ELUCIDATING THE FUNCTION OF INOSITOL TETRAKISPHOSPHATE 1-KINASE IN *TRYPANOSOMA CRUZI*

by

SABRINA ELIZABETH CLINE

(Under the Direction of Roberto Docampo)

ABSTRACT

Trypanosoma cruzi is an obligate intracellular parasite that is the causative agent for Chagas Disease, a chronic bloodborne infection endemic to the Americas. Inositol-tetrakisphosphate 1-kinase (ITPK1) has been described as an enzyme that may mediate an alternative lipid-independent inositol polyphosphate synthesis pathway through phosphorylation of inositol monophosphate and other intermediates. Bioinformatics research has identified hypothetical protein TcCLB.503885.50 as a potential *T. cruzi* ITPK1 (*Tc*ITPK1) homolog and is orthologous to other higher-order eukaryotic ITPK1. The ability of *Tc*ITPK1 to act as the mediator for this alternative pathway has been established through yeast complementation assays and SAX-HPLC analysis of radioactively-labeled inositol phosphate species in complemented yeast. Immunofluorescence assays of *T. cruzi* epimastigotes with endogenously tagged *Tc*ITPK1 localizes to the cytosol. Additionally, preliminary data demonstrates *Tc*ITPK1 may be essential to *T. cruzi* survival as multiple CRISPR/Cas9 knockout experiments have resulted in no viable *in vitro* parasites in the epimastigote stage.

INDEX WORDS: Alternative Inositol Polyphosphate Pathway, Inositol Polyphosphate Pathway, Inositol Tetrakisphosphate 1-kinase, ITPK1, *Trypanosoma cruzi*

ELUCIDATING THE FUNCTION OF INOSITOL TETRAKISPHOSPHATE 1-KINASE IN *TRYPANOSOMA CRUZI*

by

SABRINA ELIZABETH CLINE

B.S., Auburn University, 2019

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2022

ELUCIDATING THE FUNCTION OF INOSITOL TETRAKISPHOSPHATE 1-KINASE IN *TRYPANOSOMA CRUZI*

by

SABRINA ELIZABETH CLINE

Major Professor:	Roberto Docampo
Committee:	Ronald Drew Etheridge
	Diego Huet

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2022

DEDICATION

To Todd Cline and Jeanine Wiltgen-Cline, my wonderful parents who always supported my lifelong, enthusiastic pursuit to further my understanding of the scientific world. While I can now tell you that the sky is blue due to the Raleigh scattering of electromagnetic radiation, I cannot begin to find the words to explain how grateful I am to have you in my life. I would not be the strong woman you know today without your constant love and support.

To my cat Luna Marie Cline, the lovely creature who has been by my side since I started my graduate school journey in 2019. Her constant presence and affection has brightened up my days during the pandemic and provided me with the best presentation practice partner.

ACKNOWLEDGEMENTS

I would like to give a huge thank you to my principal investigator and advisor Roberto Docampo for his unwavering support of my scientific journey. The past few years of research have been under unique circumstances due to the pandemic and my personal health adversities; however, Roberto has been a constant positive beacon of support. I have truly enjoyed being a member of the Docampo Laboratory and for all the research opportunities it has provided me.

I also want to deeply thank all the members of the Moreno-Docampo laboratory for their support and willingness to sit down and answer my many, many questions. I would especially like to thank Miguel Chiurillo for his mentorship during my initial rotation and entrance to the laboratory. His continued support of my research while I was on a medical leave of absence allowed for my project to continue until I was able to resume it safely.

I would like to thank Ronald Drew Etheridge and Diego Huet for agreeing to be members on my committee. Their unique perspectives and insights on my research strengthened my project and I appreciate their guidance over the past two years.

I would like to acknowledge and thank our longtime collaborator, Adolfo Saiardi from the University College London, for his generous gift of pCA45 plasmids and BY4741 WT/*plc1*Δ yeast strains and his assistance with the yeast complementation project.

Additionally, I would like to thank the GACRC and Casey Bergman for bolstering my bioinformatics skills to be able to analyze and mine for novel findings in data archives. I would also like to thank Rodrigo P. Baptista for his assistance when constructing my phylogenetic trees.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1:

INTRODUCTION AND LITERATURE REVIEW

Purpose of the Study

Inositol polyphosphates are involved with a myriad of biological processes including phosphate homeostasis, energy metabolism, calcium storage and release, and secondary messenger signaling[1–5]. Inositol trisphosphate (IP_3) is a soluble metabolite and can be generated from two unique biosynthetic pathways: (1) phospholipase C (PLC) hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) in a lipid-dependent mechanism and (2) repeated inositol tetrakisphosphate 1-kinase (ITPK1) activity on inositol monophosphate species generated from cellular metabolites in a lipid-independent mechanism. Functional redundancy of these inositol polyphosphate (IPP) synthesis pathways supports the theory that IPP metabolites and related enzymes play vital regulatory and energy metabolomic roles in eukaryotes. Regardless of their biosynthetic pathway origin, IP_3 molecules can generate evolutionarily conserved inositol polyphosphate ($IP_{4\text{-}6}$) and inositol pyrophosphate species ($IP_{7\text{-}8}$). Inositol phosphate kinases such as inositol polyphosphate multikinase (IPMK) and inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IP5K) generate $IP_{4\text{-}6}$ molecules from the consecutive phosphorylation of IP_3. Once the metabolite IP_6 is generated, inositol pyrophosphate kinases such as inositol hexakisphosphate kinase (IP6K) and diphosphoinositol pentakisphosphate kinase (PPIP5K) can generate the inositol pyrophosphate molecules, $IP_{7\text{-}8}$, that feature high-energy phosphoanhydride bonds and can act as phosphate donors in phosphotransferase reactions[6,7]. The

lipid-independent mechanism of inositol polyphosphate and inositol pyrophosphate synthesis is identical to the lipid-dependent mechanism with the exception that ITPK1 mediates the phosphorylation of I-1,3,4-P_3 to I-1,3,4,5-P_4 due to the isomeric configuration of the phosphate functional groups.

This study investigates the lipid-independent inositol polyphosphate synthesis pathway in *Trypanosoma cruzi*. Inositol tetrakisphosphate 1-kinase (ITPK1) is an evolutionary conserved, multifunctional enzyme with roles that include phosphorylation of endogenous inositol monophosphate species formed from enzyme-catalyzed reactions of glucose-6-phosphate and inositolphosphoceramide with inositol-3-phosphate synthase (INO1) and inositol phosphosphingolipids phospholipase C (ISC1), respectively, and phosphorylation of other inositol phosphate species (I-1,3-IP_2, I-1,3,4-P_3) in higher-order eukaryotes. Human ITPK1 (*Hs*ITPK1) is shown to mediate an alternative inositol polyphosphate pathway and rescue IP_6 synthesis in *plc1*Δ yeast complementation experiments[1]. We hypothesize that *T. cruzi* utilizes this pathway due to the continued synthesis of inositol phosphate species after the knockout of *Tc*PI-PLC, the mediator of the classical inositol polyphosphate pathway in trypanosomes. Our laboratory has been interested in understanding the inositol phosphate synthesis pathway in *T. cruzi* as these molecular species and related enzymes have been implicated to be essential for parasite viability and, therefore, targets of interest for therapeutic design.

This thesis has been divided into six chapters. Chapter 1 provides the necessary background information and literature review related to *Trypanosoma cruzi* and the current understanding of the inositol phosphate synthesis pathways. Chapter 2 describes phylogenetic and bioinformatic studies that identified and predicted *Tc*ITPK1 characterization. Chapter 3 explores the impacts of organometallic compounds on the *T. cruzi* lipid-independent IPP

synthesis pathway in a retrospective RNA-seq analysis. Chapter 4 focuses on CRISPR/Cas9 experiments aiming to knockout, knockdown, and localization of *Tc*ITPK1 in *T. cruzi* Y strain (C6) epimastigote parasites in culture. Yeast complementation experiments that characterize the rescue of inositol polyphosphate synthesis in *T. cruzi* are described in Chapter 5. Finally, Chapter 6 will discuss concluding remarks of the research and future points of study.

Background and Significance

Trypanosoma cruzi is an obligate intracellular parasite that is the causative agent for Chagas Disease (CD), a chronic bloodborne infection that has infected eight million individuals and puts another eighty million at risk of infection in the Americas[8]. Benznidazole and nifurtimox are the only two pharmaceutical drugs used for the treatment of CD[9]. Both benznidazole and nifurtimox have been approved for pediatric use in the U.S. and can be used in adults via an investigational new drug protocol[9–12]. Unfortunately, these treatments only target the acute phase of infection and leave those with the chronic stage at risk of developing life-threatening symptoms such as cardiac arrhythmia, myocarditis, megacolon, or megaesophagus[9,13,14]. It is estimated that two to three million infected individuals experience these life-threatening complications, with 10,000 deaths attributed to this disease annually[15]. Therefore, it is imperative to develop novel therapeutics and vaccines to combat this neglected tropical disease.

While originally endemic to South and Central America, the United States has recently seen an increase in CD prevalence. It is estimated that there are 300,000 individuals infected with CD in the United States alone[14]. The vector of this bloodborne disease, triatome insects, are reported in the southern regions of the United States; however, the prevalence of this disease can

be primarily attributed to globalization and mass migrations seen over the last few decades[9,13,14]. In the United States, it is estimated that less than one percent of individuals infected with CD are diagnosed and treated for this disease[13,14]. The ability of the United States healthcare system to target this prolific parasitic infection is further hampered by limited public and healthcare knowledge of this disease, asymptomatic or mild febrile cases, and different diagnostic methods for the two stages, i.e., thick blood smears with Giemsa staining or molecular detection with polymerase chain reaction (PCR) for acute stage infection and FDA-cleared CD IgG serological assays ($n \geq 2$) for chronic stage infection[9,13]. This gap in disease awareness in the United States puts individuals at higher risk of developing chronic stages and severe cardiomyopathic and gastrointestinal complications. Some control efforts are in place to stop the spread of this disease in the United States, namely the FDA-approved Ortho *T. cruzi* ELISA Test system used to screen donated blood which has been in use since January 2007[16–18]. While there are no FDA-approved CD diagnostic methods, there are four FDA-cleared IgG serological assays that physicians utilize for CD diagnosis: Ortho *T. cruzi* enzyme-linked immunosorbent assay (ELISA) (Ortho Clinical Diagnostics, Raritan, NJ), Hemagen Chagas' kit ELISA (Hemagen Diagnostics, Inc., Columbia, MD), Wiener Chagatest Recombinante v.3.0 ELISA (Wiener Laboratories, Rosario, Argentina), and InBios Chagas Detect Plus (CDP) rapid test (InBios International, Inc, Seattle, WA)[18]. Outside of this limited surveillance and diagnostic effort, the United States is ineffective in monitoring and treating CD in acute and chronically infected individuals. This further demonstrates the growing need for public awareness and continued research on novel treatments for this disease.

Life Cycle of *Trypanosoma cruzi*

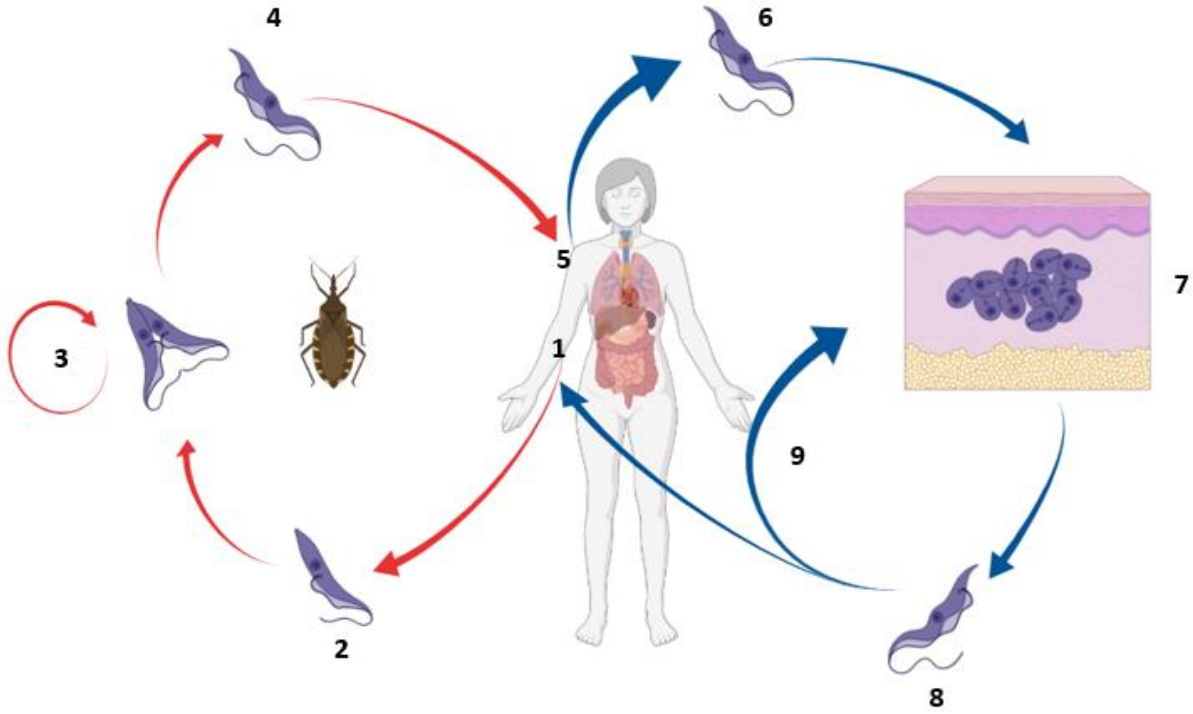


Figure 1. Life Cycle of *Trypanosoma cruzi*. After a triatomine bug takes a blood meal (**1**), *T. cruzi* trypomastigotes transform into epimastigotes in the midgut (**2**) where they replicate by binary fission (**3**). As epimastigotes travel down the insect's gastrointestinal tract, they transform to metacyclic trypomastigotes (**4**). After an infected triatomine bug takes another blood meal (**5**), metacyclic trypomastigotes are then passed out through insect feces and enter the vertebrate host via skin abrasion or mucosal membrane (**6**) where they invade host cells and transform into amastigotes (**7**). After several rounds of replication, amastigotes will undergo another transformation into trypomastigotes (**8**) where they can either invade further host cells (**9**) or be taken up by a triatomine bug (**1**) to restart the cycle.

T. cruzi is a kinetoplastid that features a diheteroxenous life cycle between an invertebrate (insect) vector and a vertebrate (mammalian) host. More than 130 insect species are capable of transmitting the parasite to vertebrate hosts, with the predominant vectors of this

disease being found in the subfamily Triatominae[19]. The life cycle of this parasitic protozoan begins when a triatomine bug bites an infected host and ingests trypomastigotes in its blood meal. After blood consumption, the trypomastigotes will transform into epimastigotes in the insect's midgut and undergo rapid asexual reproduction by binary fission. As the epimastigotes move down the insect's digestive tract and arrive in the hindgut and rectum region, they will undergo a second transformation into the highly infectious metacyclic trypomastigote stage. The triatomine bug will then pass these metacyclic trypomastigotes in its feces upon its next blood meal. Typically, vertebrate hosts can get infected by this parasite-ladened excrement in two routes: (1) metacyclic trypomastigotes enter the bite wound or other skin abrasions or (2) metacyclic trypomastigotes enter the host through intact mucosal membranes. It is important to note that there are other non-vector-specific routes of infection including parasite-contaminated blood transfusion or organ transplants, parasite-contaminated food or beverages, congenital (vertical) transmission of an infection from a mother to her fetus, and laboratory accidents[8].

Once the metacyclic trypomastigotes infect in the vertebrate host, parasites will invade cells around the site of infection and transform into amastigotes. After several rounds of division, the amastigotes will differentiate into infectious trypomastigotes and be released into the bloodstream to infect new cells or be taken up in a blood meal by the insect vector. Clinical manifestations of CD are typically seen in this acute infective stage when the newly formed trypomastigotes burst from the infected cells and enter circulation.

The role of inositol polyphosphate synthesis in these various stages is critical for *Trypanosoma* spp. parasite survival. Inositol polyphosphate multikinase (IPMK) in *T. brucei*, a related bloodborne pathogen, is reported to be an essential enzyme in the soluble inositol polyphosphate synthesis pathway for the bloodstream forms of parasitic infection[20]. Genetic

knockouts and chemical inhibition of *Tb*IPMK leads to defects in acidocalcisome morphology and disrupts inositol polyphosphate synthesis[20–22]. In *T. cruzi*, chemical inhibition of IPMK results in decreased intracellular amastigotes proliferation in host cells ($p < 0.05$)[22]. These discoveries stress the importance of researching the inositol polyphosphate pathways in *Trypanosoma* spp., especially in *T. cruzi* due to the lack of satisfactory therapeutics to target chronic infections.

Classical Inositol Phosphate Synthesis Pathway

Inositol phosphates are important biological molecules ubiquitous to eukaryotic life. These molecules utilize the structurally stable *myo*-inositol ring and form phosphoester bonds with phosphate functional groups to operate a variety of intracellular and extracellular biological roles. Classically, a lipid-dependent pathway generates inositol triphosphate (IP_3) and diacylglycerol (DAG) by phospholipase C (PLC) hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) (**Figure 2**, light yellow). The resultant IP_3 molecule plays a crucial secondary messenger signaling role by binding to the IP_3 gated calcium channel, known as the inositol trisphosphate receptor (IP_3R), found in the endoplasmic reticulum in mammals or the acidocalcisome in *T. cruzi*[3,23]. This binding subsequently activates the release of calcium stores into the cytosol and initiates a signaling cascade to modulate cellular activation and proliferation pathways. IP_3 molecules are either degraded by phosphatases to recycle inositol molecules or further phosphorylated by kinases to form higher-order inositol phosphate species. In addition to calcium signaling, inositol phosphate species are imperative for a myriad of biological roles and activities such as the function of phosphate homeostasis[2,24–26], RNA editing[27], and disease pathogenicity[3,20,28–34].

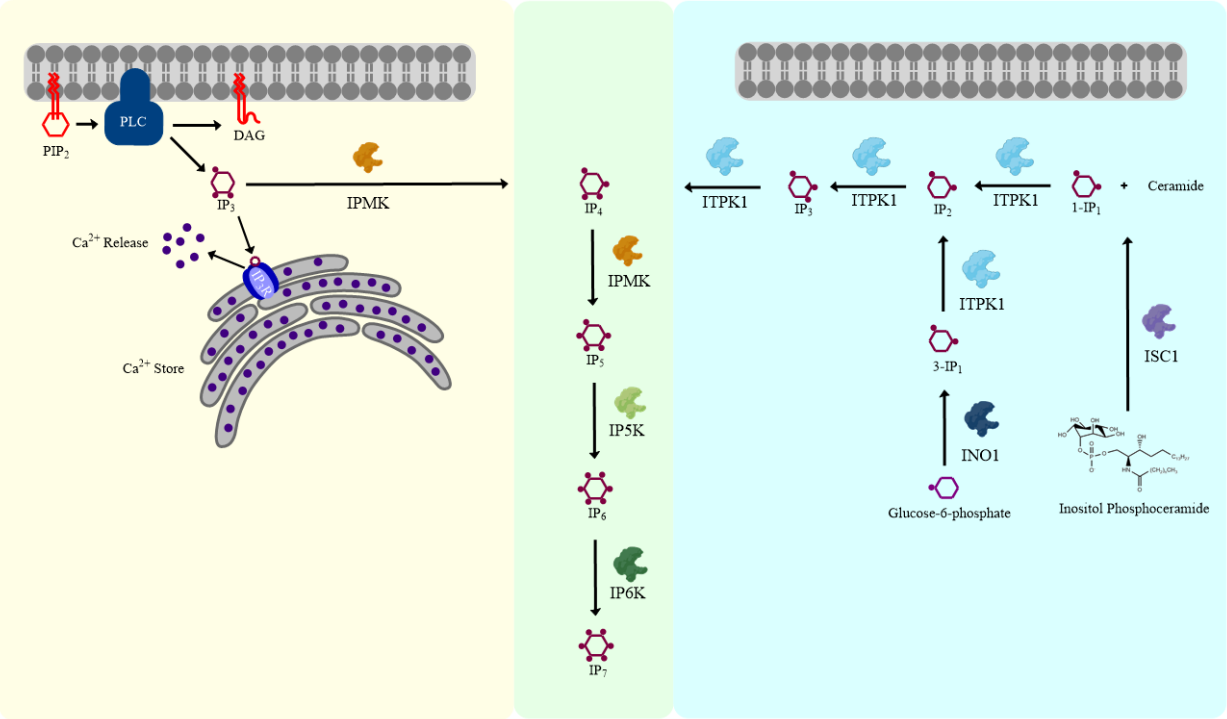


Figure 2. Generation of inositol phosphate species in eukaryotic organisms. Inositol polyphosphate species can be generated by two unique pathways: (1) a lipid-dependent pathway that utilizes phospholipase C to hydrolyze PIP_2 into DAG and I-1,4,5-P_3 (light yellow) and (2) a lipid-independent pathway that uses cellular metabolites such as glucose-6-phosphate and inositol phosphoceramide to form inositol monophosphate isomers and ITPK1 to form I-1,3,4-P_3 (light blue). The isomeric inositol triphosphate molecules from each pathway can then be driven into the universal IPP synthesis pathway to form $IP_{6\text{-}7}$ (light green). The enzymes that mediate these IPP synthesis pathways – inositol-3-phosphate synthase (INO1), inositol phosphosphingolipids phospholipase C (ISC1), inositol tetrakisphosphate 1-kinase (ITPK1), inositol polyphosphate multikinase (IPMK), inositol pentakisphosphate kinase (IP5K), inositol hexakisphosphate kinase (IP6K) – are also represented in this schematic.

Dependent on the isomeric configuration of the three phosphate functional groups, inositol polyphosphate multikinase (IPMK) or an inositol tetrakisphosphate 1-kinase (ITPK1) homolog can generate inositol tetrakisphosphate (IP_4) from IP_3 metabolites (**Figure 2**). IP_4 can be further phosphorylated by IMPK and inositol pentakisphosphate kinase (IP5K) to generate

inositol pentakisphosphate (IP_5) and inositol hexakisphosphate (IP_6) metabolites, respectively. While IP_6 metabolites serve important cellular functions as phosphate storage[35] and signal transduction[36,37], these inositol polyphosphate species can be further phosphorylated by inositol hexakisphosphate kinase (IP6K) and diphosphoinositol pentakisphosphate kinase (PPIP5K) to form inositol pyrophosphate species (IP_{7-8}), important metabolites that regulate cellular pathways by directly phosphorylating its ATP- or GTP-primed targets with phosphate functional groups that are more acid-labile and resistant to phosphatases when compared to ATP phosphorylated targets[6,7,20,35].

