sequence	Product Use/Additional Notes
SfiI-RP1_Fluc_Forward	ga **ggccacctgggcc** atggaagatgccaaaaacattaagaagg	Cloning of firefly luciferase gene into the pRPs. SfiI sites are in bold with 3'-overhangs underlined.
SfiI-RP2_Fluc_Reverse	ga **ggccagcccggcc** ctaccacccacggcgatcttgc	
XmaI_BstXI-B_LmTUB_Forward	ggagaa *cccggg* a **ccaacactatgg** ccaccgcgactgtgtgtgtg	Introduction of nonidentical BstXI sites into the pRP-LA and pRP-VH 3'-UTRs. After whole plasmid amplification, the pRPs were circularized via an XmaI-site (italics). BstXI sites are in bold with 3'-overhangs underlined.
XmaI_BstXI-A_LmTUB_Reverse	ggagaa *cccggg* t **ccatctccttgg** gcacacgcacgcgtgcac	
BstXI-A_LdNT3 SL_Forward	cctcct **ccaaggagatgg** gcaacacgctacaagacaatcccatc	Cloning of the LdNT3 stem-loop into pRP vector. BstXI sites are in bold with 3'-overhangs underlined.
BstXI-B_LdNT3 SL_Reverse	cctcct **ccatagtgttgg** gtgagtgtggtgtgagttgttggtgag	

Table A.12. Primers for step-wise construction of Plasmid 3.1F (i.e. *LdNT2-LdNT4* 5'-UTR replacement).

Primer name	Sequence	Product Use/Additional Notes
LdNT2_SwaI_pBB_Forward	**cgcaacagcc atttaaat** tggaacaacactcaaccctatctcg	Primers were used to amplify *LdNT2/NLuc* at *LdNT2* locus (flanked by *LdNT2* 5'- and 3'-IGRs) from gDNA. PCR products were cloned into pBB via Gibson assembly. Sequence indicated in bold corresponds to pBB, allowing for vector assembly. SwaI sites (part of pBB backbone; used for construct linearization) are underlined.
LdNT2_SwaI_pBB_Reverse	**atcactgcagc atttaaat** gaaacctgtcgtgccagctgc	
LdNT2/NLuc_ATG_Forward	atgacgggccaatctgctgc	Primers exclude the *LdNT2* 5'-UTR from *LdNT2/NLuc* constructs via whole plasmid amplification. Product was assembled with the *LdNT4* 5'-region (see primers below) via Gibson assembly reaction. Sequence added to direct Gibson assembly is in bold.
LdNT2/NLuc_SA_Reverse	**gaggg** cttaacttcgctcgtgtg	
LdNT4_5'-UTR_Forward	**cgaagttaagccctc** agctcttccacgc	Used to amplify 250 nt of the *LdNT4* IGR, immediately upstream of ATG, from gDNA. PCR product was inserted in place of *LdNT2* 5'-UTR in the *LdNT2/NLuc* construct by Gibson assembly reaction. Sequence added to direct Gibson assembly is in bold.
LdNT4_5'-UTR_Reverse	**gcccgt** catcgctcgcactgcgcc	

Table A.13. Primers to generate overlapping deletions in the *LdNT2* 3'-UTR. *LdNT2/NLuc* constructs were modified as depicted in Figure 3.2. NotI/XbaI sites used for plasmid circularization are indicated in bold and italics, respectively.