Inositol phosphate and higher-order inositol polyphosphate species are important molecules in all three forms of the *T. cruzi* parasitic life cycle. Genetic knockouts of inositol phosphate pathway enzymes result in a decrease of parasite viability and defects in parasitic growth, differentiation, and infectivity[29,32]. *T. cruzi* inositol polyphosphate multikinase (*Tc*IPMK) knockout cultures result in a two-fold reduction of metacyclic forms, with the *TcIPMK* ablated trypomastigotes and amastigotes being less infective to host cells and fewer parasites egressing after their intracellular cycle, respectively[32]. Ablation of *T. cruzi* inositol pentakisphosphate kinase (*TcIP5K*) gene using CRISPR/Ca9 results in abnormal epimastigote morphology and significant reduction in parasite growth[32]. Therefore, the *T. cruzi* inositol phosphate synthesis pathway is of key interest to understanding parasite biology and serves as a potential therapeutic target.

Impact of CRISPR/Cas9 Knockout of *Trypanosoma cruzi* Phosphoinositide-specific Phospholipase C

Phosphoinositide-specific phospholipase C (PI-PLC) is an important component of the inositol polyphosphate (IPP) synthesis pathway and the IP_3/DAG signaling pathway in *T. cruzi*. A previous study highlights that this plasma membrane-bound protein is involved in the trypomastigote to amastigote transformation[38]. Initial *Tc*PI-PLC knockout (KO) attempts utilized active site replacements with the neomycin (*neo*) resistance gene and the addition of a glyceraldehyde 3-phosphate dehydrogenase (*gapdh*) splice acceptor site in one design, enhancing neomycin resistance gene expression through the addition of a poly(A) tail. The two independent attempts to generate a *Tc*PI-PLC KO failed to yield live epimastigotes six weeks post transfection[38]. Other common molecular methods such as RNA interference and switching to the closely related parasite model of *T. brucei* were ruled out due to lack of functional RNAi in *T. cruzi* and lack of intracellular stage in *T. brucei*, respectively. The lack of viable parasites in knockout gene conditions originally led to the conclusion that *Tc*PI-PLC is essential for the survival of parasites in culture.

However, a successful *Tc*PI-PLC knock out in *T. cruzi* Y C6 strain epimastigotes using CRISPR/Cas9-based methodology was recently generated by replacing the 2163 bp gene with the Blasticidin S deaminase (*Bsd*) gene (Miguel Chiurillo and Roberto Docampo, unpublished data). While the growth of transfected epimastigotes is not inhibited in the *Tc*PI-PLC KO lines (**Figure 3A**), the ablation of the *TcPI-PLC* gene significantly impacts host cell invasion ($p < 0.01$) (**Figure 3B**) and intracellular replication ($p < 0.001$) (**Figure 3C**) (Miguel Chiurillo and Roberto Docampo, unpublished data). Due to *Tc*PI-PLC's key role in the IP_3/DAG signaling

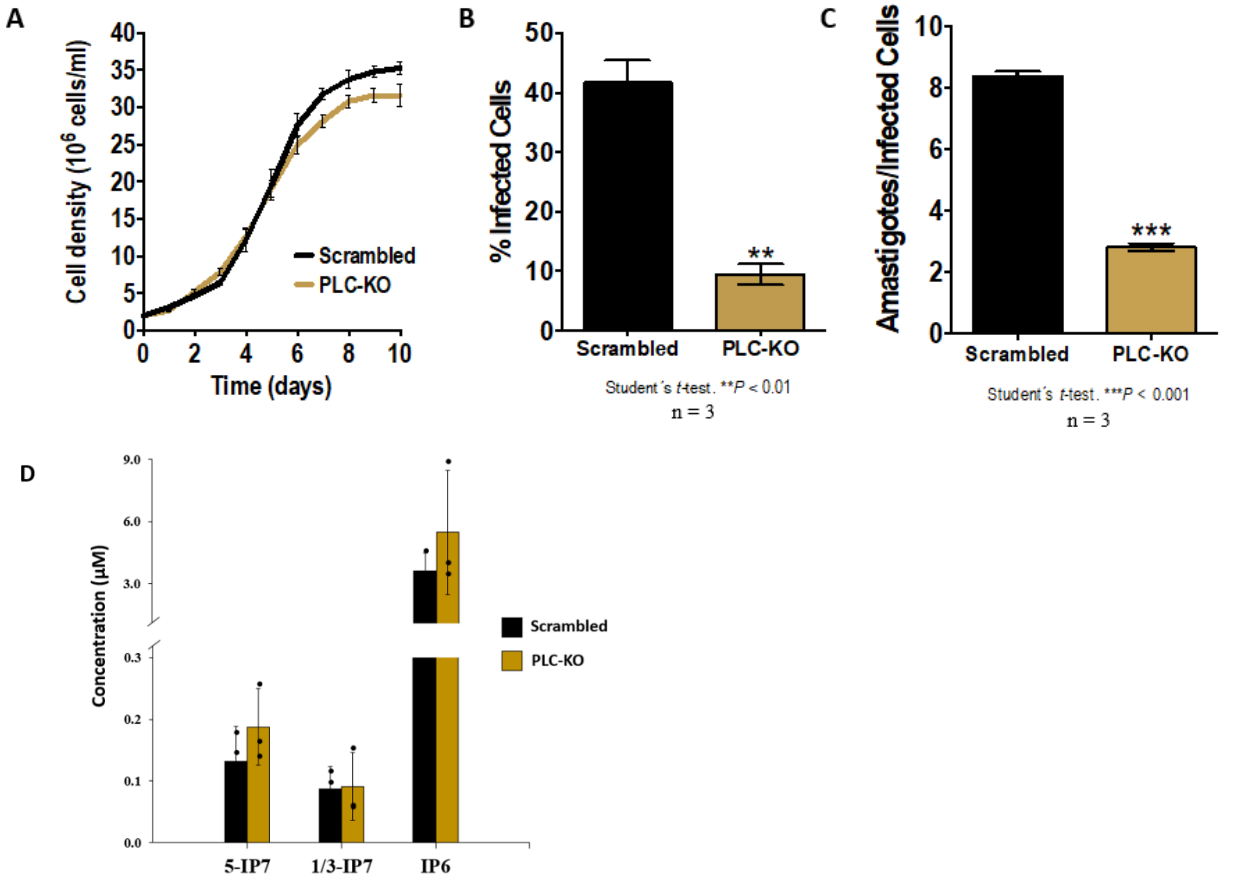


Figure 3. Impact of the knockdown of PI-PLC in *Trypanosoma cruzi*. A. *In vitro* growth assay of *TcPI-PLC* KO and scrambled sgRNA/Cas9-expressing transfected (SCRAMBLED) control epimastigotes cultured in LIT medium until stationary phase. No significant difference between *TcPI-PLC* ablated parasites and control (n = 3 independent test).[A] **B.** *TcPI-PLC* KO and SCRAMBLED trypomastigote infection of Vero cells. There was a significant difference in the percentage of infected Vero cells, **P < 0.01, n = 3 independent tests (student t-test).[A] **C.** *TcPI-PLC* KO and SCRAMBLED *in vitro* intracellular replication assay resulted in significant differences, ***P<0.001, n = 3 (student t-test).[A] **D.** Synthesis of IP_6 and IP_7 Persists in TcPI-PLC-KO *Trypanosoma cruzi* epimastigotes.[B] Both wild type (WT) and *Tc*PI-PLC KO epimastigotes underwent inositol phosphate extraction with perchloric acid and were enriched with TiO_2beads and then analyzed by CE-ESI-MS analysis. *Tc*PI-PLC KO does not significantly decrease the formation of inositol phosphate species IP_6, 5-IP_7, and 1,3-IP_7 when compared to WT (n = 3 independent tests).

[A] Work completed by Miguel Chiurillo and Roberto Docampo. [B] Work completed by Logan Crowe, Dayne Qui, Henning Jessen, and Roberto Docampo.

pathway, we would expect KO lines to have a decrease in the formation of these inositol phosphates and inositol polyphosphate species. Interestingly, capillary electrophoresis coupled to electrospray ionization mass spectrometry (CE-ESI- MS) analysis reveals *Tc*PI-PLC KO epimastigotes did not have a significant impact on the synthesis of inositol phosphate and inositol polyphosphate species when compared to the WT (**Figure 3D**) (Logan Crowe, Dayne Qui, Henning Jessen, and Roberto Docampo, unpublished data). This finding demonstrates that *Tc*PI-PLC is not essential for the formation of inositol polyphosphates and provides evidence to support an alternative inositol polyphosphate synthesis pathway in *T. cruzi*.

Role of Inositol Tetrakisphosphate 1-Kinase in Eukarya

Inositol tetrakisphosphate 1-kinase (ITPK1) is described as a conserved, multifunctional enzyme that may mediate a lipid-independent IP_6 soluble synthesis pathway for higher-order eukaryotes through the phosphorylation of IP_1 and intermediates in the cytosol (**Figure 2**, light blue)[1]. One exception of protein evolutionary conservation in eukaryotes is the notable lack of a *ITPK1* homolog in the *Saccharomyces cerevisiae* genome; however, this absence allows for the characterization of ITPK1 in yeast complementation experiments. Originally, *Homo sapiens* ITPK1 (*Hs*ITPK1) was only identified as an IP_3 5/6 kinase[39]; however, this protein is also found to have enzymatic flexibility with its ability to interconvert IP_4 and IP_5 and act as a 1-phosphatase[5,40]. A recent study demonstrates a new functionality of the *Hs*ITPK1: the mediator of the lipid-independent pathway. *S. cerevisiae plc1*Δ yeast complemented with *Homo sapiens* ITPK1 (*Hs*ITPK1), *Arabidopsis thaliana* ITPK1 (*At*ITPK1), and *Oryza sativa* (*Os*ITPK1) rescues IP_6 formation to that of wild type levels[1]. Strains lacking ITPK1 complementation failed

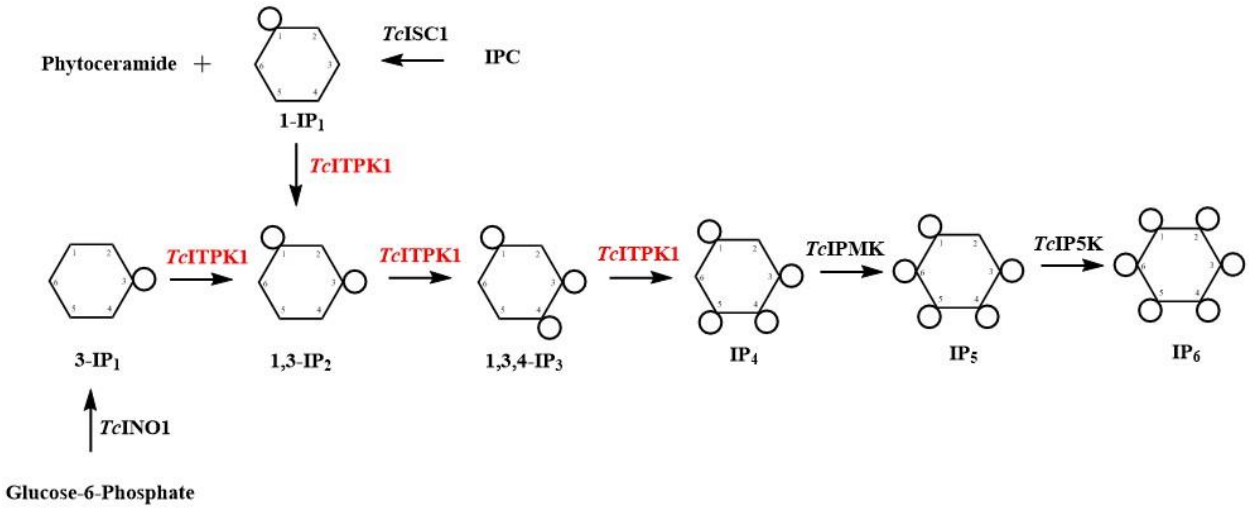


Figure 4. **Proposed Lipid-Independent Inositol Phosphate Synthesis Pathway in *T. cruzi***. When *TcPI-PLC* is knocked out or parasites experience P_i starvation, *T. cruzi* may mediate an alternative inositol phosphate synthesis pathway utilizing inositol monophosphate generated from glucose-6-phosphate (G6P) and inositolphosphoceramide (IPC) by *Tc*INO1 and *Tc*ISC1, respectively. *Tc*ITPK1 will phosphorylate this metabolite multiple times to create the IP_4 metabolite. *T. cruzi* inositol phosphate multikinase (*Tc*IPMK) will phosphorylate IP_4 to IP_5. Then, *T. cruzi* inositol pentakisphosphate kinase (*Tc*IP5K) phosphorylates the last remaining position of IP_5 to make IP_6.

to generate IP_6, supporting the hypothesis of a lipid-independent mechanism of IP and IPP formation[1]. Glucose-6-phosphate (G6P) and inositolphosphoceramide (IPC) are identified as the sources of inositol monophosphate in this lipid-independent pathway in a series of experiments; phosphate starved HCT116 cells demonstrated *de novo* synthesis of IP_6 from G6P while *plc1*Δ*isc1*Δ yeast failed to form IP_6 even with ITPK1 complementation, respectively[1].

Since IPPs are still formed in *TcPI-PLC* ablated epimastigotes (**Figure 3D**), this data strongly supports the hypothesis that *T. cruzi* also mediates a lipid-independent pathway (**Figure 4**). Orthologs of inositol-3-phosphate synthase (*Tc*INO1) and inositol phosphosphingolipids phospholipase C (*Tc*ISC1)–proteins that mediate the formation of endogenous inositol monophosphate via conversion of G6P into 3-IP_1 or IPC into ceramide and 1-IP_1, respectively–

have been identified using TriTrypDB, further supporting this proposed model (Roberto Docampo, unpublished data). The main focus of this study is to identify and characterize *Tc*ITPK1 in order to determine the presence of a lipid-independent IPP synthesis pathway in *T. cruzi*.

CHAPTER 2:

UTILIZING A BIOINFORMATICS TOOLSET TO ANALYZE *Tc*ITPK1 *IN SILICO*

Cline, S.E. and Docampo, R. To be submitted to *Bioinformatics and Biology Insights*.

Abstract

Novel advancements in bioinformatics have created the opportunity to glean deeper insights into molecular evolution, biochemical and biophysical interactions, and structural biology of biomolecules. The push for open-access of raw experimental data from publications and "big data" repositories has resulted in the development of an investigatory pipeline that can mine data in order to identify biomolecules of interest for study. TcCLB.503885.50 was identified on TriTrypDB as a potential homolog of inositol tetrakisphosphate 1-kinase (ITPK1) in *Trypanosoma cruzi*. TcCLB.503885.50 was established as *Tc*ITPK1 after identifying two highly-conserved and three lesser-conserved IP and ATP-binding residue sites when compared to other ITPK1 homologs and featured high orthologous conservation across the parasitic kinetoplastid clade. *Tc*ITPK1's three-dimensional structural biology was resolved by AlphaFold-2.1.1 and COFACTOR predicted its molecular function to include inositol-1,3,4-trisphosphate 5-kinase (GO:0052726), inositol-1,3,4-trisphosphate 6-kinase (GO:0052725), inositol tetrakisphosphate kinase (GO:0051765), and inositol tetrakisphosphate 1-kinase (GO:0047325) activities.

Introduction

The interdisciplinary intersection of molecular biology, evolutionary biology, computational sciences, and novel mathematical and statistical algorithms has allowed for the rapid growth of bioinformatics to predict novel biochemical and biophysical *in silico* relationships based on sequence alone. The past decade has seen a rapid expansion in the bioinformatics toolkit and availability of "big data" repositories as a result of large-scale sequencing efforts, array-based biomolecule profiling, and next-generation sequencing profiling

to understand epigenetics, transcriptomics, and biomolecule interactome[41–43]. TriTrypDB (https://tritrypdb.org/) is a kinetoplastid informatic resource part of the Eukaryotic Pathogen, Vector and Host Informatics Resources (VEuPathDB, https://veupathdb.org/) that aims to promote public access to genomic information and large-scale datasets for eukaryotic kinetoplastid parasites (n = 81) including but not limited to *Trypanosoma cruzi*, *Trypanosoma brucei gambiense*, and *Leishmania braziliensis*[44,45]. This web-based informatics tool allows for researchers to access various omics data without the barrier of learning bioinformatic scripting languages such as Python, R, and BASH. Query searches on TriTrypDB identify TcCLB.503885.50 as a potential homolog of inositol tetrakisphosphate 1-kinase in the organism *Trypanosoma cruzi*.

One field that has embraced this bioinformatics revolution is evolutionary biology. This field regularly employs an array of analytical tools to determine orthologous relationships and reconstruct the evolutionary relationship among a group of related organisms. BLAST, Clustal Omega, and HMMER are a few primary sequence alignment and analysis tools that have arisen in response to the publication of "big data" repositories[46–49]. Molecular Evolutionary Genetics Analysis (MEGA), Geneious Prime, and PHYML are used to construct and analyze phylogenetic trees in an assortment of methods, i.e. maximum likelihood, maximum parsimony, and distance-matrix methods such as neighbor joining[50–53]. This chapter will utilize a variety of primary sequencing and phylogenetic analytical tools to glean a better understanding of the evolutionary relationship of *Tc*ITPK1, our protein of interest, in protozoan parasites.

The applications of line scripting coupled with novel algorithms has evolved a diverse bioinformatics toolset that allows for *in silico* prediction of the structural biology and interactome of biomolecules. AlphaFold-2.1.1 has emerged as a novel artificial intelligence (AI)

program that can accurately predict protein three-dimensional structures and atomic details using a single amino acid sequence and an AI-based algorithm, even if no homologous structure is available[54]. At the Critical Assessment of protein Structure Prediction (CASP-14), AlphaFold-2.1.1 was demonstrated to be highly accurate to a median backbone accuracy of 0.96Å RMSD[54]. This system will be utilized in this chapter to elucidate structural biology evolution of *Tc*ITPK1. COFACTOR will utilize the three-dimensional structure prediction of *Tc*ITPK1 to identify enzyme commission, gene ontology, and ligand-binding sites[55,56].

The major aim of this chapter is to employ an expansive bioinformatics toolset to elucidate the evolutionary relationship, structural biology, and potential interactome of *Tc*ITPK1 *in silico*. First, phylogenetic analyses will verify that TcCLB.503885.50 is the protein of interest, *Tc*ITPK1, and describe its evolutionary relationship in protozoan parasites. AlphaFold-2.1.1 will be verified as an accurate predictor of ITPK1 molecules through a series of comparative experiments based on AI-predicted *Tc*ITPK1 and *Hs*ITPK1 structural comparisons to X-Ray diffraction-resolved *Hs*ITPK1 structure. This data will provide a preliminary basis of future experiments related to retrospective analyses (**Chapter 3**), CRISPR/Cas9 genetic manipulation (**Chapter 4**), and yeast complementation experiments (**Chapter 5**).

Materials and Methodology

Clustal Omega Analysis: Nine homologous ITPK1 amino acid sequences of interest were identified from organisms *Lokiarcheum candidatus*, *Naegleria fowleri*, *Entamoeba histolytica*, *Monosiga brevicollis*, *Homo sapiens*, *Danio rerio*, and *Trichomonas vaginalis* from a previous publication[1]. A potential *Tc*ITPK1 homolog was identified in TriTrypDB (version 45, https://tritrypdb.org/) as TcCLB.503885.50[44]. A fasta file containing all ten homologous amino

acid sequences was uploaded to Clustal Omega Multiple Sequence Alignment (https://www.ebi.ac.uk/Tools/msa/clustalo/) and default multiple sequence alignment parameters were used[49].

Constructing ITPK1 Phylogenetic Tree: The initial amino acid sequence for *T. cruzi* inositol tetrakisphosphate 1-kinase (*Tc*ITPK1) was obtained from TriTrypDB (version 55, https://tritrypdb.org/)[44]. TriTrypDB and OrthoMCL search analyses were performed using the full-length protein sequence of *T. cruzi* ITPK1 protein (TcCLB.503885.50) and orthology group number (OG6_147480)[44,57]. In addition, previously published *Tc*ITPK1 orthologous outgroups from two amoebozoan organisms and one mammal organism were added to the compilation of identified orthologs for a total of 29 unique sequences. Amino acid sequences were aligned using MAFFT (v7.475) with the L-INS-I algorithm, a very slow but accurate alignment algorithm for < 200 sequences[58,59]. The maximum likelihood tree was created with the Molecular Evolutionary Genetics Analysis (MEGAX) program (v-10.0) with the JTT matrix-based model[50]. Bootstrap analysis of 1000 replicates was used for the construction of the phylogenetic tree to improve tree readability.

AlphaFold-2.1.1 Prediction and Modeling of TcITPK1 and HsITPK1: AlphaFold Version 2.1.1. was installed on the Sapelo2 cluster by GACRC using EasyBuild following the steps in the dockerfile available at https://github.com/deepmind/alphafold[54]. The run_alphafold.sh bash script was obtained from https://github.com/kalininalab/alphafold_non_docker and related documentation is available at that URL. The run_alphafold.sh bash script was edited to run the

fasta file containing the amino acid sequence for either *Tc*ITPK1 or *Hs*ITPK1 using the monomer_casp14 modeling parameters.

*Tc*ITPK1

Fasta File containing amino acid sequence for *Tc*ITPK1: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%202/TcITPK1.fasta. Git revision for this file is 0a5796a9e836493b06bc9522904bcf104d1d98ed.

Final Bash Script for AlphaFold-2.1.1 *Tc*ITPK1 is available at: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%202/AlphaFoldTcITPK1.sh. Git revision for this file is 0a5796a9e836493b06bc9522904bcf104d1d98ed.

*Hs*ITPK1

Fasta File containing amino acid sequence for *Hs*ITPK1: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%202/HsITPK1.fasta. Git revision for this file is 0a5796a9e836493b06bc9522904bcf104d1d98ed.

Final Bash Script for AlphaFold-2.1.1 *Hs*ITPK1 is available at: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%202/AlphaFoldHsITPK1.sh. Git revision for this file is 0a5796a9e836493b06bc9522904bcf104d1d98ed.

Output files can be found on the Sapelo2 cluster at /scratch/sec84829/DocampoLaboratory/AlphaFoldTcITPK1 folder or /scratch/sec84829/DocampoLaboratory/AlphaFoldHsITPK1 folder for *Tc*ITPK1 and *Hs*ITPK1, respectively. The ranking_debug file contains the five CASP14 modeling pLDDT outputs and provides information about the best predicted model. The best models for *Tc*ITPK1 (Model 3 – 83.0469) and *Hs*ITPK1 (Model 1 – 83.2896) were selected for downstream analysis. These models were then uploaded to iCn3D, a web-based 3D structure viewer available at NCBI (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html)[60]. Protein ribbon style and AlphaFold Confidence coloring was selected for the model.

Protein Structural Comparison Analysis: Pairwise Structure Alignment is a tool on the RCSB Protein Data Bank website that allows for the comparison of two or more protein structures (https://www.rcsb.org/alignment)[61–67]. As a built-in control for AlphaFold-2.1.1 prediction accuracy, the best model for *Hs*ITPK1 predicted by AlphaFold-2.1.1 Monomer_Casp14 algorithm was compared to the X-Ray diffraction-resolved *Hs*ITPK1 (2ODT) structure (https://www.rcsb.org/structure/2ODT) using the jFATCAT (flexible) parameter[68,69]. Pairwise Structural Alignment output a table describing the RMSD, TM-Score, Score, SI%, SS%, and overall length of the structures selected/uploaded for superposition. Structural comparison was repeated at FATCAT (https://fatcat.godziklab.org/) using the FATCAT (flexible) to find the significance of structural comparisons[66]. This protein structural comparison analysis pipeline was repeated for the best AlphaFold-2.1.1 predicted model for *Tc*ITPK1 and X-Ray diffraction-resolved *Hs*ITPK1 (2ODT) structure (https://www.rcsb.org/structure/2ODT).

COFACTOR Analysis and Ligand-Protein Modeling of TcITPK1: COFACTOR employed a comparative algorithm (I-TASSER_FUNCTION) that combines three dimensional structure, amino acid residue sequence, and predicted protein-protein interactions to described protein function in a web-based format available at Zhang Lab at the University of Michigan (https://zhanggroup.org/COFACTOR/)[55,56]. The PDB file for used in structure-based functional prediction study is found at /scratch/sec84829/DocampoLaboratory/AlphaFoldTcITPK1/TcITPK1/relaxed_model_3 on the GACRC Sapelo2 cluster. After uploading the file into the online query, COFACTOR identified function sites and homologies by matching local and global structures to the BioLiP protein function database. COFACTOR then utilized UniProt-GOA and STRING to provide gene ontology prediction based on sequence-profile alignments and protein-protein interaction inferences, respectively. Predicted ligand binding site(s) of *Tc*ITPK1were downloaded to local computer as a complex1.pdb file. This file was then uploaded to iCn3D, a web-based 3D structure viewer available at NCBI (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html)[60]. Protein ribbon style and AlphaFold Confidence coloring was selected for the model.

Results

ITPK1 is a conserved inositol phosphate kinase and recently studies have demonstrated that *Lokiarcheum candidatus* had structural and homology similarities with *Homo sapiens* ITPK1 and *Entamoeba histolytica* ITPK1 with highly conserved residues at H167, K199, and R212 found near the IP-contact residues and proximal to the ATP binding site[1]. Additionally, these ITPK1 homologs share lesser conserved residues at K18, K59, H162, and G301[1]. A Clustal Omega multiple sequence alignment of previously established homologs including species from

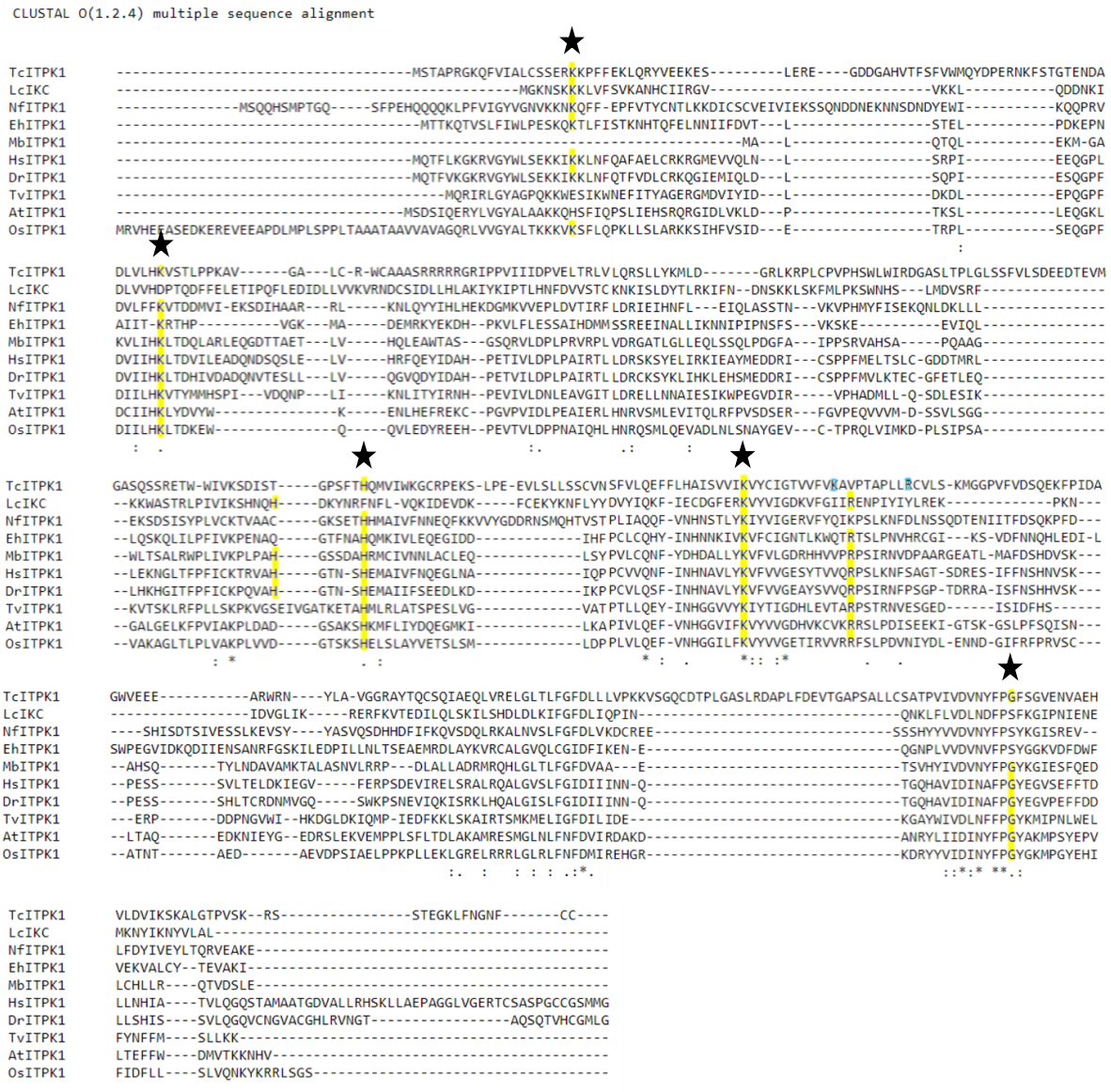


Figure 5. Clustal Omega Multiple Sequence Alignment of ITPK1 Orthologs with Proposed *Tc*ITPK1 (TcCLB.503885.50) protein. Tc: *Trypanosoma cruzi* (protozoan), Lc: *Lokiarcheum candidatus* (archaea), Nf: *Naegleria fowleri* (amoeboflagellate excavate), Eh: *Entamoeba histolytica* (amoebae), Mb: *Monosiga brevicollis* (Choanoflagellate), Hs: *Homo sapiens* (animalia), Dr: *Danio rerio* (animalia), Tv: *Trichomonas vaginalis* (protozoan), At: *Arabidopsis thaliana*, Os: *Oryza sativa*. Yellow highlighted amino acid residues denote evolutionary conserved IP and ATP binding site of ITPK1. Blue highlighted residues denote probable residue substitutions of highly conserved residues for *Tc*ITPK1. Black stars denote conserved residues of known ITPK1 homologs with TcCLB.503885.50.

groups archaea (*Lokiarcheum candidatus*), plantae (*Arabidopsis thaliana*, *Oryza sativa*), animalia (*Homo sapiens*, *Danio rerio*), amoebozoa (*Entameoba histolytica*), amoeboflagellate excavate (*Naegleria fowleri*), and other miscellaneous groups (*Trichomonas vaginalis*, *Monosiga brevicollis*) and the hypothetical protein TcCLB.503885.50 was utilized to determine if this protein has these conserved IP and ATP contact residues. TcCLB.503885.50 is shown to share the H167 and K199 highly conserved *Hs*ITPK1 residue sites and a few of the lesser conserved *Hs*ITPK1 residue sites at K18, K59, and G301 (**Figure 5**). The highly conserved *Hs*ITPK1 R212 residue is hypothesized to be replaced by either the lysine residue at position 242 or the arginine residue at position 253 in TcCLB.503885.50. The CLUSTAL Omega alignment reinforces the hypothesis that the hypothetical protein TcCLB.503885.50 is *Tc*ITPK1.

Once TcCLB.503885.50 was established as *Tc*ITPK1, a phylogenetic evolutionary tree was constructed to understand the evolution of ITPK1 and, by association, the lipid-independent IPP pathway in protozoan parasites with outgroups of free-living kinetoplastid *Bodo saltans*, amoeboflagellate excavate *Naegleria fowleri*, amoebozoan parasite *Entamoeba histolytica*, and *Homo sapiens* (**Figure 6**). Within the kinetoplastid parasite branches, there is high conservation of ITPK1 with all but one branching having values of 80 or above (bootstrap values of 70 or higher are ideal for mapping evolutionary processes because these provide more accurate, confident results). *T. theileri* and *T. grayi* ITPK1 may be less conserved when compared to other kinetoplastid ITPK1 orthologs due to evolutionary pressure related to their restricted vertebrate host range of bovine and African crocodiles, respectively[70,71]. The comparison of parasitic and free living kinetoplastids still demonstrates a high level of evolutionary conservation (Bootstrap value = 99). Outgroup orthologs, expectedly, did not share the same level of evolutionary conservation.

Preliminary bioinformatic predictions of *Tc*ITPK1 structural biology were contradictory with the known function and expected subcellular localization of the lipid-independent IPP pathway. Both InterPro and PROTTER analyses predicted that there is a transmembrane domain for *Tc*ITPK1; however, none of the known orthologs have this domain and the lipid-independent IPP pathway occurs in the cytosol[1,72,73]. To better understand the protein structure and cellular localization of *Tc*ITPK1, a series of bioinformatics studies (**Figure 8**) and CRISPR/Cas9 endogenous C-terminal tagged *Tc*ITPK1 epimastigotes immunofluorescent assay (IFA) studies (**Chapter 4**) were conducted to resolve these conflicting predictions.

The field of structural bioinformatics has rapidly evolved over the past decade allowing for the creation of AlphaFold-2.1.1, an artificial intelligence algorithmic network that utilizes the queried amino acid sequence to predict the three-dimensional structure of proteins[54]. To validate AlphaFold-2.1.1 as a method of accurately predicting ITPK1 structure, a pilot experiment was conducted to compare the structure of *Hs*ITPK1 predicted by AlphaFold-2.1.1 with the X-ray diffraction resolved structure of *Hs*ITPK1, known as 2ODT in the RCSB PDB, using the Pairwise Structure Alignment Tool[61–69]. Both the full structural (**Figure 7A**) and aligned residue (**Figure 7B**) comparison showed high conservation in the secondary protein structure and tertiary spacing. The only major difference noted between the two structures was between residues 322-414 of AlphaFold-2.1.1 predicted *Hs*ITPK1 which is seen in grey (**Figure 7A**). This region of the AlphaFold-2.1.1 prediction has a low predicted local-distance difference test (pLDDT) (<50) and is not highly conserved.

The table showing structural alignment characterizations further validates AlphaFold-2.1.1 as an accurate predictor of ITPK1 proteins (**Figure 7**). When comparing the bioinformatic-

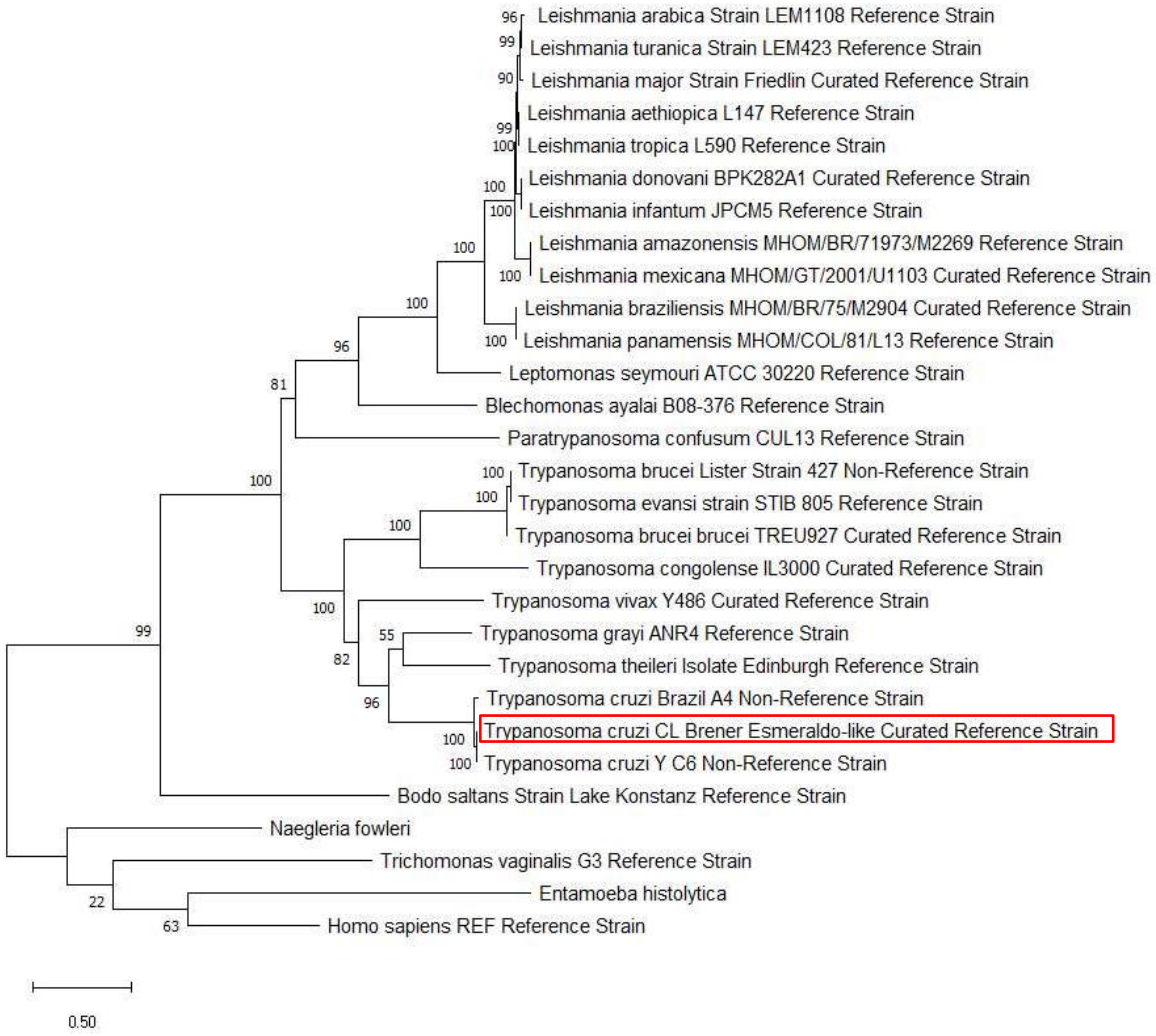
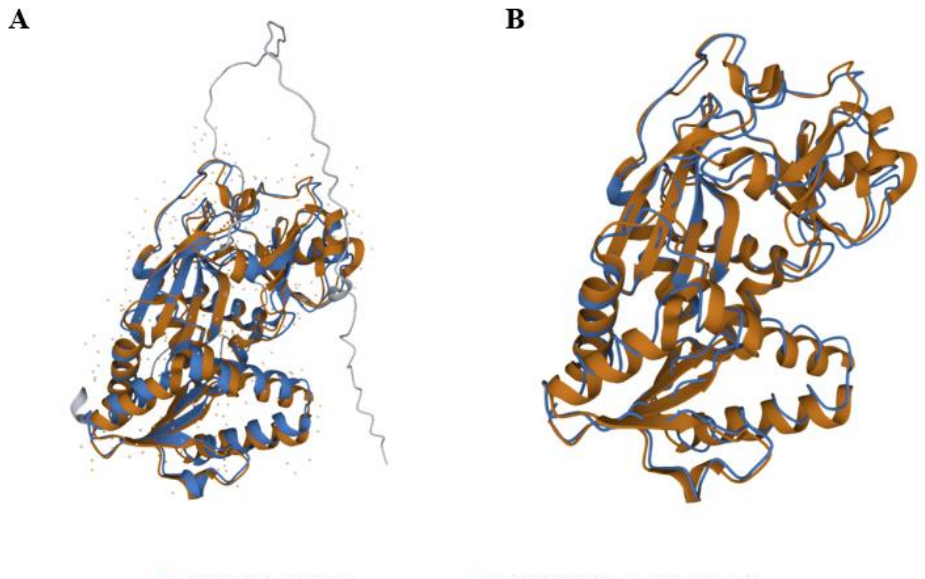


Figure 6. Phylogenetic Study of ITPK1 in Protozoan Parasites. Protozoan parasite ITPK1 phylogenetic tree was created using orthologs identified in TriTrypDB, OrthoMCL, and previously published manuscripts using reference strains and select non-reference strains pertinent to the protein of interest (TcCLB.503885.50). The addition of parasitic and non-parasitic *Eukarya* ITPK1 ortholog outgroups strengthen the phylogenetic study. The orthologs were aligned using the latest version of MAFFT (v.7.475) and employing MEGA-X (v.10.0) for maximum likelihood estimation. Bootstrap analysis of 1000 replicates was used for the construction of the phylogenetic tree to improve tree readability.



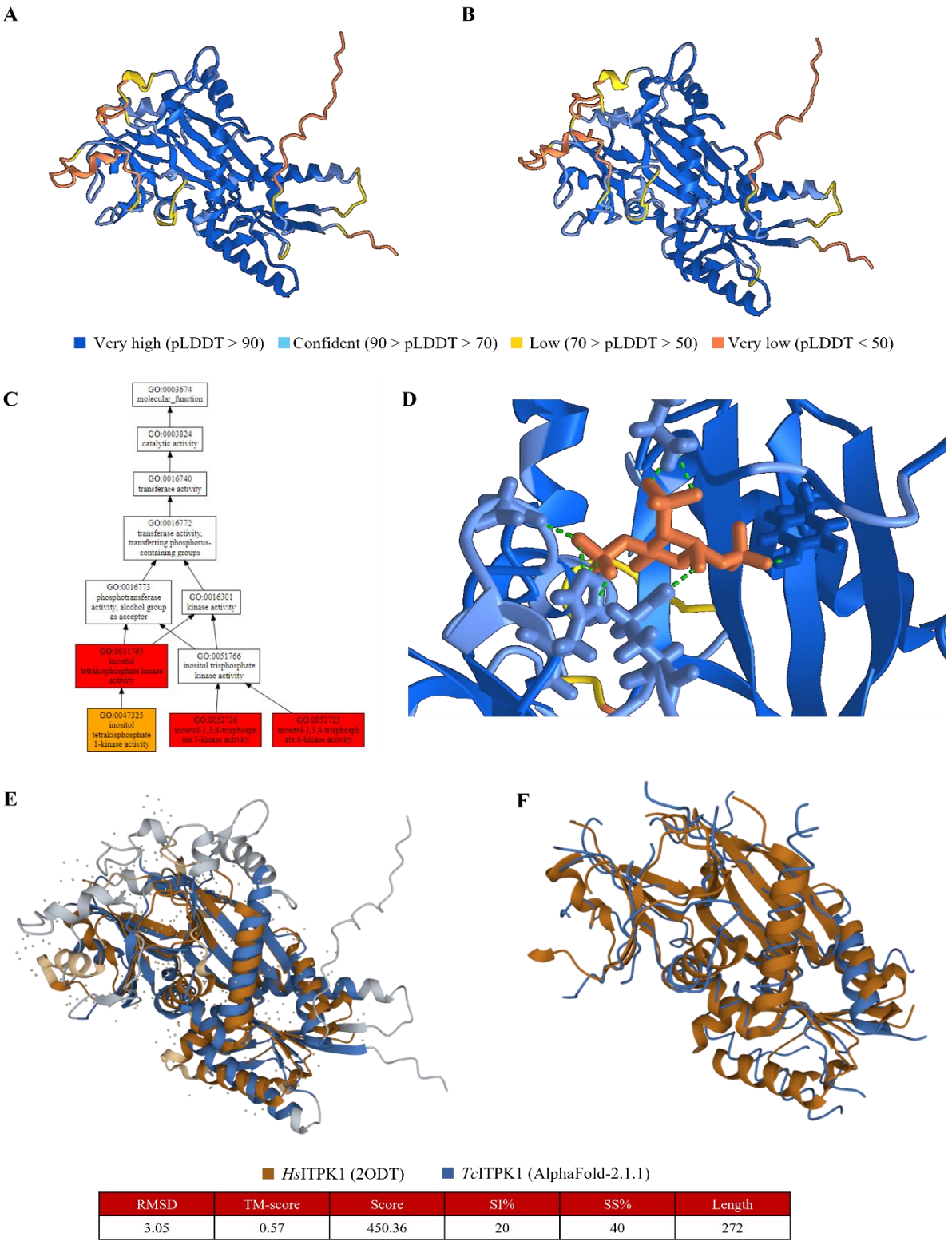
RMSD	TM-score	Score	SI%	SS%	Length
1.76	0.72	905.14	100	100	315

Figure 7. Experimental Validation of AlphaFold-2.1.1 to model ITPK1 proteins. AlphaFold-generated *Hs*ITPK1 PDB structure (relaxed) was compared to an X-Ray diffraction resolved *Hs*ITPK1 (2ODT) structure using the pairwise structure alignment available on RCSB PDB with the resulting structural alignment characterization scores. The resolved structures above show the (**A**) full structural comparison and (**B**) aligned residue comparison.

predicted and experimentally-determined structures of *Hs*ITPK1, the two structures have 315 equivalent positions with a root mean standard deviation (RMSD) of 1.76Å without twists, denoting strong alignment between the pair of structures. The sequence identity percentage (SI%) and sequence similarity percentage (SS%) comparison further demonstrates the high similarities of these two *Hs*ITPK1 structures, both at 100%. The template modeling score (TM-score) of 0.72 indicates that the two proteins have the same protein folding topology, with scores of 0.5 being the threshold of similarity. The two structures are found to be significantly similar with a jFATCAT (flexible) score of 905.14 and a p-value of 0.00e+00.