Primer name	Sequence	Deletion name	Position deleted, relative to stop
delNT2_NotI_171_Forward	ggttgt **gcggccgc** ggttttcaaagaatcaaacgcctacgatg	Δ1	42 - 171
delNT2_NotI_42_Reverse	ggttgt **gcggccgc** aagagaggtagggcagaagctc		
delNT2_NotI_232_Forward	ggttgt **gcggccgc** gtctgctgtcaggtcctcgatg	Δ2	85 - 232
delNT2_NotI_85_Reverse	ggttgt **gcggccgc** ggagtctccgtgctcatacaactc		
delNT2_NotI_323_Forward	ggttgt **gcggccgc** gtgtgactcaaacgtatctcactcg	Δ3	170 - 323
delNT2_NotI_170_Reverse	ggttgt **gcggccgc** attggttcggcagttgcag		
delNT2_NotI_385_Forward	ggttgt **gcggccgc** cgagagatgcaccacgtgtc	Δ4	248 - 385
delNT2_NotI_248_Reverse	ggttgt **gcggccgc** gacctgacagcagacatcagtagc		
delNT2_NotI_481_Forward	ggttgt **gcggccgc** gtgtctgtttcgcaccac	Δ5	349 - 481
delNT2_NotI_349_Reverse	ggttgt **gcggccgc** gagtgagatacgtttgagtcacac		
delNT2_NotI_537_Forward	ggttgt **gcggccgc** gaacttatgtatggcaaagaatggttccgtatac	Δ6	408 -537
delNT2_NotI_408_Reverse	ggttgt **gcggccgc** ggacacgtggtgcatctctcg		
delNT2_XbaI_650_Forward	ggaggc *tctaga* ctttcgtgtgtgcgtgctgc	ΔCU	572 - 650
delNT2_XbaI_572_Reverse	ggaggc *tctaga* gtatacggaaccattctttgccatacataagttc		
delNT2_XbaI_702_Forward	ggaggc *tctaga* cttctagtgtgcttttccgtcctctg	Δ7	674 - 702
delNT2_XbaI_674_Reverse	ggaggc *tctaga* ccagcagcacgcacacacgaaag		
delNT2_NotI_808_Forward	ggttgt **gcggccgc** tgtattagccagtgctcatccttgc	Δ8	704 - 808
delNT2_NotI_704_Reverse	ggttgt **gcggccgc** ctgaagattaaacacgaatagatcaacg		
delNT2_NotI_1004_Forward	ggttgt **gcggccgc** tgaaaggagtgacgcaacag	Δ9	833 - 1004
delNT2_NotI_833_Reverse	ggttgt **gcggccgc** aaggatgagcactggctaatacacg		
delNT2_NotI_1067_Forward	ggttgt **gcggccgc** ggcacacacggacagaagttacc	Δ10	938 - 1067
delNT2_NotI_938_Reverse	ggttgt **gcggccgc** tgcttgcagccgagcacttc		
delNT2_NotI_1110_Forward	ggttgt **gcggccgc** cactggcatgctccactttc	Δ11	1027 - 1110
delNT2_NotI_1027_Reverse	ggttgt **gcggccgc** tgttgcgtcactcctttcagc		

Table A.14. Primers to delete putative DRBD3 binding sites from the *LdNT2/NLuc* construct. Appended sequences used to facilitate circularization of PCR products is indicated in bold.

Primer name	Sequence
LdNT2_delDRBD3_Forward	**tttgcattcatctttttttttttgaacttttttt** ctttcgtgtgtgcgtgctgc
LdNT2_delDRBD3_Reverse	**caaaaaaaaaaaaacagggagaagctttttttt** gtatacggaaccattctttgccatac

Table A.15. Primers to generate overlapping deletions in *LdNT1* 3'-UTR. *Fluc-NEO* constructs were modified as depicted in Figure 3.5.