After the previous validation experiments, AlphaFold-2.1.1 was then utilized to predict the structural confirmation of the 419 aa protein, *Tc*ITPK1. Both the relaxed (**Figure 8A**) and unrelaxed (**Figure 8B**) *Tc*ITPK1 ribbon models demonstrate high structural confidence in the core of the protein near the IP-contact and ATP binding site residues with pLDDT values of greater than 90, denoted in dark blue. There are some lower confidence regions of *Tc*ITPK1 protein structure (denoted by yellow and orange); however, this lesser confidence score may be due to evolutionary diversity outside of the ITPK1's active site.

After three-dimensional resolution of *Tc*ITPK1 by AlphaFold-2.1.1, the PDB file was analyzed by COFACTOR to glean structural-based function predictions for gene ontology (**Figure 8C**) and ligand binding partners and ligand-binding sites (**Figure 8D**)[55,56]. The predicted molecular function and gene ontology (GO) of *Tc*ITPK1 with the activity of other orthologous ITPK1 molecules: inositol-1,3,4-trisphosphate 5-kinase activity (GO:0052726, $Cscore^{GO}$ 1.00), inositol-1,3,4-trisphosphate 6-kinase activity (GO:0052725, $Cscore^{GO}$ 1.00), inositol tetrakisphosphate kinase activity (GO:0051765, $Cscore^{GO}$ 0.92), and inositol tetrakisphosphate 1-kinase activity (GO:0047325, $Cscore^{GO}$ 0.89) (**Figure 8C**). COFACTOR-determined GO term predictions agreed with previously predicted GO terms on TriTrypDB (version 49, https://tritrypdb.org/); however, TriTrypDB analysis also predicted ATP binding (GO:0005524) and magnesium ion binding (GO:0000284) as molecular functions[44,55,56]. In addition to molecular functional predictors, COFACTOR predicted that the molecule D-myo-inositol-1,3,4-trisphosphate would bind with *Tc*ITPK1 at residues K20, T192, G193, H198, K242, Y244, Q277, N274, N374, P377, and G378 (**Figure 8D**).



RMSD	TM-score	Score	SI%	SS%	Length
3.05	0.57	450.36	20	40	272

Figure 8. Utilizing a bioinformatics toolkit to understand *Tc*ITPK1 structural and functional biology.

Structural prediction of *Tc*ITPK1 using AlphaFold-2.1.1 for both (**A**) relaxed and (**B**) unrelaxed ribbon models

generated by iCn3D (NCBI) with AlphaFold Confidence indicator. Employment of COFACTOR structure-based functional prediction to show *Tc*ITPK1 functional insights including (**C**) predicted gene ontology and (**D**) predicted ligand binding site for D-myo-inositol-1,3,4-trisphosphate (I3S) with *Tc*ITPK1 generated by iCn3D (NCBI) with AlphaFold Confidence indicator. Structural comparison of AlphaFold-predicted *Tc*ITPK1 (relaxed) with X-Ray diffraction resolved *Hs*ITPK1 (2ODT) structure using the pairwise structure alignment available on RCSB PDB with the resulting structural alignment characterization scores for (**E**) full structural comparison and (**F**) aligned residue comparison.

The AlphaFold-2.1.1-predicted *Tc*ITPK1 was also structurally compared to the X-ray diffracted *Hs*ITPK1 (2ODT) using the Pairwise Structural Alignment tool for the full structural (**Figure 8E**) and aligned residue (**Figure 8F**) comparison[61–69]. Both ITPK1 proteins had highly conserved secondary structures and minor spatial modulation of alpha-helical structures. When comparing *Tc*ITPK1 with *Hs*ITPK1 (2ODT), the two structures were found to have 272 equivalent positions with a RMSD of 3.05Å without twists, denoting strong alignment between the eukaryotic orthologous ITPK1 structures (**Figure 8**). jFATCAT (flexible) comparison of the two ITPK1 proteins predicted a significantly similar structure with a p-value of 5.76e-12 and a score of 450.36. The 0.57 TM-score demonstrated a similar protein structure between AlphaFold-2.1.1-*Tc*ITPK1 and *Hs*ITPK1 (2ODT), but lesser than the prediction of AlphaFold-2.1.1-*Hs*ITPK1 and *Hs*ITPK1 (2ODT) (TM-score = 0.72, **Figure 7**). The sequence identity and similarity between *Tc*ITPK1 and resolved *Hs*ITPK1 were expectedly lower due to the evolutionary differences in ITPK1 amino acid residue composition with scores of 20% and 40%, respectively.

Discussion

The hypothetical protein TcCLB.503885.50 is predicted to be inositol tetrakisphosphate 1-kinase (ITPK1) in the protozoan parasite *Trypanosoma cruzi*. This chapter uses a bioinformatic toolset to establish the phylogenetic evolution and structural conservation of *Tc*ITPK1. Initial results of Clustal Omega alignment demonstrates that *Tc*ITPK1 shares two out of three highly conserved residues and three out of four lesser conserved residues required for the IP and ATP-binding functionality of the protein. Ligand-binding predictions of *Tc*ITPK1 based on a three-dimensional protein structure determined by AlphaFold-2.1.1 and COFACTOR (I-Tasser_Function) algorithm bolsters this residue conservation, i.e. residues at position K20, H198, K242, and G378 of *Tc*ITPK1 correspond to residues K18, H168, R212, and G301 in *Hs*ITPK1. COFACTOR ligand-binding site predicts an additional seven residues required for D-myo-inositol-1,3,4-trisphosphate binding, but there does not appear to be high conservation between species for these specific residues. Both TriTrypDB and COFACTOR structural-based prediction algorithm predict that *Tc*ITPK1 is involved with inositol-1,3,4-trisphosphate 5-kinase, inositol-1,3,4-trisphosphate 6-kinase, inositol tetrakisphosphate kinase, and inositol tetrakisphosphate 1-kinase activities (GO:0052726, GO:0052725, GO:0051765, and GO:0047325, respectively).

This chapter's phylogenetic study demonstrates that ITPK1 is highly conserved in kinetoplastids (bootstrap values > 80) and other higher order eukaryotes. The phylogenetic tree aimed to study the evolution of all protozoan parasites; however, there was a notable lack of ITPK1 in apicomplexans such as *Toxoplasma gondii* and *Plasmodium* spp. This finding was also noted in a recent review on the IPP pathway in higher order eukaryotes[74]. *T. cruzi* is found is the Discoba supergroup whereas *Homo sapiens* are a member of the Opisthokonta supergroup[75].

While both of these organisms are in different eukaryotic supergroups, we believe that their last eukaryotic common ancestor had this particular lipid-independent, soluble pathway to synthesize inositol polyphosphates from glucose-6-phosphate without the need for PLC. We believe that the absence of this pathway in *Apicomplexans* is a recent loss of that clade and does not reflect the overall evolution of this pathway.

While this bioinformatics toolset provides a strong foundational basis of understanding *Tc*ITPK1, there are limitations to what can be elucidated using bioinformatics. Predictions can only provide scientists with preliminary results, as correlation does not always mean causation. Therefore, laboratory experiments are needed to fully understand the activity of *Tc*ITPK1 in the lipid-independent pathway.

CHAPTER 3:

UTILIZING KALLISTO-SLEUTH TO DETERMINE THE IMPACTS OF ORGANOMETALLIC COMPOUNDS ON THE *TRYPANOSOMA CRUZI* LIPID-INDEPENDENT INOSITOL POLYPHOSPHATE SYNTHESIS PATHWAY

Cline, S.E. and Docampo, R. To be submitted to *Bioinformatics and Biology Insights*.

Abstract

Current standard treatments for Chagas Disease are unsatisfactory, only targeting the acute stages of the parasitic infection and leading to unfavorable side effects. Pd-dppf-mpo and Pt-dppf-mpo are organometallic compounds that have trypanocidal activity and can successfully target both the infective (acute) and intracellular (chronic) forms, but the exact pharmacological mode of action is still unknown. Both compounds have been shown to modulate *TcIP3R*, an essential gene that encodes the *T. cruzi* IP_3 receptor, which plays a key role in inositol polyphosphate pathway (IPP) and calcium signaling. Previous attempts to disrupt *TcIP3R* expression led to significant parasitic defects related to growth, transformation, and infectivity[76]. This chapter focuses on the retroactive analysis of organometallic compounds' modulation of the *T. cruzi* lipid-independent inositol polyphosphate pathway genes *TcITPK1* (TcCLB.611903.250:mRNA), *TcISC1* (TcCLB.511655.50), and *TcINO1* (TcCLB.507609.60:mRNA) using a Kallisto-Sleuth RNA-seq analysis pipeline. While the impact of Pd-dppf-mpo and Pt-dppf-mpo on *TcISC1* still remains to be resolved, gene transcription of *TcINO1* was significantly downregulated in both treatments while *TcITPK1* gene expression was unaffected.

Introduction

For decades, benznidazole and nifurtimox have been the standard antiparasitic treatment options for acute Chagas Disease (CD). In 2020, these oral therapeutics received FDA approval for the treatment of CD in pediatric patients with benznidazole treating adolescents 2-12 years of age and nifurtimox treating individuals younger than 18 years of age (weight $\geq$ 2.5 kg)[10,12,77]. While these therapeutics cause a decrease or clearance of parasitic infection in acutely infected

patient CD IgG serological assays, these treatments have unfavorable side effects. Genotoxicity studies of both medications report a two- and thirteen-fold increase in chromosomal aberrations for benznidazole and nifurtimox, respectively[10,12]. Additionally, common side effects of these medications include - but are not limited to - hypersensitivity responses, decreased appetite and weight loss, acute porphyria attacks, peripheral neuropathy, and embryo-fetal toxicity[10,12]. Approximately 20% of patients undergoing treatment have to stop treatment due to these adverse side effects[11,78,79]. The limitations of these two therapeutic treatments to target the acute phase of CD combined with patient tolerability concerns highlight the importance of continued drug discovery for this tropical, neglected disease.

Recently, two organometallic compounds, 1,1′-bis(diphenylphosphino) ferrocene pyridine-2-thiolate-1-oxide Pd(II) hexafluorophosphate (Pd-dppf-mpo) and 1,1′-bis(diphenylphosphino) ferrocene pyridine-2-thiolate-1-oxide Pt(II) hexafluorophosphate (Pt-dppf-mpo) (**Figure 10A**), are identified to modulate gene transcription, affect *T. cruzi* parasite proliferation, and decrease parasite survival[80–83]. Both of these compounds demonstrate a better IC_{50} than nifurtimox (IC_{50} = 1.8 µM) with IC_{50} values of 0.28 µM and 0.64 µM for the Pt-dppf-mpo and Pd-dppf-mpo compound, respectively[80,81]. However, one important piece of information about these compounds is their capability to target parasite proliferation and survival in the infective and chronic intracellular forms of parasite, potentially providing a lead into drug discovery for chronic Chagas Disease. After treatment with organometallic compounds, *T. cruzi* epimastigotes experience interruptions in molecular pathways involved with DNA binding, protein metabolism, transmembrane transport, oxidative defense, and ergosterol biosynthesis[83].

This chapter is a retroactive bioinformatics study on how organometallic compounds, Pd-dppf-mpo and Pt-dppf-mpo, impact the lipid-independent IPP pathway through the utilization of

previously published sequence read archive (SRA) data available on NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra)[83]. *T. cruzi* epimastigotes were incubated with six times the IC_{50} for each targeted organometallic compound and then underwent RNA-seq analysis to determine modulated transcriptomics. One significantly modulated gene reported in the original manuscript is *TcIP3R* (TcCLB.509461.90), an essential gene related to inositol polyphosphate pathway (IPP). The original data analysis did not delve into understanding the effects of the organometallic compounds on the lipid-independent IPP, instead focusing on understanding the multimodal effects of the novel treatments on parasite biochemical pathways and gene expression profiles[83]. Furthermore, the original manuscript's high-throughput analysis eliminates all fragmented and pseudogenes from their pipeline and excludes discussion on specific hypothetical protein transcripts. As described in the previous chapter, *TcITPK1* (TcCLB.511903.250), a lipid-independent IPP GOI, is noted to be a hypothetical protein in TriTrypDB (version 55, https://tritrypdb.org/) and, therefore, was not analyzed in the original publication[44,82]. *TcINO1* (TcCLB.507609.60) is notated as an inositol-3-phosphate synthase since its annotation in TriTrypDB (version 9.0, released January 14, 2015); however, while it was analyzed in the original dataset, its modulation or lack thereof is not discussed. TriTrypDB annotates *TcISC1* (TcCLB.511655.50) as an inositol phospholipid phospholipase C (pseudogene) and, therefore, was not analyzed in the initial manuscript due to its exclusionary pipeline. The major aim of this chapter is to retroactively determine if lipid-independent IPP genes *TcITPK1* (TcCLB.511903.250), *TcINO1* (TcCLB.507609.60), and *TcISC1* (TcCLB.511655.50) are significantly modulated by either Pd-dppf-mpo or Pt-dppf-mpo treatment using Kallisto-Sleuth RNA-seq analysis techniques.

Materials and Methodology

Mosquillo et al. 2020 Preparation of RNA for Transcriptomic Assays and Data Analysis[83]: The original study utilized CL Brener Non-Esmeraldo *T. cruzi* epimastigotes cultured in fresh BHI medium supplemented with 10% Fetal Bovine Serum (FBS) and antibiotics penicillin (100 units mL^{-1}) and streptomycin (100 µg mL^{-1}) and maintained at 28°C in the exponential growth phase. Organometallic compounds Pd-dppf-mpo and Pt-dppf-mpo were added to the parasitic culture to a final concentration of 1.5 µM and 0.3 µM corresponding to five times their experimental IC_{50}, respectively, and incubated for 6 hours. Both untreated (WT) and treated (Pd-dppf-mpo or Pt-dppf-mpo) epimastigotes underwent Trizol Reagent RNA extraction (Life Technologies) and RNA cleanup using a DNA-free kit (Life Technologies) according to manufacture guidelines. Each condition had three independent replicates. Isolated RNA samples were then prepared with the Illumina TruSeq™ RNA Sample Preparation Kit v2 and HiSeq 2500 for PoliA + RNA paired-end sequencing. The original RNA-seq results were analyzed using Bowtie2 and HTSeq/DESeq2 pipeline[84–86].

Transcriptomic Analyses: RNA-seq data for wild type (WT) control epimastigotes (SRX7822256-SRX7822258), Pd-dppf-mpo treated epimastigotes (SRX7822259-SRX7822261), and Pt-dppf-mpo treated epimastigotes (SRX7822262-SRX7822264) is available on NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) and was obtained using the SRA-Toolkit (SRA-Toolkit/2.9.6-1-centos_linux64)[87]. CL Brener Non-Esmeraldo *Trypanosoma cruzi* annotated CDS fasta file was obtained from TriTrypDB (version 55, https://tritrypdb.org/)[44]. This CDS fasta file was indexed for quantification using Kallisto (kallisto/0.46.1-foss-2019b)[88]. Kallisto was then used for the quantification of each RNA-seq transcriptomic experiment using

100 bootstrap replicates. Differential expression analysis was performed with Sleuth under a variety of conditions[89]. The top ten modulated transcripts of all conditions were obtained with Sleuth using the likelihood ratio test (lrt) with a cutoff *q*-value of 0.05. The top ten modulated transcripts for each specific condition, i.e. WT control versus Pd-dppf-mpo-treated epimastigotes or WT control versus Pt-dppf-mpo-treated epimastigotes, were obtained by Sleuth differential expression analysis using the Wald test with a cutoff *q*-value of 0.05. To see transcripts of interest, I utilized the WT control versus Pd-dppf-mpo-treated epimastigotes Sleuth differential analysis using Wald test with a cutoff *q*-value of 0.05 and called upon the individual transcripts using their annotated name, i.e. TcCLB.509461.90:mrNA (*TcIP3R*), TcCLB.611903.250:mRNA (*TcITPK1*), and TcCLB.507609.60:mRNA (*TcINO1*). Unfortunately, due to *TcISC1* (TcCLB.511655.50:mRNA) being a pseudogene, quantification and differential analysis was not able to be conducted due to repeated computational errors in the Sapelo2 cluster. The initial manuscript also excluded pseudogenes from their analysis.

- Final BASH script is available at https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/project.sh. Git revision for this BASH code is 0a5796a9e836493b06bc9522904bcf104d1d98ed.
- Final R script is available at https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/projectRscript.r. Git revision for this code is 0a5796a9e836493b06bc9522904bcf104d1d98ed.

Pd-dppf-mpo		Pt- dppf-mpo	
Gene ID	**Protein**	**Gene ID**	**Protein**
TcCLB.506201.39	**25 kDa translation elongation factor 1-beta, putative**	**TcCLB.511899.40**	**2-amino-3-ketobutyrate coenzyme A ligase, putative**
TcCLB.506789.280	**methylthioadenosine phosphorylase, putative**	**TcCLB.505183.30**	**malic enzyme, putative**
TcCLB.510445.50	S-adenosylmethionine synthetase, putative	**TcCLB.510947.50**	**cyclophilin (rotamase), putative**
TcCLB.510725.40	hypothetical protein, conserved (TMEM14 family)	**TcCLB.506193.60**	**ascorbate peroxidase**
TcCLB.510943.50	**delta-1-pyrroline-5-carboxylate dehydrogenase**	TcCLB.507649.70	nucleolar RNA-binding protein, putative
TcCLB.511229.50	S-adenosylhomocysteine hydrolase, putative	TcCLB.511025.70	chaperonin alpha subunit, putative
TcCLB.511289.70	mitochondrial ADP/ATP carrier protein 5, putative	TcCLB.508175.9	hypothetical protein, conserved (fragment)
TcCLB.511903.40	3-hydroxy-3-methylglutaryl-CoA synthase, putative	TcCLB.510445.50	S-adenosylmethionine synthetase, putative
TcCLB.509065.60	glyceraldehyde 3-phosphate dehydrogenase, putative	**TcCLB.507913.20**	**phosphomevalonate kinase protein, putative**
TcCLB.509065.70	glyceraldehyde 3-phosphate dehydrogenase, putative	**TcCLB.507583.20**	**nucleolar protein, putative**

Table 1. **Top ten modulated transcripts for Pd-dppf-mpo and Pt-dppf-mpo treatments as analyzed in the Kallisto-Sleuth pipeline.** All gene IDs and their associated protein names in bold are reported to be significantly modulated from the original manuscript published by the Péréz-Díaz group in 2020[83].

Results

After analysis through the Kallisto-Sleuth pipeline, the top ten significantly modulated transcripts for each treatment group, Pd-dppf-mpo and Pt-dppf-mpo, were reported (**Table 1**). Both Pd-dppf-mpo and Pt-dppf-mpo treatments were shown to highly modulate transcripts associated with protein translation and conformational folding (GO:0006414, GO:0006457, GO:0000413), ergosterol biosynthesis pathway (GO:0006696), and S-adenosylmethionine biosynthetic and one-carbon metabolic processes (GO:0006556, GO:0006730). Top Pd-dppf-mpo-modulated transcripts were also associated with proline catabolic activity (GO:0006562) and glucose metabolic processes (GO:0006006); whereas top Pt-dppf-mpo-modulated transcripts were associated with RNA metabolic processes and post-transcriptional modifications (GO:0016070, GO:0001510), energy metabolism (GO:0006108), threonine amino acid catabolism (GO:0006567), and responses to oxidative stress (GO:0006979). These results agreed

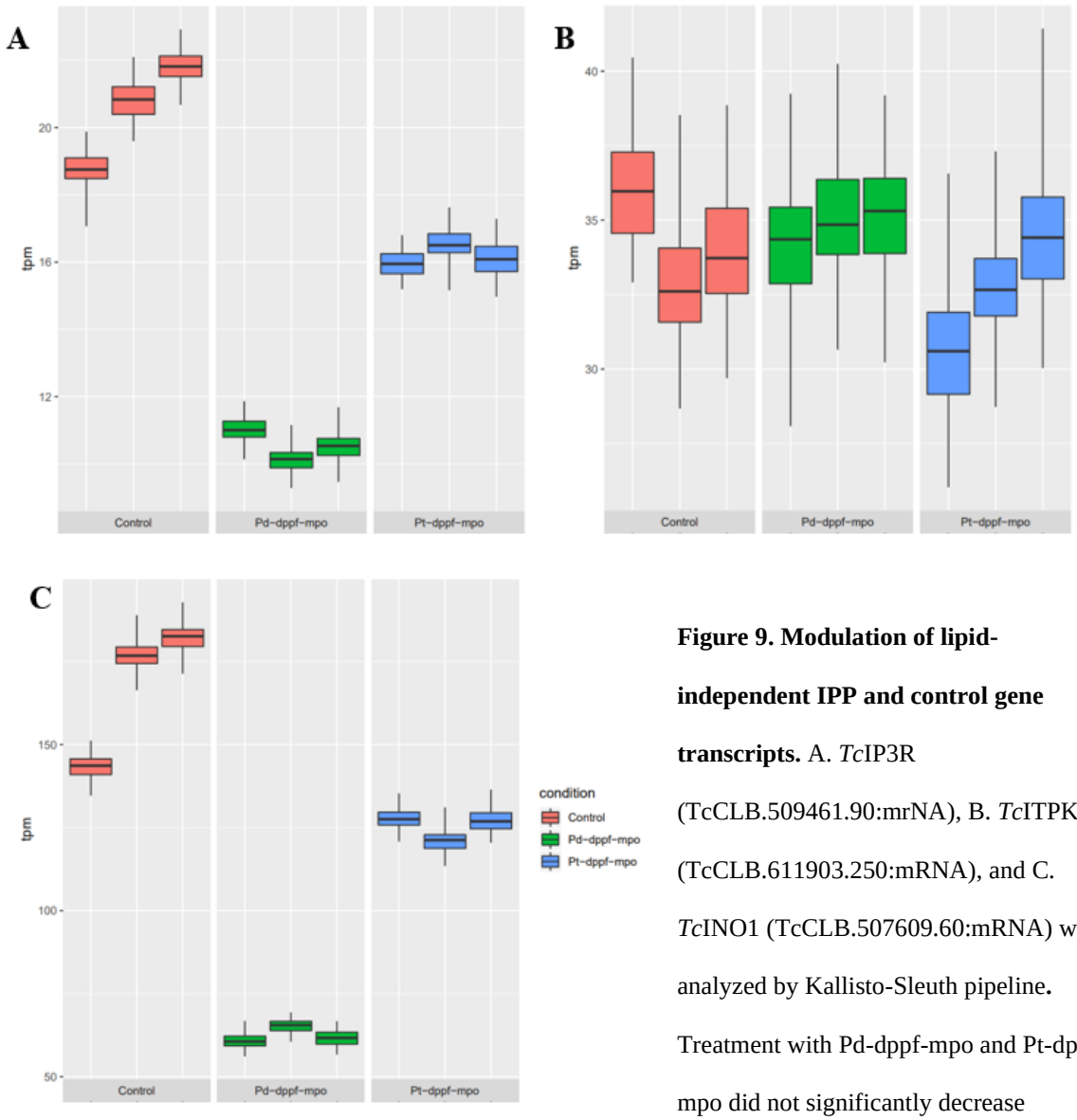


Figure 9. Modulation of lipid-independent IPP and control gene transcripts. A. *Tc*IP3R (TcCLB.509461.90:mRNA), B. *Tc*ITPK1 (TcCLB.611903.250:mRNA), and C. *Tc*INO1 (TcCLB.507609.60:mRNA) were analyzed by Kallisto-Sleuth pipeline**.** Treatment with Pd-dppf-mpo and Pt-dppf-mpo did not significantly decrease *TcITPK1* transcription but did have a significant impact on *TcINO1* transcription. Results for *TcISC1* are not shown due to status as pseudogene. Treatment conditions are denoted in red, green, and blue for control, Pd-dppf-mpo, and Pt-dppf-mpo, respectively.

with the findings of the original manuscript, with three and six of the top ten transcripts matching for Pd-dppf-mpo and Pt-dppf-mpo, respectively (**Table 1**)[83]. The IPP pathway, inositol metabolism, or other associated pathways were not noted in this table; however, this absence

does not imply that the lipid-independent IPP pathway was not significantly modulated by organometallic compound treatments.

The three GOIs related to the lipid-independent IPP - *TcITPK1, TcINO1,* and *TcISC1* - and the inclusion of a positive control from the original transcript, *TcIP3R* were analyzed individually using the Wald test with a cutoff *q*-value of 0.05. The Kallisto-Sleuth pipeline demonstrated that, similar to the original manuscript, *TcIP3R* was significantly modulated in both treatment groups (**Figure 9A**). Lipid-independent IPP *TcITPK1* did not have significant down or upregulation of gene expression when exposed to either treatment groups (**Figure 9B**). However, *TcINO1* did have significant modulation of gene transcription when exposed to either Pd-dppf-mpo or Pt-dppf-mpo treatment (**Figure 9C**). Transcription of *TcINO1* decreased from approximately 160 transcripts per million (tpm) to 60 tpm and 125 tpm after Pd-dppf-mpo and Pt-dppf-mpo treatment, respectively. Treatment with Pd-dppf-mpo was more successful than Pt-dppf-mpo in downregulating the expression of both *TcIP3R* and *TcINO1* (**Figure 9A, 9C**). Unfortunately, there were issues related to the created Kallisto-Sleuth pipeline for the *TcISC1* pseudogene that did not allow for differential expression for this GOI. The created bioinformatics analysis pipeline can handle annotated and hypothetical genes; however, it is incapable of running pseudogene analysis. Therefore, the *TcISC1* was excluded from the final analysis.

All resulting output pdf files are also available on the BINF8940 GitHub Repository (git revision 0a5796a9e836493b06bc9522904bcf104d1d98ed) at the following URL locations:

- Kallisto-Sleuth Pipeline (LTR) for top ten transcripts: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/ProjectSleuthControlResults.pdf
- Kallisto-Sleuth Pipeline (Wald, q-value = 0.05) for top ten Pd-dppf-mpo transcripts: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/Pd-dppf-mpo_Results.pdf
- Kallisto-Sleuth Pipeline (Wald, q-value = 0.05) for top ten Pt-dppf-mpo transcripts: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/Pt-dppf-mpo_Results.pdf
- Kallisto-Sleuth Pipeline for *TcITPK1*: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/TcCLB.511903.250mRNA.pdf
- Kallisto-Sleuth Pipeline for *TcINO1*: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/TcCLB.507609.60mRNA.pdf
- Kallisto-Sleuth Pipeline for *TcIP3R*: https://github.com/sabrinaecline/Thesis/blob/0a5796a9e836493b06bc9522904bcf104d1d98ed/Chapter%203/TcCLB.509461.90mRNA.pdf

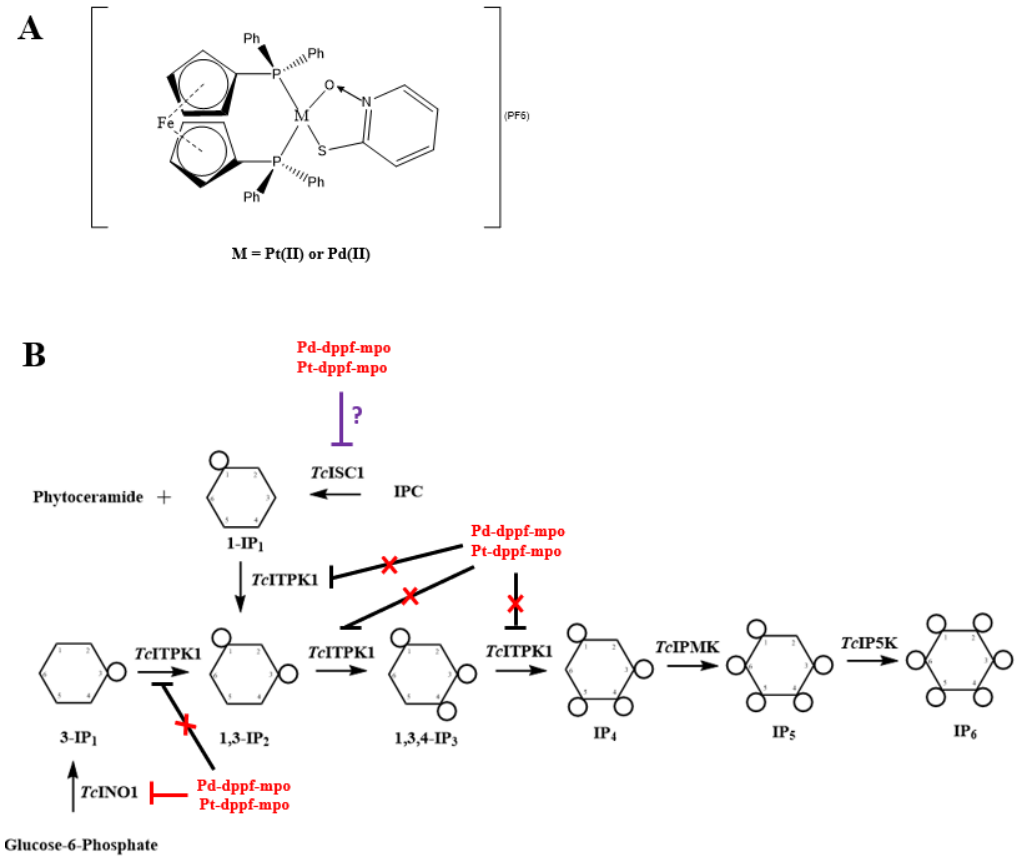


Figure 10. **Effects of Organometallics on the *T. cruzi* Alternative Inositol Polyphosphate Synthesis Pathway.** A. Schematic representation of Pd-dppf-mpo and Pt-dppf-mpo organometallic compounds. B. Impact of organometallic compound treatment on the *T. cruzi* lipid-independent IPP synthesis pathway. Organometallic compounds significantly downregulate *TcINO1* gene expression but do not significantly affect *TcITPK1* gene expression. Affect of compounds on *TcISC1* is still unknown.

Discussion

The use of bioinformatics to re-examine previously published datasets is an exciting new approach to mine for novel findings and preliminary results without heavy time and cost burden of traditional experimentation. In the original analysis of this dataset, the authors demonstrated an interruption of a variety of biosynthetic and metabolic pathways including those involved

with protein metabolism, transmembrane transport, oxidative defense, and ergosterol biosynthesis[83]. The authors report that Pt-dppf-mpo was able to significantly modulate 201/10773 genes detected whereas the Pd-dppf-mpo compound was able to modulate 2327/10785 genes detected (with significant fold changes being greater than |1.5| and p-values adjusted for a FDR of 0.01)[83]. As a built in control for the project, the top ten transcripts from each treatment group and compared these to the original results (**Table 1**). There is some agreement with the original manuscript for significantly modulated genes, particularly with the Pt-dppf-mpo compound; however, the authors used TriTrypDB (version 29, https://tritrypdb.org/)[44] and filtered out any hypothetical proteins from their analysis and result pipeline[83]. The gene annotation for CL Brener Non-Esmeraldo *Trypanosoma cruzi* in TriTrypDB has been updated multiple times so some of these non-bolded proteins may have previously been reported as hypothetical and subsequently filtered out of the original manuscript. This project still reports two hypothetical proteins - TcCLB.510725.40 and TcCLB.508175.9 - as highly modulated transcripts when epimastigotes are treated with Pd-dppf-mpo and Pt-dppf-mpo, respectively.

Other variations may be a result of differences in quantification and differential expression analysis pipelines. The original manuscript employed Bowtie2[86] in very-sensitive mode with quantification determined by HTSeq-count[85] and DESeq2[84] to determine differentially expressed genes between experimental and control groups whereas this manuscript employed a Kallisto-Sleuth pipeline. Bowtie2 is a comparative genomic tool that aligns sequencing reads to long reference sequences and has been a standard in ChIP-seq, RNA-seq, and BS-seq analyses for the past decade. After short read alignment to the genome, Bowtie2-generated SAM files are analyzed through HTSeq count for quantification of the high-throughput

sequencing data and then Deseq2 for differential gene expression analysis. However, as the bioinformatics community has improved its toolset, Kallisto-Sleuth pipeline has emerged as a new standard for quantification and differential analysis. This pipeline skips the need for sequence alignment, instead focusing on utilizing pseudoalignment of reads based on k-compatibility class with the transcriptome de Bruijin graph for quantifying the abundance of transcripts[88]. These Kallisto index output files are then committed to Sleuth which allows for the differential gene analysis and comparison of multiple RNA-seq experiments.[89] Since there were three experimental conditions with three independent replicates, the Kallisto-Sleuth pipeline offered a fast, highly accurate, and efficient method to retroactively analyze the multi-sample dataset and identify transcripts of interest.

The original manuscript of this project's RNA-seq dataset does not delve into how treatment of *T. cruzi* epimastigotes with organometallic compounds impacted the lipid-independent IPP synthesis pathway. This chapter illustrates that, while gene expression for *TcITPK1* was not modulated under either treatment condition, *TcINO1* expression was highly modulated with treatments by both Pd-dppf-mpo and Pt-dppf-mpo (**Figure 10B**). Additionally, organometallic compound treatment of epimastigotes did show significant modulation of glyceraldehyde-3-phosphate dehydrogenase (TcCLB.509065.60 and TcCLB 509065.70) genes, which encode a key enzyme in the glycolysis and gluconeogenesis pathways, and malic enzyme (TcCLB.505183.30) gene, which encodes a key enzyme that enhances the pentose phosphate pathway (**Table 1**)[90]. Since *Tc*INO1 utilizes glucose-6-phosphate to synthesize inositol-1-monophosphate, the downregulation might be as a result of dysregulation of those metabolic pathways. Further refinement of the Kallisto-Sleuth bioinformatic pipeline to include pseudogene analysis would allow for the impact of organometallic compounds on *TcISC1* and

the lipid-independent IPP pathway to be resolved (**Figure 10B**). Both *Tc*INO1 and *Tc*ISC1 proteins are hypothesized to provide cytosolic inositol monophosphate isomers to the start of the lipid-independent pathway and so it would be interesting to see if the *TcISC1* transcript was also modulated in a similar fashion to *TcINO1*.

While the impact of Pd-dppf-mpo and Pt-dppf-mpo compounds on the lipid-independent IPP pathway is still not fully resolved, these compounds show immense potential for drug discovery for the treatment of *T. cruzi*. Significant modulation of both *TcINO1* and *TcIP$_3$R* by Pd-dppf-mpo implicates inositol polyphosphate synthesis as a potential druggable target. Previous research has demonstrated that *Tc*IP$_3$R is essential for *T. cruzi* life cycle with genetic disruptions leading to the death of the parasite. The role of INO1 in *T. cruzi* still remains to be resolved; however, *T. brucei* bloodstream forms show that *TbINO1* is an essential gene in conditional knockouts[76,83,91]. Therefore, this study demonstrates that these organometallic compounds can successfully downregulate kinetoplastid essential genes in the inositol phosphate synthesis and regulatory pathways.

CHAPTER 4:

EMPLOYING CRISPR/CAS9 STRATEGIES TO STUDY *Tc*ITPK1 *IN VITRO*

Cline, S.E., Chiurillo, M.A., and Docampo, R. To be submitted to *Frontiers in Cellular and Infection Microbiology*.

Abstract

While there is a diverse genetic toolset available for eukaryotic organisms, the protozoan parasite *Trypanosoma cruzi* poses special limitations due to its unique biological properties. The development and species-specific optimization of CRISPR/Cas9 technologies has rapidly advanced the understanding of *T. cruzi* cellular biology in targeted genomic studies. *Tc*ITPK1 is hypothesized to be the mediator of the lipid-independent inositol polyphosphate (IPP) synthesis pathway in *T. cruzi*. Utilizing CRISPR/Cas9 strategies such as endogenous C-terminal tagging and ablation of *TcITPK1* gleans a deeper understanding of this enzyme's biological role. Immunofluorescence assays have shown *TcITPK1* to localize to the cytosol, the site of the lipid-independent IPP synthesis pathway. Two published knockout strategies failed to produce viable *TcITPK1*-KO parasites after antibiotic selection, suggesting that *TcITPK1* is essential for parasite survival. A genetic knockdown of *Tc*ITPK1 using the CRISPR/Cas9 riboswitch (*glmS/M9*) method has been developed and verified, but further experiments are required to study the impacts of *TcITPK1* ablation *in vitro*.

Introduction

Genetic manipulation in the protozoan parasite, *Trypanosoma cruzi*, has always had its unique set of challenges. RNA interference (RNAi) methodology is a functional genomics technique utilized for protozoan parasites, namely in *Trypanosoma brucei*; however, other closely related kinetoplastids *Leishmania major* and *T. cruzi* lack the necessary RNAi cellular machinery to allow for the dsRNA-mediated downregulation of gene expression[92]. A tetracycline-inducible expression vector pTcINDEX allowed for the modulation of gene expression; however, this strategy was noted to have issues related to expression leakage in the

absence of tetracycline[93,94]. The discovery and optimization of CRISPR/Cas9 genomic editing has revolutionized gene editing and improved the genetic toolset available to study *T. cruzi*. CRISPR/Cas9 strategies allow for highly efficient, targeted genomic editing by utilizing a Cas9 nuclease and a small site-specific, guide RNA to modulate eukaryotic genomic DNA[95]. CRISPR/Cas9 methodologies have been developed to allow for C-terminal endogenous tagging, overexpression, knockout, and knockdown of genes in *T. cruzi*, rapidly expanding the genomic toolset that we can use to study these biologically relevant parasites[96–100].

As discussed in Chapter 1, phosphoinositide-specific phospholipase C (PI-PLC) is an important enzyme in the lipid-dependent inositol polyphosphate (IPP) synthesis pathway. Original strategies to knockout *TcPI-PLC* with neomycin resistance (*neo*) gene with and without the addition of a glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) gene to enhance resistance gene expression failed to produce viable parasites[38]. The introduction of CRISPR/Cas9-based methodology allowed for successful *TcPI-PLC* knockout and generated viable epimastigotes *in vitro* (Miguel Chiurillo and Roberto Docampo, unpublished data). This successful knockout demonstrated that ablation of *TcPI-PLC* significantly impacted host cell invasion and intracellular replication (**Chapter 1, Figure 3B, 3C**). Unexpectedly, there was no significant impact on IPP synthesis after *TcPI-PLC* knockout (**Chapter 1, Figure 3D**). This finding demonstrates that *Tc*PI-PLC is not essential for IPP synthesis and the existence of an alternative pathway.

Inositol tetrakisphosphate 1-kinase (ITPK1) is the mediator of the lipid-independent IPP synthesis pathway in higher-order eukaryotes[1]. Phylogenetic analysis identified TcCLB.503885.50 as the eukaryotic ITPK1 ortholog in *T. cruzi*. *Tc*ITPK1 which is hypothesized to facilitate this alternative IPP synthesis pathway. The biological impacts of *TcITPK1* ablation

and subcellular localization are of key interest to understanding the lipid-independent IPP synthesis pathway in this protozoan parasite.

In this chapter, we detail our attempts to utilize CRISPR/Cas9 strategies to modify *Tc*ITPK1: (1) endogenously C-terminal tagged *Tc*ITPK1, (2) *TcITPK1* knockout with *HX1-Bsd-gapdh* cassette, (3) *TcITPK1* knockout with *Bsd* cassette, and (4) riboswitch (*glmS/M9*) glucosamine-induced *TcITPK1* knockdown. Endogenous C-terminal tagging of *TcITPK1* is of importance for *in vitro* subcellular localization studies and will illuminate whether the protein is membrane-bound as predicted by Protter[72] and InterPro[73] or if it is found in the cytosol as expected in higher-order eukaryotic lipid-independent IPP synthesis. Ablation of *Tc*ITPK1 will highlight its role in *T. cruzi* cellular processes, most importantly its role in IP_{6-8} synthesis.

Materials and Methodology

T. cruzi Culture Methods: *T. cruzi* (Y strain) epimastigotes were maintained at 28°C in LIT media supplemented with 10% fetal bovine serum (FBS), penicillin (100 U mL^{-1}), and streptomycin (100 μg mL^{-1}). After verification, CRISPR/Cas9 transfected epimastigotes were also cultured in LIT media supplemented with 10% FBS, penicillin (100 U mL^{-1}), and streptomycin (100 μg mL^{-1}) with the addition of the following transfection-specific antibiotics. Classic *Bsd* and *HX1-Bsd-gapdh TcITPK1* knockout transfected epimastigotes received antibiotics geneticin (G418) and blasticidin to a final concentration of 250 μg mL^{-1} and 10 μg mL^{-1}, respectively. Riboswitch (*glmS/M9*) *TcITPK1* knockdown transfected epimastigotes received antibiotics geneticin (G418) and hygromycin to a final concentration of 250 μg mL^{-1} and 150 μg mL^{-1}, respectively. Endogenously C-tagged *TcITPK1* transfected epimastigotes received antibiotics geneticin (G418) and puromycin to a final concentration of 250 μg mL^{-1} and

5 μg mL^{-1}, respectively. SCRAMBLE transfected epimastigotes received antibiotics geneticin (G418) to a final concentration of 250 μg mL^{-1}.

CRISPR/Cas9 TcITPK1 Endogenous C-terminal Tagging: Targeted sgRNAs for endogenous C-terminal tagging (3'-end) of *TcITPK1* were amplified in a one-step PCR reaction using a forward oligonucleotide primer (**Table 2**, Primer 9), universal reverse primer (**Table 2**, Primer 1), and pUC_sgRNA plasmid as a template. This PCR step was followed by PCR purification. *Tc*ITPK1-specific sgRNA was cloned into the Cas9/pTREX-n vector (Addgene plasmid #68708) (Lander *et al., 2015*) and alignment was verified by Sanger sequencing (**Table 2**, Primer 13). Donor DNA was synthesized for endogenous C-terminal tagging in a one-step PCR reaction using the pMOTag23M DNA vector as the DNA template and 75 nucleotide-long ultramers (**Table 2**, Primer 10-11). After visualization of donor DNA on an agarose gel, DNA was purified by phenol/chloroform/isoamyl alcohol (25:24:1) extraction and quantified by nanodrop. Transfection methodology was completed as detailed below.

CRISPR/Cas9 TcITPK1 Classical and HX1-Bsd-gapdh Knockout: Targeted sgRNAs for *TcITPK1* KO were amplified by PCR in a one-step PCR reaction using two designs of forward oligonucleotide primers (**Table 2**, Primer 2-3), a universal reverse primer (**Table 2**, Primer 1), and pUC_sgRNA plasmid as a template. This PCR step was followed by PCR purification. *TcITPK1*-specific sgRNA was cloned into the Cas9/pTREX-n vector (Addgene plasmid #68708) and alignment was verified by Sanger sequencing (**Table 2**, Primer 8)[98]. Donor DNA was synthesized for KO reactions in a two-step PCR reaction using the classic Bsd cassette or HX1-Bsd-gapdh cassette as the DNA template and 100 nucleotide-long ultramers (**Table 2**, Primer 10-

11). After visualization of donor DNA on an agarose gel, DNA was purified by phenol/chloroform/isoamyl alcohol (25:24:1) extraction and quantified by nanodrop. Transfection methodology was completed as detailed below.

CRISPR/Cas9 TcITPK1 Riboswitch Knockdown: Targeted sgRNAs for *Tc*ITPK1 knockdown were amplified a one-step PCR reaction using a forward oligonucleotide primer (**Table 2**, Primer 9), universal reverse primer (Primer 1), and pUC_sgRNA plasmid as a template. This PCR step was followed by PCR purification. *Tc*ITPK1-specific sgRNA was cloned into the Cas9/pTREX-n vector (Addgene plasmid #68708) and alignment was verified by Sanger Sequencing (**Table 2**, Primer 13)[98]. Donor DNA was synthesized for knockdown cassettes using either pMOTag-*glmS*-4H and pMOTag-*M9*-4H vectors (Addgene plasmids #106378 and #106379) as the DNA templates and 75 nucleotide-long ultramers (**Table 2**, Primers 10-11)[96,101]. After visualization of donor DNA on an agarose gel, DNA was purified by phenol/chloroform/isoamyl alcohol (25:24:1) extraction and quantified by nanodrop. Transfection methodology was completed as detailed below.