Primer name	Sequence	Deletion name	Position deleted, relative to stop
SfiIC_LdNT1 3'-UTR_Forward	ga ggcctctgtggcc aggtagcaggcagcacgtaaagag	NA; Universal F primer	
SfiID_LdNT1 3'-UTR_Reverse	ga ggcctgactggcc gctggtggatggcgattctg	NA; Universal R primer	
SfiIA_LdNT1 3'-UTR_792_Reverse	ga ggcctaggtggcc ctgtagacacacacacacacacac	Δ1	792 - 838
SfiIA_LdNT1 3'-UTR_838_Forward	ga ggccacctaggcc tccctccctctctctcgatgaac		
SfiIA_LdNT1 3'-UTR_810_Reverse	ga ggcctaggtggcc tcggaagatttgcaatcacctgtag	Δ2	810 - 855
SfiIA_LdNT1 3'-UTR_855_Forward	ga ggccacctaggcc gatgaacccacccttctcgtagc		
SfiIA_LdNT1 3'-UTR_834_Reverse	ga ggcctaggtggcc ggacggatggtgctgtgtg	Δ3	834 - 887
SfiIA_LdNT1 3'-UTR_887_Forward	ga ggccacctaggcc gctgccgcacttacacg		
SfiIA_LdNT1 3'-UTR_873_Reverse	ga ggcctaggtggcc gtgggttcatcgagagagagg	Δ4	873 - 909
SfiIA_LdNT1 3'-UTR_909_Forward	ga ggccacctaggcc gtatcgacacttcgttgaactcttc		
SfiIA_LdNT1 3'-UTR_885_Reverse	ga ggcctaggtggcc ggaggggctacgagaagggtg	Δ5	885 - 935
SfiIA_LdNT1 3'-UTR_935_Forward	ga ggccacctaggcc taatgctatctgtgttgatgctttca		
SfiIA_LdNT1 3'-UTR_908_Reverse	ga ggcctaggtggcc gcgtgtaagtgcggcag	ΔUE1	908 - 958
SfiIA_LdNT1 3'-UTR_958_Forward	ga ggccacctaggcc tcattttatttttaacatccccctc		
SfiIA_LdNT1 3'-UTR_935_Reverse	ga ggcctaggtggcc aagagttcaacgaagtgtcgatacgc	Δ6	935 - 1000
SfiIA_LdNT1 3'-UTR_1000_Forward	ga ggccacctaggcc cagccatggcatgcttac		
SfiIA_LdNT1 3'-UTR_958_Reverse	ga ggcctaggtggcc aagcatcaacacagatagcattagaagagttc	ΔCU	958 - 1000
SfiIA_LdNT1 3'-UTR_1000_Forward	ga ggccacctaggcc cagccatggcatgcttac		
SfiIA_LdNT1 3'-UTR_986_Reverse	ga ggcctaggtggcc gggatgttaaaaaataaaatgaaagcatcaacacagatagc	Δ7	986 - 1050
SfiIA_LdNT1 3'-UTR_1050_Forward	ga ggccacctaggcc aacggcttccagcgcatac		
SfiIA_LdNT1 3'-UTR_1024_Reverse	ga ggcctaggtggcc caggtaagcatgccatggctgg	Δ8	1024 - 1088
SfiIA_LdNT1 3'-UTR_1088_Forward	ga ggccacctaggcc cacacacacacgcacgtac		
SfiIA_LdNT1 3'-UTR_1060_Reverse	ga ggcctaggtggcc gaagccgttaacatggcacagtcacatg	Δ9	1060 - 1110
SfiIA_LdNT1 3'-UTR_1110_Forward	ga ggccacctaggcc tatatatgtgcacggactctagagc		
SfiIA_LdNT1 3'-UTR_1077_Reverse	ga ggcctaggtggcc gggaggtgtatgcgctggaag	Δ10	1077 - 1143
SfiIA_LdNT1 3'-UTR_1143_Forward	ga ggccacctaggcc gtgtatgtgtgtgtgggtaagatag		
SfiIA_LdNT1 3'-UTR_1123_Reverse	ga ggcctaggtggcc tgcacatatatatatgtacgtgcgtgtg	Δ11	1123 - 1169
SfiIA_LdNT1 3'-UTR_1169_Forward	ga ggccacctaggcc gaaagcaagatgggtagatgcg		
SfiIA_LdNT1 3'-UTR_1151_Reverse	ga ggcctaggtggcc catacacactctctcgctctagagtcc	Δ12	1151 - 1210
SfiIA_LdNT1 3'-UTR_1210_Forward	ga ggccacctaggcc gaggggtggtgtgctgtc		

Table A.16. Primers for *LdNT2-LdNT1* polypyrimidine tract replacement in the *LdNT2/NLuc* construct.

Primer name	Sequence	Product Use/Additional Notes
NT2_delCU_XbaI-F	ggaggc **tctaga** ctttcgtgtgtgcgtgctgc	Primers used to delete the *LdNT2* CU tract from *LdNT2/NLuc* vectors via whole plasmid amplification. PCR products were recircularized using XbaI sites encoded in primers (bold)
NT2_delCU_XbaI-R	ggaggc **tctaga** gtatacggaaccattctttgccatacataagttc	
NT1_CU_XbaI-F:	cctcct **tctaga** ttctaatgctatctgtgttgatgctttc	Used to amplify *LdNT1* polypyrimidine tract from gDNA for insertion into the *LdNT2/NLuc* construct at XbaI restriction sites (bold)
NT1_CU_XbaI-R:	cctcct **tctaga** gcaagtgtgcaggtaagcatg	