Cell Transfection: *T. cruzi* epimastigotes (Y strain) grown to a density of $1\text{-}2\cdot10^7$ epimastigotes mL^{-1} were used for transfection of various plasmid and linear DNA donor constructs as detailed above. Parasite cultures were washed once with cold PBS, pH 7.4 and resuspended in ice-cold Cytomix (120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4, 2 mM EDTA, 5 mM $MgCl_2$, pH 7.6) to a density of $1\cdot10^8$ epimastigotes mL^{-1}. Then, $0.4\cdot10^8$ epimastigotes (400 µL) were added to a 4-mm cuvette with 25 µg donor DNA and 25 µg plasmid. Epimastigotes were pulsed at 1.5 kV and 25 µF three times with the Bio-Rad Gene

Pulse Xcell Electroporator, allowing for cells to recover for at least one minute on ice between pulses. Pulsed epimastigotes were incubated for 15 minutes at RT before adding them to 5 mL of LIT media supplement with 20% fetal bovine serum, penicillin, and streptomycin and incubated at 28°C. After 24 hours, antibiotics for specific constructs were added to allow for selection of transfected parasites. Classic and HX1-Bsd-gapdh *TcITPK1* knockout cassettes received geneticin (G418) and blasticidin to a final concentration of 250 μg mL^{-1} and 10 μg mL^{-1}, respectively. Riboswitch (*glmS*/*M9*) *TcITPK1* knockdown cassettes received geneticin (G418) and hygromycin to a final concentration of 250 μg mL^{-1} and 150 μg mL^{-1}, respectively. Endogenously C-tagged *TcITPK1* cassettes received geneticin (G418) and puromycin to a final concentration of 250 μg mL^{-1} and 5 μg mL^{-1}, respectively. After PCR verification of construct incorporation (Primers 6-8,12-13), epimastigotes were cultured in 5 mL of LIT media supplement with 10% fetal bovine serum, penicillin, and streptomycin with the addition construct-specific antibiotics at 28°C.

Western Blot Analyses: Transfected *T. cruzi* epimastigote cultures ($1\text{-}2 \cdot 10^7$ epimastigotes/mL) were spun down and washed twice with cold 1X PBS, pH 7.4. The pellets were then incubated in a microcentrifuge tube with 100 μL RIPA Buffer while being agitated at 4°C for 60 minutes. After lysis, microcentrifuge tubes were spun down at 13.2k rpm at 4°C. Supernatant was collected into a fresh microcentrifuge tube and put on ice. Pellet was discarded. Protein concentrations were standardized using a Bradford assay with BSA standard curve. Approximately 5-10 μg of each lysate was mixed with 4x Laemmli sample buffer and incubated at 95°C for 5 minutes. Extracted proteins were then subjected to electrophoresis using a 12%

Primer Number	Primer Name	Sequence (5'→3')
1	Common sgRNA Reverse Primer	CAGTGGATCCAAAAAAGCACCGACTCGGTG
2	TcITPK1 sgRNA Forward Knockout #1	GATCGGATCCTGTGTCTTGTCGAAGATG GGGTTTTAGAGCTAGAAATAGC
3	TcITPK1 sgRNA Forward Knockout #2	GATCGGATCCCAGTTCCTCACTCTTGGCT GGTTTTAGAGCTAGAAATAGC
4	TcITPK1 Knockout 100 nt Forward Ultramer	TCTGTGTGAAGCGTCCCTTCTGGTATTTTG TTGTTGTTGTTTTTGTTTGGTTTTTCCCACT TCCCAGCGTGGTGCTGAAGAGGAGAAGCA AACTCTGTACCCCAACGAGTTTCTTCAAAA TATGC
5	TcITPK1 Knockout 100 nt Reverse Ultramer	ACACACACACACACAGAAAAATGCCTAAAT GTTTCGCCCTTCATCTCCACTTTTAGCAGCAG AAGTTGCCGTTAAACAGTTTTCCCTCAGTTG AGGATCTATTGGCTGCAGGGTCGCTC
6	TcITPK1 Knockout Forward Check 1	CAGAAGTCTTATTGTCTACTTCCC
7	TcITPK1 Knockout Forward Check 2	ATGGCCAAGCCTTTGTCTCAAG
8	TcITPK1 Knockout Reverse Check	TTTACGGCAGTAGATGCATTTG
9	TcITPK1 C-Tag/Riboswitch sgRNA	GATCGGATCCAGATGAAGGGCGAAACATTTG TTTTAGAGCTAGAAATAGC
10	TcITPK1 C-Tag/Riboswitch 75 nt Forward Ultramer	TCAAAGGCACTGGGTACTCCCGTCTCGAAAAG ATCCTCAACTGAGGGAAAACTGTTTAACGGCA ACTTCTGCTGCGGTACCGGGCCCCCCCTCGAG
11	TcITPK1 C-Tag/Riboswitch 75 nt Reverse Ultramer	CAGTAGATGCATTTGCTCACACGCATAGAGC ATCTTAAAACACACACAAACACACACACACA CAGAAAAATGCCTTGGCGGCCGCTCTAGAAC TAGTGGAT
12	TcITPK1 C-Tag/Riboswitch Forward Check	TGAGCATGTGTTGGATGTCATTAAG
13	TcITPK1 C-Tag/Riboswitch Reverse Check	GTAGAGATATAACACGGCGTTTACG
14	TcITPK1 RT PCR Forward	ACTGCTCCAAGAGGGAAGC
15	TcITPK1 RT PCR Reverse	AAAAGGTGACATGAGCGCC
16	Tubulin RT PCR Forward	GAGGGCATGGACGAGATG
17	Tubulin RT PCR Reverse	CTCCTCCTCGTCGAACTCA

Table 2. Primers Required for *TcITPK1* CRISPR/Cas9 and Related Experiments.

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) at 50 V for 30 minutes, 110V until marker goes to the end of gel (60-120 minutes). BLUelf Prestained Protein Ladder was used to calculate molecular weights. Electrophoresed proteins were transferred to a nitrocellulose membrane using a Bio-Rad transblot apparatus for 5 hours at 30 V at 4°C. Following transfer, membrane blots were blocked with 5% nonfat milk in PBS-T at room temperature for 60 minutes. Blots were probed with mouse anti-c-Myc monoclonal antibody (1:100 dilution in PSB-T), mouse anti-HA monoclonal antibody (1:5,000 dilution in PBS-T), or mouse or rabbit anti-

tubulin monoclonal antibody (1:10,000 dilution in PBS-T) at room temperature for 60 minutes. Blots were washed with PBS-T for 5 minutes, three times. Then, the secondary antibody of either horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG antibody (1:10,000 dilution in PSB-T), conjugated goat anti-rabbit antibody (1:20,000), or conjugated goat anti-mouse antibody (1:20,000) was incubated for 60 minutes, protected from light. Additionally, HRP-conjugated membranes (only) were processed with SuperSignal™ West Pico chemiluminescent substrate according to manufacturer's guidelines. Membranes were then imaged and analyzed using the Odyssey infrared system software (LICOR Biosciences) or a ChemiDoc™ Imaging System (Bio-Rad).

Immunofluorescence Assays: *T. cruzi* epimastigotes transfected with endogenous C-tag construct were grown up to 1-2·10^7 epimastigotes/mL (log phase) and were washed once with 1X PBS, pH 7.4 at room temperature and fixed with 4% paraformaldehyde in PBS for 30 minutes at room temperature. Parasites were allowed to adhere to poly-L-lysine-coated coverslips for 30 minutes and washed three times with 1X PBS, pH 7.4. Coverslips were then permeabilized with 0.3% Triton X-100 for 3 minutes and washed twice with 1X PBS (pH 7.4). Permeabilized epimastigotes were then blocked overnight at 4°C in blocking media (5% goat serum, 50 mM NH_4Cl, 3% bovine serum albumin (BSA), and 1% fish gelatin in 1X PBS, pH 7.4). After blocking, primary antibody of mouse anti-c-Myc monoclonal antibody (1:10 dilution, 1% BSA in 1X PBS, pH 8.0) for *TcITPK1*-C-Tagged epimastigotes was added to the coverslip and incubate for 60 minutes at room temperature. Washed coverslip with 1% BSA in 1X PBS, pH 8.0 three times. Added secondary antibody Alexa Fluor 488-conjugated goat anti-mouse antibody (1:1,000 dilution, 1% BSA in 1X PBS, pH 8.0) to coverslip and incubated for 60

minutes at room temperature, protected from light. After incubation, coverslips were washed three times with 1% BSA in 1X PBS, pH 8.0. Coverslips were washed one additional time in 1X PBS, pH 8.0 before counterstaining with 4',6-diamidino-2-phenylindole (DAPI, 5 μg mL^{-1}) in Fluoromount-G mounting medium on a glass slide. Slides were imaged using the Delta Vision Elite deconvolution microscope (Applied Precision).

Quantitative PCR: *TcITPK1* riboswitch (*glmS/M9*) transfected epimastigotes had the addition of 10 mM exogenous glucosamine to 5 mL cultures. Total RNA was isolated from epimastigotes Day 0-3 using TRI® reagent (Sigma) by following manufacturer's guidelines. Extracted RNA was treated with DNase I to remove any genomic DNA contamination. cDNA was synthesized using SuperScript™ III First-Strand Synthesis System (Invitrogen) and RNase H removed any remaining RNA contamination. After RNase step, cDNA sample concentrations were determined by nanodrop and then samples were diluted to a concentration of 100 ng μL^{-1}. qPCR reactions were completed using the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad) in a 96-well PCR plate with 1 μL cDNA (100 ng μL^{-1}), 5 μL iQ™ SYBR ® Green Supermix (Bio-Rad), and 4 μl of nuclease-free water with primers at a final concentration of 300 nM (Primers 13-16). Activation of polymerase was performed at 95°C for 2 min. PCR cycling conditions included 39 cycles of denaturation at 95°C for 10 s, and annealing and extension at 60°C for 30 s. All sample reactions were done in triplicate. Normalization of genes was done using *T. cruzi* α-tubulin. Relative quantification of *TcITPK1* was normalized to α-tubulin according to ΔCT method. qPCR experiment was repeated for a total of three independent experiments.

Results

Endogenous tagging allows for the visualization of proteins of interest without disrupting subcellular localization. CRISPR/Cas9-mediated endogenous C-terminal tagging of *TcITPK1* (**Figure 11A**) allowed for the incorporation of 3x c-*Myc* tag. Incorporation of this cassette design was verified in a two-pronged method: incorporation of design by PCR analysis (**Figure 11B**) and tagged-protein expression by western blot (**Figure 11C**). Both verification methods demonstrated successful CRISPR/Cas9 tagging for *TcITPK1-3xc-Myc* (**Figure 11B, 11C**). Transfected epimastigotes were then selected to undergo a series of immunofluorescent assays. *TcITPK1* IFA images localized this protein of interest to the cytosol of the parasite (**Figure 11D**). There does appear to have a punctate distribution of the *Tc*ITPK1 in the cytosol, potentially localizing in granules, so additional colocalization IFAs with established cytosolic and granular proteins should be conducted to further validate its localization.

To study the ablation of *TcITPK1* in epimastigotes, two cassettes were utilized along with two unique *TcITPK1*-targeted sgRNA designs (**Figure 12A, 12B**). Each of the four unique designs were done in technical duplicates and the complete co-transfection experiments were repeated for a total of three independent times. Each of the transfections resulted in mixed populations of *TcITPK1* ablated and WT epimastigotes (**Figure 12C**). However, after limited dilution to isolate *TcITPK1*-KO parasites, no epimastigotes survived the selection process. Since these targeted designs continued to demonstrate a death phenotype upon *TcITPK1* knockout, a riboswitch method of knocking down *TcITPK1* was explored (**Figure 13A**). The first co-transfection resulted in a death phenotype for both *TcITPK1-glmS* and *TcITPK1-M9* and this result was attributed to failure of cassette incorporation. The second transfection resulted in

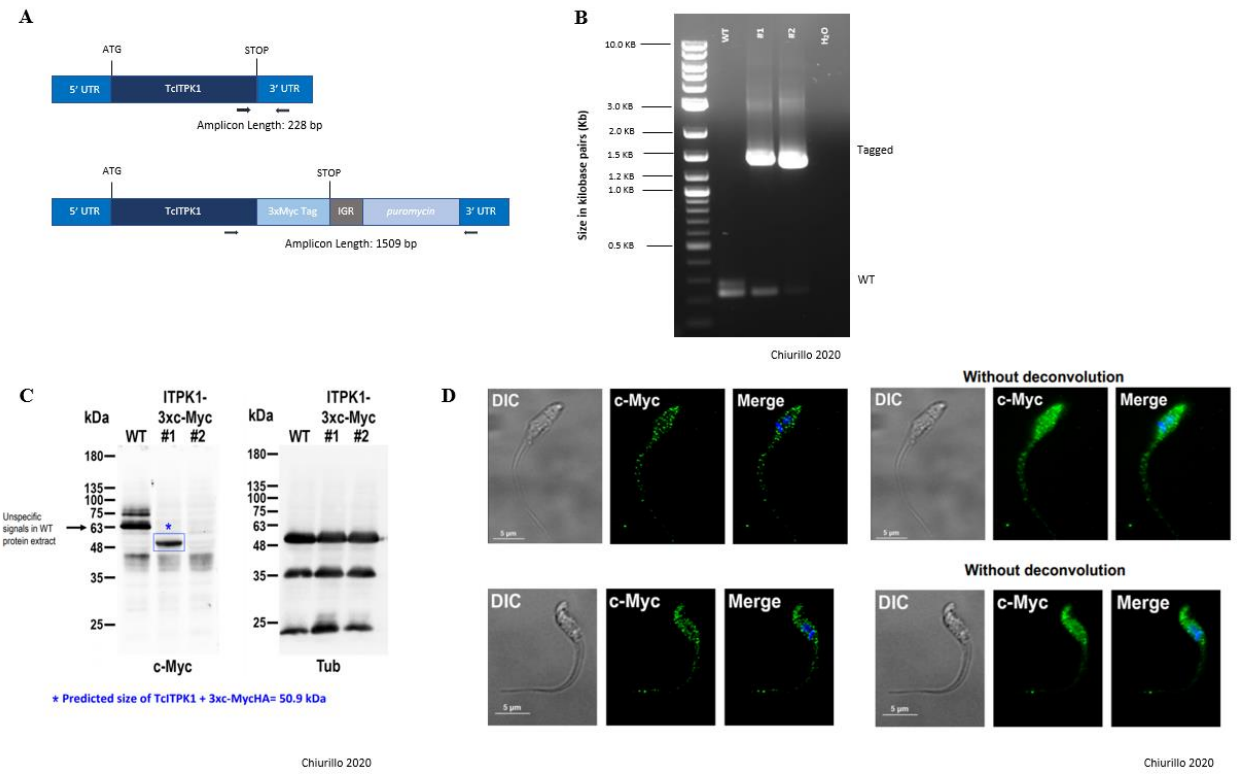


Figure 11. CRISPR/Cas9 Endogenous C-Terminal Tagging of *TcITPK1* (TcCLB.503885.50). *T. cruzi* epimastigotes were endogenously tagged with 3xMyc Tag using CRISPR/Cas9 strategies (**A**). Incorporation of the design was verified by both PCR (**B**) and Western Blot (**C**). (D). Immunofluorescence assays were conducted using DAPI to visualize the nucleus and anti-c-Myc (GFP) to visualize tagged *TcITPK1*.

viable parasites that successfully incorporated the cassettes (**Figure 13B, 13C**). Both *TcITPK1-glmS* and *TcITPK1-M9* have similar expression on the Western Blot verification because it was completed without the addition of exogenous glucosamine 6-phosphate to verify successful incorporation of the cassette; therefore, *TcITPK1* ablation would not occur because the *glmS* riboswitch does not have enough of the ligand to turn on its self-cleaving activity (**Figure 13C**). Further qPCR and Southern Blot validation experiments need to be conducted to validate the downregulation of *TcITPK1* in *TcITPK1-glmS* epimastigotes.

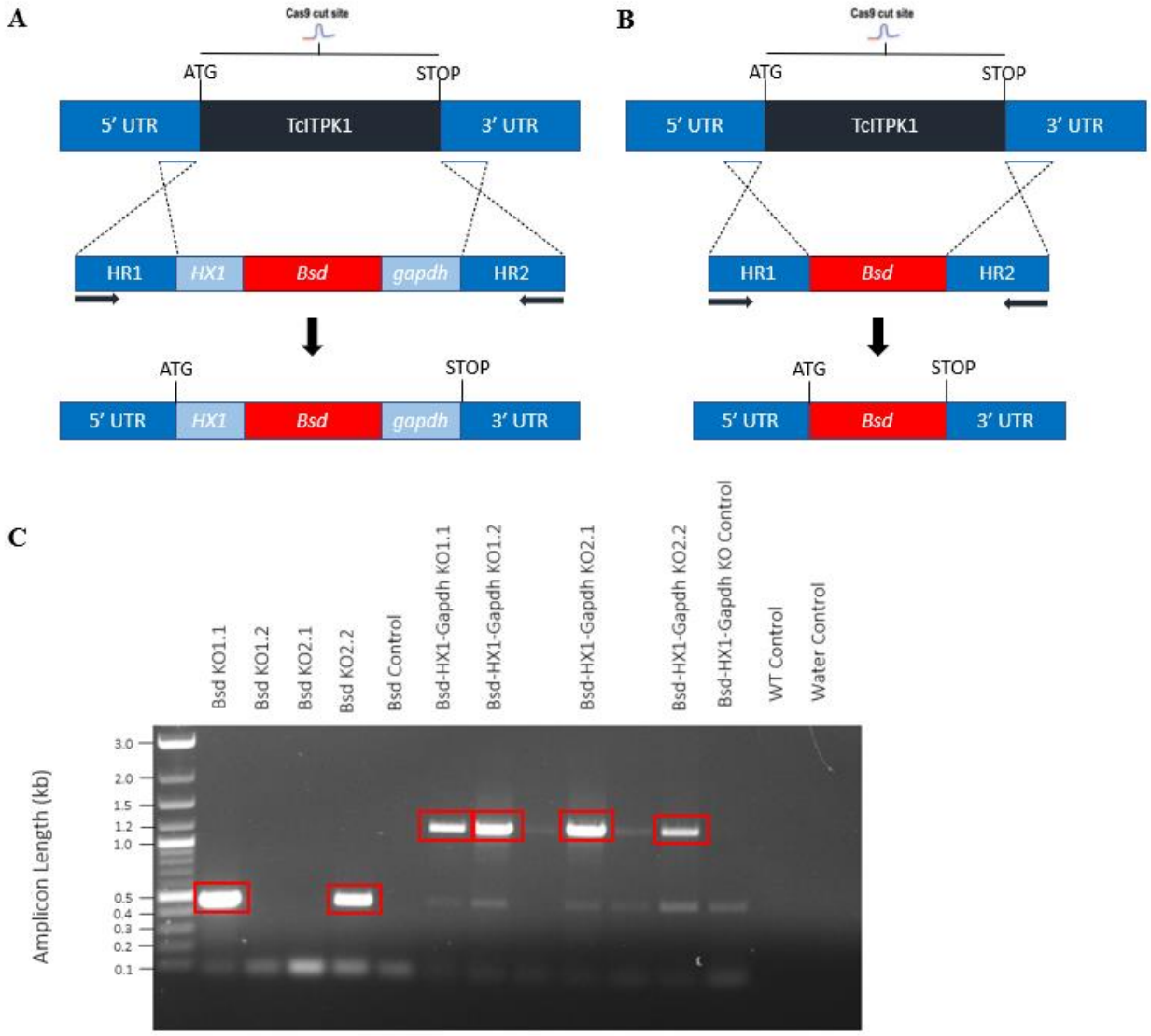


Figure 12. **CRISPR/Cas9 Knockout of *TcITPK1* using *HX1-Bsd-gapdh* and Classic *Bsd* Cassette Design**. For the *TcITPK1*-targeted knockout in *T. cruzi* epimastigotes, we designed two sgRNAs and utilized two unique cassettes designs to replace the endogenous gene: *HX1-Bsd-gapdh* cassette (**A**) and classic *Bsd* cassette (**B**). Each design had two technical replicates and was repeated in three independent experiments. PCR verification of *TcITPK1* ablation on the third independent *HX1-Bsd-gapdh* and classic *Bsd* cassette KO attempt is shown above (**C**). There are mixed populations of WT and *TcITPK1* ablated epimastigotes (correct incorporation of the cassette is denoted in red). Primer dimer is seen at approximately 0.1 kb on the gel.

Discussion

While bioinformatics provides a fantastic preliminary toolset to predict the functions of hypothetical proteins, the technological advances in genetic engineering using CRISPR/Cas9 allow for the elucidation of molecular function of biomolecules. Additionally, these experiments help to resolve dissonance between hypothesized biochemical mechanisms and bioinformatical predictions. For example, *Tc*ITPK1 was predicted to have a transmembrane domain according to InterPro and Protter; however, TMHMM2.0 predicted that there was only a 7.194% chance that it had this domain[72,73,102]. Orthologs of *Tc*ITPK1 are found consistently in the cytosol where the lipid-independent IPP pathway occurs. Immunofluorescence assays of CRISPR/Cas9 endogenously tagged *TcITPK1* demonstrated that TcITPK1 localized to the cytosol (**Figure 11D**), bolstering the hypothesis that this protein mediates the lipid-independent IPP pathway.

The generation of a CRISPR/Cas9-mediated *TcITPK1* knockout and knockdown had its challenges. Three cassette strategies were created to generate an ablated *Tc*ITPK1 population: (1) *HX1-Bsd-gapdh* cassette (**Figure 12A**), (2) *Bsd* cassette (**Figure 12B**), and (3) riboswitch *glmS*/*M9* cassette (**Figure 13A**). The first challenge faced in this study is the expression levels of *TcITPK1* in epimastigotes based on the four life-cycle stages of *T. cruzi* – *TcITPK1* had a -0.75 downregulated expression value (log2 ratio) – meaning that *TcITPK1* expression is downregulated during this stage[44,103]. The first knockout cassette design, *HX1-Bsd-gapdh*, included *HX1*, for *trans*-splicing, and *gapdh*, for the addition of the poly(A) tail, to allow for better expression of the Blasticidin S. Two protospacer regions were designed for the Cas9 knockout in case one of the knockout designs were non-specific or did not ablate *Tc*ITPK1. The first co-transfection of this cassette generated mixed populations by PCR verification (data not shown); however, when epimastigotes were cloned by limited dilution, the *TcITPK1*-KO

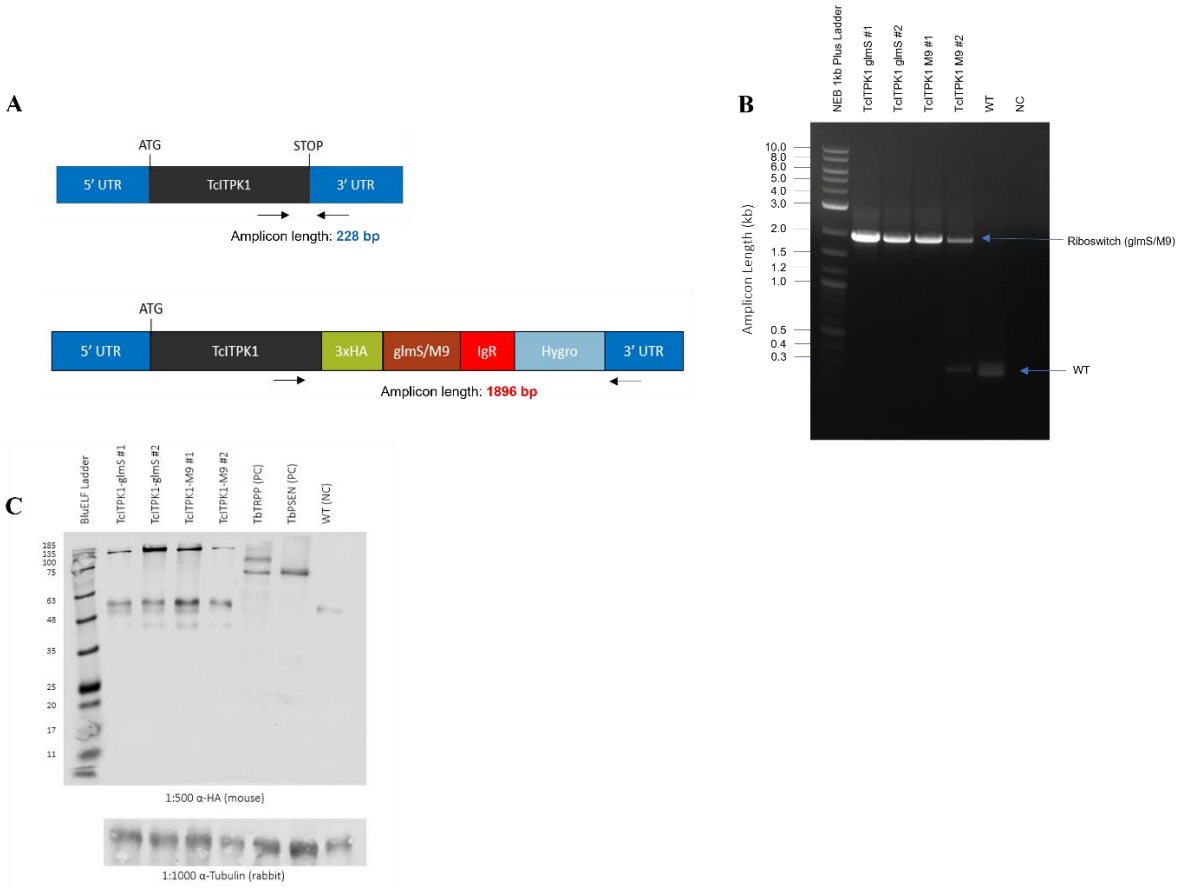


Figure 13. **CRISPR/Cas9 Knockdown of *TcITPK1* using *glmS*/*M9* Riboswitch Method**. For the *TcITPK1*-targeted knockout in *T. cruzi* epimastigotes, we utilized the previously published riboswitch method, co-transfecting epimastigotes with *TcITPK1*-targeted sgRNA and either the *glmS* or *M9* donor DNA **(A)**. **B.** PCR verification of *TcITPK1-glmS/M9* knockdown design in *T. cruzi* epimastigotes. Expected size of design is 1896 bp while WT is 228 bp. **C.** Western Blot verification of *TcITPK1-glmS/M9* knockdown design. Expected size of TcITPK1-glmS and TcITPK1-M9 is 50.6 kDa. Non-specific anti-HA banding is seen at between 135 and 185 kDa. This banding is most likely the specific band for HA-tagged Cas9 which is observed at a molecular weight of 158 kDa.

epimastigote populations were lost. Two additional *HX1-Bsd-gapdh* cassette transfections with the two independent sgRNAs were conducted and resulted in no viable knockouts in the populations after six weeks. After the failure of this cassette, the classic *Bsd* cassette transfection

was conducted three independent times but also failed to yield viable parasite populations. Slow growth of parasites and ultimate death phenotype was seen in all designs in both 20% FBS supplemented LIT medium and conditioned 20% FBS supplemented LIT medium with their appropriate antibiotics. Parasites integrating the *HX1-Bsd-gapdh* and *Bsd* cassettes were eventually lost after subsequent rounds of antibiotic selection and no *TcITPK1* ablated cultures lasted past six weeks. This result suggests that *TcITPK1* may be an essential gene in *T. cruzi* epimastigotes and that the lipid-independent pathway plays an important role in parasite survival.

The optimization of the riboswitch (*glmS/M9*) CRISPR/Cas9 strategy for *T. cruzi* allows for the study of essential proteins[96,101]. This method employs a *Bacillus subtilis* riboswitch (*glmS*) to control gene expression in response to the addition of exogenous glucosamine 6-phosphate, triggering its self-cleaving activity; the inactive mutant riboswitch (*M9*) serves as a negative control for gene expression in response to exogenous glucosamine 6-phosphate. Utilizing this strategy to target *TcITPK1* (**Figure 13A**) allows for the severe downregulation of *TcITPK1* mRNA transcription in active *glmS* riboswitch transfections while the inactive *M9* riboswitch transfection cell line serves as a negative control. Both cultures have viable parasites. Current qPCR experiments are focused on quantifying the level of *TcITPK1* downregulation in the *TcITPK1-glmS* cultures when compared to the *TcITPK1-M9* cultures of wild type and scramble epimastigotes – all supplemented with glucosamine 6-phosphate. If parasites expressing *TcITPK1-glmS* demonstrate significant downregulation as compared to the *TcIPTK1-M9* control, inositol polyphosphate concentrations should be measured by CE-ESI-MS analysis to determine this impact on *T. cruzi* inositol polyphosphate synthesis.

Overall, this chapter entails the creation of three CRISPR/Cas9-mediated cell lines: (1) endogenously tagged *TcITPK1*, (2) *TcITPK1-glmS* active riboswitch knockdown, and (3)

TcITPK1-M9 inactive riboswitch control. Immunofluorescence assays of *T. cruzi* epimastigotes with endogenously tagged *Tc*ITPK1 have shown this protein to localize to the cytosol, the site of the lipid-independent IPP synthesis pathway, and bolsters the hypothesis that this enzyme mediates this pathway. Additionally, *TcITPK1* appears to be an essential enzyme to *T. cruzi* epimastigotes as four distinct knockout strategies all resulted in slowed epimastigote growth and, ultimately, death phenotype within six weeks of selection. The creation of a *TcITPK1* knock down with the riboswitch cassette shows immense possibilities for understanding *T. cruzi* lipid-independent IPP synthesis pathway *in vitro*; however, qPCR analysis is still required for the quantifiable validation of genetic knockdown.

CHAPTER 5:

*Tc*ITPK1 MEDIATES THE LIPID-INDEPENDENT INOSITOL POLYPHOSPHATE SYNTHESIS PATHWAY IN *TRYPANOSOMA CRUZI*

Cline, S.E., Saiardi, A., and Docampo, R. To be submitted to *PNAS*.

Abstract

Inositol phosphate molecules are essential for eukaryotic survival, facilitating biological roles related to energy metabolism, calcium storage and release, and signaling cascades. While inositol phosphate homeostasis was originally thought to be solely facilitated in a lipid-dependent phospholipase C (PLC)-driven mechanism, recent studies have highlighted a highly conserved, eukaryotic cytosolic lipid-independent pathway mediated by inositol tetrakisphosphate 1-kinase (ITPK1). The protozoan parasite *Trypanosoma cruzi* is hypothesized to use this soluble pathway due to continued synthesis of inositol polyphosphate species even after CRISPR/Cas9 genetic knockout of *Tc*PI-PLC. *Saccharomyces cerevisiae plc1*Δ genetic screens were used to characterize the molecular function of *Tc*ITPK1. Complementation of *S. cerevisiae plc1*Δ with *Tc*ITPK1 was able to rescue growth deficiencies in *myo*-inositol deficient media to that of wild type levels. Additionally, *Tc*ITPK1 complementation rescued IP_6 formation as demonstrated by SAX-HPLC assays in *S. cerevisiae plc1*Δ. These results provide strong evidence that *Tc*ITPK1 is capable of mediating this lipid-independent pathway and that this pathway is conserved in a multitude of higher order eukaryotes.

Introduction

Inositol phosphates are important for a myriad of biological processes including phosphate homeostasis, secondary messengers to trigger signaling cascades, energy metabolism, and disease pathogenicity[3,4,23,31,40]. These ubiquitous small molecules can be synthesized by two distinct mechanisms: (1) lipid-dependent generation through the phospholipase C (PLC) hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) into diacylglycerol (DAG) and inositol trisphosphate (IP_3) and (2) lipid-independent generation through the phosphorylation by inositol tetrakisphosphate 1-kinase (ITPK1) of inositol monophosphates formed from glucose-6-phosphate (G6P) and inositolphosphoceramide (IPC). The latter inositol polyphosphate (IPP) synthesis pathway is conserved in higher-order eukaryotic organisms[1].

Recent CRISPR/Cas9 experiments in our laboratory focused on characterizing phosphoinositide-specific phospholipase C (PI-PLC) in the protozoan parasite *Trypanosoma cruzi*, a neglected, tropical parasite endemic to the Americas that infects a total of approximately eight million people[8]. Original experiments failed to produce viable genetic knockouts of *TcPI-PLC*, leading to speculation on this gene being essential to *T. cruzi* [38]. However, recent CRISPR/Cas9 experiments not only produced viable knockouts of *TcPI-PLC* but *T. cruzi* epimastigotes demonstrated their capacity to synthesize highly phosphorylated inositol species (Miguel Chiurillo and Roberto Docampo, unpublished data; **Chapter 1**). This data demonstrates that *T. cruzi* parasites may mediate both IPP synthesis pathways for inositol phosphate homeostasis and parasite survival.

Saccharomyces cerevisiae (yeast) genetic screens have been essential for finding enzymes involved in IPP synthesis pathways[1,4,104–108]. In *S. cerevisiae,* there is a notable lack of *ITPK1* and the lipid-independent IPP synthesis pathway, relying on their homologous PI-PLC

gene, *plc1,* to synthesize different inositol phosphate species in a lipid-dependent manner[1]. *S. cerevisiae plc1*Δ failed to produce IPP species unless rescued by genetic complementation with a ITPK1 homolog. Human ITPK1 (*Hs*ITPK1) functionality was determined through a series of yeast genetic screening experiments and demonstrated its ability to function as a mediator of the lipid-independent pathway, rescuing IP_6 and IP_7 synthesis in complemented yeast. This chapter utilizes yeast genetic screening/complementation techniques to characterize *Tc*ITPK1 and determine if it can mediate the lipid-independent IPP synthesis pathway.

Materials and Methodology

Yeast Transformation: A colony was incubated in 3 mLYPD media at 30°C with shaking at 250 rpm for 8 hours. Fifty µL of the culture were seeded in a fresh flask containing 50 mL YPDA and incubated at 30°C with shaking at 250 rpm for 16-20 hours. When OD_{600} of 0.15-0.3 was reached yeast cells were spun down and supernatant was discarded. Resuspend yeast cells in 100 mL YPD media and incubate for 3-5 hours. Wash once with sterile deionized water and resuspend cells in 3 mL 1.1x TE/LiAc solution and split between two microcentrifuge tubes. Centrifuge down pellet and discard supernatant. Resuspend in 600 µL 1.1x TE/LiAc solution and add 0.1 mg salmon sperm DNA and 0.1 µg plasmid DNA (pCA45, pCA45-*Tc*ITPK1, or pCA45-*HsI*TPK1) to a new microcentrifuge tube. Pipette 100 µL competent yeast cells and vortex briefly. Add 600 µL sterile PEG/LiAc solution. Incubate at 30°C shaking at 250 rpm for 30 minutes. Add 70 µL DMSO and mix by gentle inversion, three times. Head shock for 15 minutes in a 42°C water bath and 2 minutes on ice. Centrifuge down yeast cells and resuspend in 1X TE Buffer. Plate 100 µL on synthetic complete uracil dropout plates (SC-URA). Incubate at 30°C until colonies form.

Western Blot Analysis: Transformed yeast were grown to mid-log phase (OD_{600} 0.3-0.8) in YPD media. Transformed yeast samples were spun down and washed twice with 1X PBS, pH 7.4. Cells were then resuspended in a 100 mM TRIS-HCl (pH 7.2), 1 mM DTT, 1 M sorbitol, and 1 mg mL^{-1} zymolyase buffer at room temperature for 30-60 minutes, checking for 90% spherical morphology every 15 minutes after 30 minutes. Samples were spun down again and resuspended in 100 μL RIPA Buffer while being agitated at 4°C for 60 minutes. After lysis, microcentrifuge tubes were spun down at 13.2k rpm at 4°C. Supernatant was collected into a fresh microcentrifuge tube and put on ice. Pellet was discarded. Protein concentrations were standardized using a Bradford assay with BSA standard curve. Approximately 5-10 μg of each lysate was mixed with 4x Laemmli sample buffer and incubated at 95°C for 5 minutes. Extracted proteins were then subjected to electrophoresis using a 12% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) at 50 V for 30 minutes, 110 V until marker goes to the end of gel (60-120 minutes). BLUelf Prestained Protein Ladder was used to calculate molecular weights. Electrophoresed proteins were transferred to a nitrocellulose membrane using a Bio-Rad transblot apparatus for 5 hours at 30 V at 4°C. Following transfer, membrane blots were blocked with 5% nonfat milk in PBS-T at room temperature for 60 minutes. Blots were probed with rabbit polyclonal hexokinase antibody (1:5,000) and rabbit polyclonal glutathione-S-transferase (GST) antibody (1:5,000) at room temperature for 60 minutes. Blots were washed with PBS-T for 5 minutes, three times. Then, secondary antibody conjugated goat anti-rabbit antibody (1:20,000) was incubated for 60 minutes, protected from light. Membranes were then imaged and analyzed using the Odyssey infrared system software (LICOR Biosciences)

Myo-Inositol Deficient Growth Assay: Transformed yeast cultures were grown overnight in 10 mL SC-URA media, shaking at 200 rpm, at 30°C until log phase (OD_{600} = 0.3-0.8). Yeast cultures were diluted to OD_{600} = 0.01 in myo-inositol deficient media, SC-URA (0 μM inositol) in a sterile 96-well assay plate for a total volume of 250 μL. Each sample was repeated in technical triplicate on the assay. For continued measurement of culture OD, the assay plate with lid was placed in Synergy H1 Hybrid Multi-Mode Microplate Reader (BioTek). Yeast cultures were incubated at 30°C while shaking and the OD_{600} was collected every 30 minutes for 48 hours. Each assay was repeated three times to achieve biological replicates.

[^{3}H]-Inositol labeling, IP Extraction, and SAX-HPLC Analysis of WT and plc1Δ Yeast: The radioactive isotope labeling, IP extraction, and SAX-HPLC analysis of pCA45-*TcITPK1* (*plc1*Δ), pCA45-*HsITPK1* (*plc1*Δ), pCA45 (*plc1*Δ), and pCA45 (WT) yeasts were performed by Dr. Adolfo Saiardi of the University College London as previously described[109]. In summary, yeast samples were grown overnight at 30°C in SC-URA media and used to seed a new flask of inositol-free SC-URA media containing 5 μCi mL^{-1} [^{3}H]-inositol. This radioactively labeled culture was grown overnight at 30°C, with shaking, to an OD_{600} = 0.5-0.9. Labeled yeast were collected by centrifugation (2,000 x *g*, 2 minutes, 4°C), washed once with ice cold water or inositol-free SC-URA media, and resuspended in ice-cold water. Yeast samples were spun down (2,000 x *g*, 2 minutes, 4°C) and resuspended in extraction buffer (1 M $HClO_4$, 3 mM EDTA, and 0.1 mg mL^{-1} IP_6) and glass bead(s). Yeast cell walls were broken by vortexing for 5 minutes at 4°C and debris removed by centrifugation (15,000 x *g*, 5 minutes). The remaining supernatant was neutralized with neutralization buffer (1 M K_2CO_3 and 3 mM EDTA) to a pH between 6.0-8.0. Tubes were incubated s on ice for 2 hours, flicking mixture every 30 minutes, and then

centrifuged down at (15,000 x *g*, 5 minutes). Supernatants could be stored at 4°C or immediately moved on to SAX-HPLC analysis. Yeast samples were separated onto the PartiSphere SAX (4.6 × 125 mm) column (Hichrom) and eluted with a gradient generated by mixing 1 mM EDTA and Buffer B (1 mM EDTA/1.3 M $(NH_4)_2HPO_4$, pH 3.8): 0 to 5 minutes, 0% Buffer B; 5 to 10 minutes, 0 to 10% Buffer B; 10 to 60 minutes, 10 to 100% Buffer B; and 60 to 80 minutes, 100% Buffer B. Experiments were done in triplicate.

Primer Number	Primer Name	Sequence (5'→3')
18	Yeast pCA45-TcITPK1 Forward Primer (SalI)	TCACGTCGACAATGAGCACTGCTCCAAGAGG
19	Yeast pCA45-TcITPK1 Reverse Primer (NotI)	ATAAGAATGCGGCCGCTTAGCAGCAGAAGTTGCC
20	Yeast pCA45 Forward Check	ACGCACCAGATGGTCATA
21	Yeast pCA45 Reverse Check	TATGACCATCTGGTGCGT

Table 3. Primers Required for *TcITPK1* Yeast Complementation Experiments.

Results

To explore if *Tc*ITPK1 could rescue IP_6 and IP_7 synthesis in complemented *S. cerevisiae* genetic screens, two yeast strains were utilized – a parental wildtype and *S. cerevisiae plc1*Δ – that were complemented with pCA45-*Tc*ITPK1, pCA45-*Hs*ITPK1, or pCA45 (**Figure 14A**). Verification of the successful plasmid incorporation and protein expression was done in a two-pronged manner, first sequence verification by Sanger sequencing as well as western blot analysis (**Figure 14A, B**). We assessed the impacts of ITPK1 complementation by a growth assay in *myo*-inositol deficient media growth assays (**Figure 15**) and strong anion exchange (SAX) high-performance liquid chromatography (SAX-HPLC) of yeast labeled with [^{3}H]-inositol (**Figure 16**).

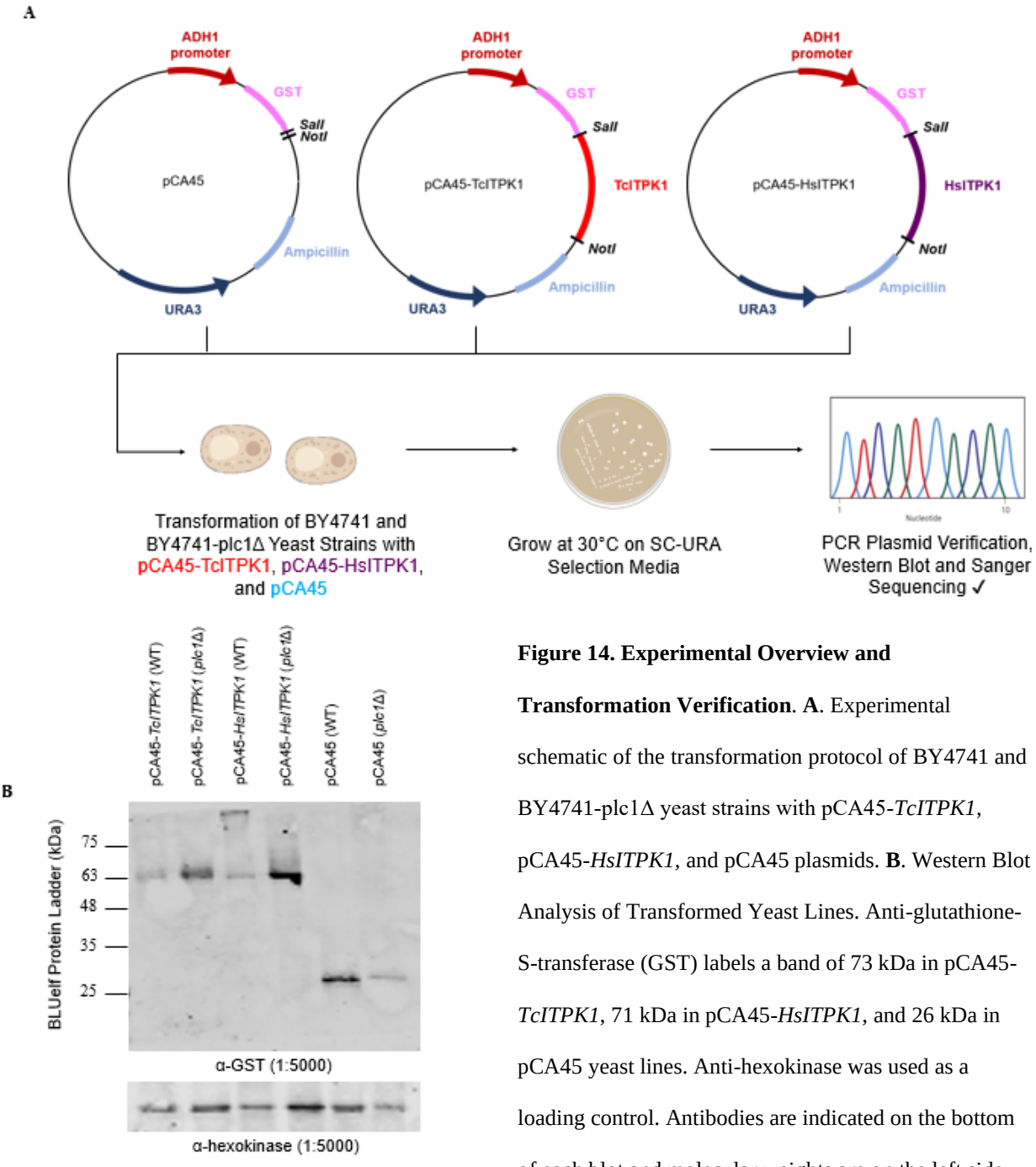


Figure 14. Experimental Overview and Transformation Verification. **A**. Experimental schematic of the transformation protocol of BY4741 and BY4741-plc1Δ yeast strains with pCA45-*TcITPK1*, pCA45-*HsITPK1*, and pCA45 plasmids. **B**. Western Blot Analysis of Transformed Yeast Lines. Anti-glutathione-S-transferase (GST) labels a band of 73 kDa in pCA45-*TcITPK1*, 71 kDa in pCA45-*HsITPK1*, and 26 kDa in pCA45 yeast lines. Anti-hexokinase was used as a loading control. Antibodies are indicated on the bottom of each blot and molecular weights are on the left side.

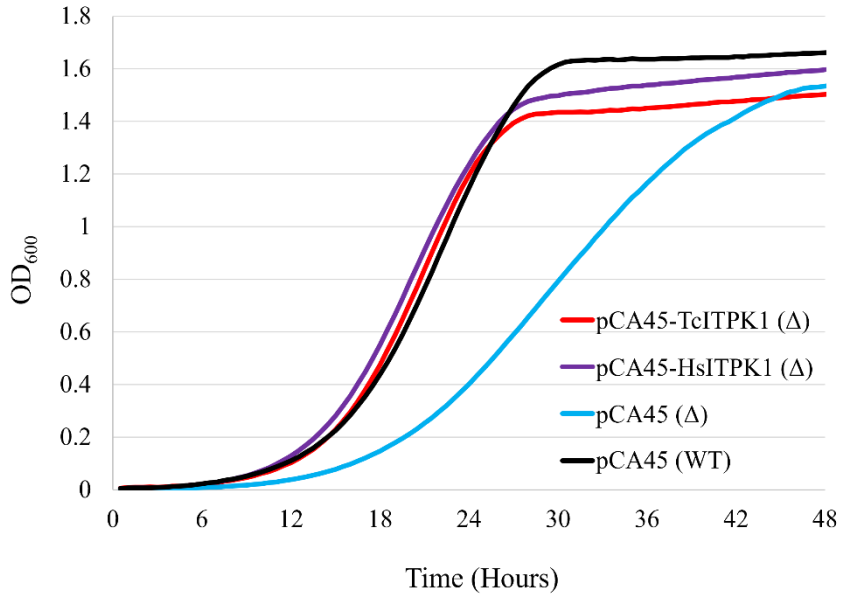
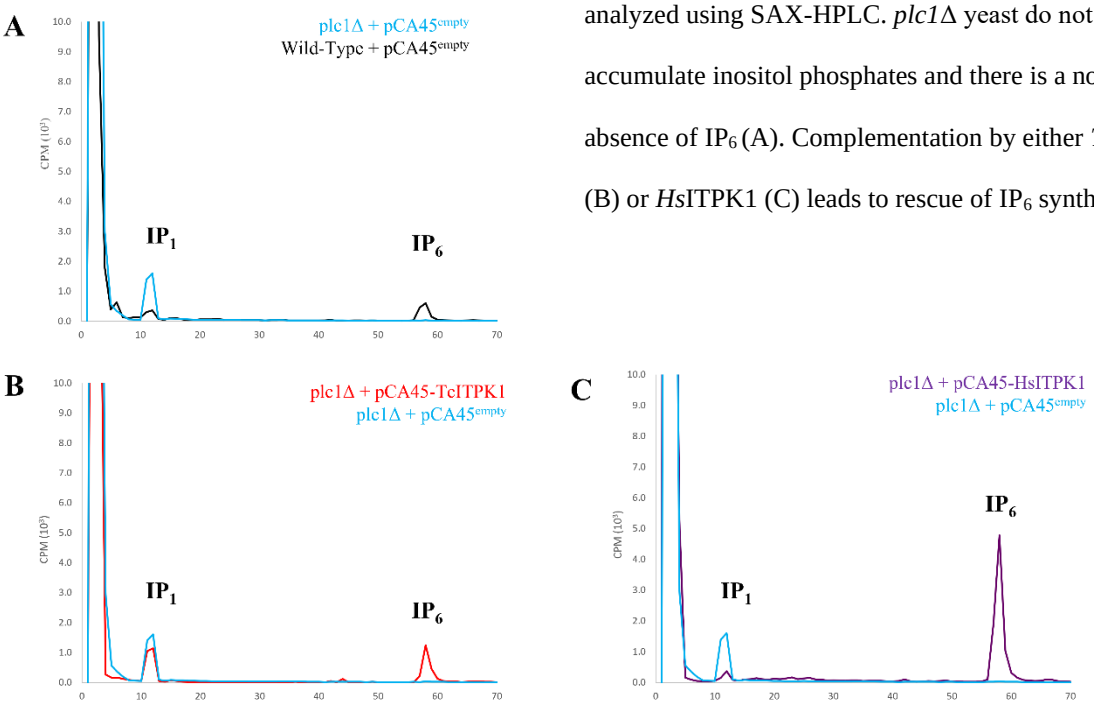


Figure 15. TcITPK1 Complementation Rescues Yeast Growth Deficiencies in PLC1 Mutant Yeast Lines in *myo*-Inositol Depleted Selection Media. BY4741 and BY4741-plc1Δ lines were grown at 30°C in SC-URA media (0 µM) for 48 hours with OD_{600} measurements taken every 30 minutes. pCA45-TcITPK1 (plc1Δ), pCA45-HsITPK1 (plc1Δ), pCA45 (plc1Δ), and pCA45 (WT) are shown in red, purple, blue, and black, respectively. N=6.

A 48 hour growth curve of transformed yeasts cultured in *myo*-inositol deficient media was done to determine if transformed yeast could rescue any potential growth deficiencies of *plc1*Δ yeast. pCA45 *plc1*Δ (light blue) yeasts had a distinct lag in their growth when compared to pCA45 wild type (black) yeast (**Figure 15**). pCA45-wild type yeast reached their stationary phase at approximately 30 hours of growth whereas pCA45 *plc1*Δ yeast took the entire 48 hours and featured a slower exponential phase. When *S. cerevisiae plc1*Δ was complemented with either *Hs*ITPK1 (purple) or *Tc*ITPK1 (red), yeast samples were able to rescue growth deficiencies to that of wild type levels.

SAX-HPLC analysis with [^{3}H]-inositol labeled wild type and *plc1*Δ transformed yeast allowed for the formation of inositol phosphate species to be quantitated. In pCA45-wild type yeast, there was formation of IP_6, seen between elution times of 55 and 60 minutes (**Figure 16A**). This highly phosphorylated inositol phosphate species peak is not seen in pCA45 *plc1*Δ yeast, which instead demonstrated an accumulation of IP_1 species (**Figure 16A**). *S. cerevisiae plc1*Δ complemented with pCA45-*Hs*ITPK1 rescued the synthesis of IP_6 formation (**Figure 16C**) in agreement with previously published results[1]. Complementation with pCA45-*Tc*ITPK1 also lead to the rescue of IP_6 (**Figure 16B**), although not to the same extent of *Hs*ITPK1.

Figure 16. *Tc*ITPK1 Rescues IP_6 Formation in PLC1 Mutant Yeast Lines. [^{3}H]-inositol labeled wild type and *plc1*Δ yeast strains complemented with *Tc*ITPK1, *Hs*ITPK1 (PC), and empty pCA45 (NC) were analyzed using SAX-HPLC. *plc1*Δ yeast do not accumulate inositol phosphates and there is a notable absence of IP_6 (A). Complementation by either *Tc*ITPK1 (B) or *Hs*ITPK1 (C) leads to rescue of IP_6 synthesis.



Discussion

*Tc*PI-PLC genetic knockout in *T. cruzi* epimastigotes does not stop the formation of higher order IPPs, demonstrating the existence of an alternative IPP synthesis pathway. *Tc*ITPK1 is hypothesized to facilitate this alternative pathway due to its orthologous relationship with *Hs*ITPK1, the mediator of a lipid-independent IPP synthesis pathway in humans. *S. cerevisiae* genetic screens were employed to determine its capabilities to phosphorylate inositol monophosphate and other inositol phosphate species.

SAX-HPLC experiments demonstrate that complementation of *S. cerevisiae plc1*Δ with *Tc*ITPK1 rescues the synthesis of IP_6 (**Figure 16B**). This result is also in agreement with previous results using the positive control *Hs*ITPK1 (**Figure 16C**). However, when there is comparison of rate of IP_6 rescue between the eukaryotic ITPK1 enzymes, *Tc*ITPK1 does not rescue to the same magnitude. This difference in rescue might be attributed to variance in the IP-binding site. There are three highly conserved residues that are important for IP- and ATP-binding – H167, K199, and R212 – in *Hs*ITPK1. *Tc*ITPK1 only shares two of these conserved residues and therefore this may explain the differences in overall rescue of IP_6 synthesis. *Tc*ITPK1 complemented yeast do, however, rescue IP_6 synthesis to that of yeast wild type levels (**Figure 16A, 16B**) so the difference between the orthologous ITPK1s may be due to difference in specific activity (**Figure 16B, 16C**).

In contrast, *S. cerevisiae* growth assays demonstrate complete rescue of growth of *S. cerevisiae plc1*Δ complemented with *Tc*ITPK1 and *Hs*ITPK1 (**Figure 15**). This impact may be due to the formation of IPP species, an important molecule to a variety of yeast cellular processes including regulation of major glycolytic transcription factor GCR1, pseudohyphal growth, and ATP concentration[4,104]. *S. cerevisiae* phosphoinositide-specific phospholipase C

gene (PLC1) is a homolog to *Tc*PI-PLC and other delta class PLC enzymes and the knockdown of *plc1* causes organisms to over accumulate PIP_2 and fail to synthesize IPP.[1] The lagging growth phenotype noted in the pCA45 (Δ) yeast grown in *myo*-inositol deficient media further demonstrates the importance of IPP species synthesis as these samples took longer to reach stationary phase and their exponential growth was less than expected (**Figure 15**). However, complementation with either eukaryotic ITPK1 enzymes – be it *Tc*ITPK1 or *Hs*ITPK1 – allowed for the rescue of this growth deficiency.

This chapter demonstrates the facilitatory role that *Tc*ITPK1 fulfills to mediate the lipid-independent pathway that is conserved amongst higher-order eukaryotes. Future directions of this study would be to conduct additional *S. cerevisiae* genetic screens to knockout the two homologous yeast enzymes that generate the inositol monophosphate species in addition to the original *plc1Δ*. *TcISC1* and *TcINO1* are predicted to mediate the formation of inositol monophosphate from glucose-6-phosphate and inositolphosphoceramide, respectively, in *T. cruzi*. Previous genetic screens of *isc1*Δ have demonstrated the continued formation of IP_6; however, there is an absence of these components in *isc1*Δ*plc1*Δ yeasts complemented with ITPK1, demonstrating that sphingolipid degradation is important for inositol phosphate synthesis.[1] The use of these genetic screens would identify which specific metabolite molecules are imperative for the *T. cruzi* lipid-independent synthesis pathway.

CHAPTER 6:

CONCLUSIONS AND FUTURE DIRECTIONS

Trypanosoma cruzi is an obligate intracellular parasite that is the causative agent for Chagas Disease, a chronic bloodborne infection that has infected eight million individuals and puts another eighty million at risk of infection in the Americas[8]. Classically, inositol polyphosphate (IPP) synthesis occurs through the phosphoinositide phospholipase C (PLC)-catalyzed hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) into inositol triphosphate (IP_3) and diacylglycerol (DAG) and further IP_3 phosphorylation by additional kinases in the cytosol. Recently, inositol-tetrakisphosphate 1-kinase (ITPK1) has been described as an enzyme that may mediate an alternative, lipid-independent IP_6 soluble synthesis pathway through phosphorylation of IP_1 and other intermediates in the cytosol.

Preliminary bioinformatics research outlined in **Chapter 2** identified hypothetical protein TcCLB.503885.50 as a potential *T. cruzi* ITPK1 (TcITPK1) homolog. Further phylogenetic studies have demonstrated that this hypothetical protein is highly conserved in parasitic kinetoplastids and to be an ortholog of human ITPK1 (*Hs*ITPK1), a protein previously established to follow the lipid-independent IPP synthesis pathway. Clustal Omega alignment of *Tc*IPTK1 and *Hs*ITPK1 demonstrates that these proteins share five of seven conserved IP-contact and ATP-binding residues in their binding pocket. Additionally, the active site and overall structure of *Tc*ITPK1 are highly conserved, as exhibited by predicted structural versus orthologous *Hs*ITPK1 experimental comparison (RMSD 3.05Å without twists, jFATCAT score

450.36, p-value 5.76e-12). While organometallic compound treatments did not modulate the *TcITPK1* transcript, other transcripts involved in the inositol polyphosphate pathway (*TcINO1*, *TcIP$_3$R*) were significantly modulated by Pd-dppf-mpo treatment (**Chapter 3**).

Advancement of CRISPR/Cas9 gene editing strategies has improved *T. cruzi* cellular biology research and aided in elucidating the function of *TcITPK1* (**Chapter 4**). This manuscript established the creation of three viable CRISPR/Cas9 *T. cruzi* epimastigote cultures: endogenously tagged *TcITPK1*, (2) *TcITPK1-glmS* active riboswitch knockdown, and (3) *TcITPK1-M9* inactive riboswitch control. Endogenously tagged *Tc*ITPK1 allowed for subcellular localization studies within the parasite. Immunofluorescence assays have shown this protein to localize to the cytosol, the site of the lipid-independent IPP synthesis pathway in other organisms. Preliminary data demonstrates that *Tc*ITPK1 may be essential to *Trypanosoma cruzi* survival as multiple CRISPR/Cas9 knockout strategies (*HX1-Bsd-gapdh* and *Bsd* cassettes) have resulted in no viable *in vitro* parasites at the epimastigote stage. Verification of successful *TcITPK1-glmS* active riboswitch knockdown, and *TcITPK1-M9* inactive riboswitch control provides a unique cell line to conduct future experiments to determine how *TcITPK1* knockdown impacts host cell invasion, intracellular replication, and inositol phosphate and inositol polyphosphate synthesis.

The ability of *Tc*ITPK1 to act as the mediator for the alternative, lipid-independent IPP synthesis assay was established in a series of complemented yeast assays (**Chapter 5**). SAX-HPLC experiments demonstrate the rescue of inositol hexakisphosphate (IP_6) after *plc1*Δ yeast are complemented with *Tc*ITPK1. Failure to complement *S. cerevisiae plc1*Δ with either of the eukaryotic ITPK1 proteins led to an overaccumulation of inositol monophosphate species, showing the specificity of this enzyme to mediate this pathway. Additional *S. cerevisiae* genetics

screens with *isc1*Δ *plc1*Δ and *ino1*Δ*plc1*Δ would allow for the identification of metabolites necessary for *Tc*ITPK1-mediated inositol polyphosphate synthesis.

Future experiments related to *Tc*ITPK1 enzymatic activity, IP binding site-directed mutagenesis, further yeast complementation and mutagenesis studies, and continued CRISPR/Cas9 experiments on *TcITPK1* riboswitch knockdown parasites will aid in advancing the characterization of *Tc*ITPK1 and the role of lipid-independent IPP synthesis in the parasitic life cycle. The characterization of this enzyme and its pathway will develop a deeper understanding of how this pathway affects calcium signaling, phosphate homeostasis, and disease pathogenicity in *T. cruzi*. Additionally, studying this parasitic kinetoplastid pathway could increase our understanding of orthologous eukaryotic IPP biology and provide insights to evolutionarily conserved processes.

Elucidating the function of inositol tetrakisphosphate 1-kinase in *Trypanosoma cruzi* is of great medical significance. Current standard treatments for Chagas Disease are unsatisfactory, only targeting the acute stages of the parasitic infection and leading to unfavorable side effects. By utilizing the mutant *T. cruzi* epimastigote lines developed by CRISPR/Cas9, we can explore if *TcITPK1* knockdown significantly affects the differentiation of epimastigotes into metacyclic trypomastigotes and intracellular amastigote replication of *T. cruzi in vitro*. If there are significant decreases in parasite viability or defects in parasitic growth, differentiation, and infectivity, *Tc*ITPK1 will become a potential therapeutic target for targeting *T. cruzi* infection in humans. While *Tc*ITPK1 does share structural identity (20%) and similarities (40%) to *Hs*ITPK1, the ITPK1 binding pockets feature key spatial and amino acid residue differences that could be exploited by small molecule therapeutics to combat Chagas Disease specifically. Utilizing the information gained by studying *Tc*ITPK1 and the lipid-independent inositol

polyphosphate pathway from these *in vitro* systems and future screening studies could accelerate a novel treatment for Chagas Disease, targeting both the acute and chronic stages of infection, in order to improve the quality of life for those infected with *Trypanosoma cruzi*.

REFERENCES

1. Desfougères, Y., Wilson, M. S. C., Laha, D., Miller, G. J. & Saiardi, A. ITPK1 mediates the lipid-independent synthesis of inositol phosphates controlled by metabolism. *Proceedings of the National Academy of Sciences* **116**, 24551–24561 (2019).
2. Wild, R. *et al.* Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. *Science (1979)* **352**, 986–990 (2016).
3. Chiurillo, M. A., Lander, N., Vercesi, A. E. & Docampo, R. IP3 receptor-mediated Ca2+ release from acidocalcisomes regulates mitochondrial bioenergetics and prevents autophagy in Trypanosoma cruzi. *Cell Calcium* **92**, 102284 (2020).
4. Szijgyarto, Z., Garedew, A., Azevedo, C. & Saiardi, A. Influence of Inositol Pyrophosphates on Cellular Energy Dynamics. *Science (1979)* **334**, 802–805 (2011).
5. Saiardi, A. & Cockcroft, S. Human ITPK1: A Reversible Inositol Phosphate Kinase/Phosphatase That Links Receptor-Dependent Phospholipase C to Ca^{2+} -Activated Chloride Channels. *Science Signaling* **1**, (2008).
6. Bhandari, R. *et al.* Protein pyrophosphorylation by inositol pyrophosphates is a posttranslational event. *Proceedings of the National Academy of Sciences* **104**, 15305–15310 (2007).
7. Barker, C. J., Illies, C., Gaboardi, G. C. & Berggren, P.-O. Inositol pyrophosphates: structure, enzymology and function. *Cellular and Molecular Life Sciences* **66**, 3851–3871 (2009).

8. Centers for Disease Control and Prevention. Parasites - American Trypanosomiasis: Epidemiology & Risk Factors.

9. U.S. Department of Health and Human Services, C. for D. C. and P. *CDC Yellow Book: Trypanosomiasis, American (Chagas Disease)*.

10. Benznidazole (Benznidazole) [package insert]. Florham Park, NJ: Exeltis USA, Inc.; 2017, 2020 (updated).

11. Torrico, F. *et al.* New regimens of benznidazole monotherapy and in combination with fosravuconazole for treatment of Chagas disease (BENDITA): a phase 2, double-blind, randomised trial. *The Lancet Infectious Diseases* **21**, 1129–1140 (2021).

12. LAMPIT (nifurtimox) [package insert]. Whippany, NJ: Bayer HealthCare Pharmaceuticals Inc.; 2020.

13. Forsyth, C. J., Stigler Granados, P., Pacheco, G. J., Betancourt, J. A. & Meymandi, S. K. Current Gaps and Needs for Increasing Access to Healthcare for People with Chagas Disease in the USA. *Current Tropical Medicine Reports* **6**, 13–22 (2019).

14. Wirtz, V. J., Manne-Goehler, J. & Reich, M. R. Access to Care for Chagas Disease in the United States: A Health Systems Analysis. *The American Journal of Tropical Medicine and Hygiene* **93**, 108–113 (2015).

15. World Health Organization (CHECK - SITE MAINTENANCE). Epidemiology.

16. CDC. Blood Donor Screening for Chagas Disease - United States, 2006-2007. *MMWR Morb Mortal Wkly Rep* **56**, 141–143 (2007).

17. Gorlin, J. *et al.* Evaluation of a new Trypanosoma cruzi antibody assay for blood donor screening. *Transfusion (Paris)* **48**, 531–540 (2008).

18. Whitman, J. D. *et al.* Chagas Disease Serological Test Performance in U.S. Blood Donor Specimens. *Journal of Clinical Microbiology* **57**, (2019).

19. Monteiro, F. A., Weirauch, C., Felix, M., Lazoski, C. & Abad-Franch, F. Evolution, Systematics, and Biogeography of the Triatominae, Vectors of Chagas Disease. in 265–344 (2018). doi:10.1016/bs.apar.2017.12.002.

20. Cordeiro, C. D., Saiardi, A. & Docampo, R. The inositol pyrophosphate synthesis pathway in *Trypanosoma brucei* is linked to polyphosphate synthesis in acidocalcisomes. *Molecular Microbiology* **106**, 319–333 (2017).

21. Cestari, I. & Stuart, K. Inositol phosphate pathway controls transcription of telomeric expression sites in trypanosomes. *Proceedings of the National Academy of Sciences* **112**, (2015).

22. Cestari, I., Haas, P., Moretti, N. S., Schenkman, S. & Stuart, K. Chemogenetic Characterization of Inositol Phosphate Metabolic Pathway Reveals Druggable Enzymes for Targeting Kinetoplastid Parasites. *Cell Chemical Biology* **23**, 608–617 (2016).

23. Ferris, C. D., Huganir, R. L., Supattapone, S. & Snyder, S. H. Purified inositol 1,4,5-trisphosphate receptor mediates calcium flux in reconstituted lipid vesicles. *Nature* **342**, 87–89 (1989).

24. López-Sánchez, U. *et al.* Interplay between primary familial brain calcification-associated SLC20A2 and XPR1 phosphate transporters requires inositol polyphosphates for control of cellular phosphate homeostasis. *Journal of Biological Chemistry* **295**, 9366–9378 (2020).

25. Lonetti, A. *et al.* Identification of an Evolutionarily Conserved Family of Inorganic Polyphosphate Endopolyphosphatases. *Journal of Biological Chemistry* **286**, 31966–31974 (2011).

26. Ried, M. K. *et al.* Inositol pyrophosphates promote the interaction of SPX domains with the coiled-coil motif of PHR transcription factors to regulate plant phosphate homeostasis. *Nature Communications* **12**, 384 (2021).

27. Macbeth, M. R. *et al.* Inositol Hexakisphosphate Is Bound in the ADAR2 Core and Required for RNA Editing. *Science (1979)* **309**, 1534–1539 (2005).

28. Lev, S. *et al.* Fungal Inositol Pyrophosphate IP $_7$ Is Crucial for Metabolic Adaptation to the Host Environment and Pathogenicity. *mBio* **6**, (2015).

29. Hashimoto, M. *et al.* Inositol 1,4,5-trisphosphate receptor regulates replication, differentiation, infectivity and virulence of the parasitic protist *Trypanosoma cruzi*. *Molecular Microbiology* **87**, 1133–1150 (2013).

30. Rao, F. *et al.* Inositol pyrophosphates promote tumor growth and metastasis by antagonizing liver kinase B1. *Proceedings of the National Academy of Sciences* **112**, 1773–1778 (2015).

31. Ahmed, I. *et al.* Huntington's disease: Neural dysfunction linked to inositol polyphosphate multikinase. *Proceedings of the National Academy of Sciences* **112**, 9751–9756 (2015).

32. Mantilla, B. S., Amaral, L. D. do, Jessen, H. J. & Docampo, R. The Inositol Pyrophosphate Biosynthetic Pathway of *Trypanosoma cruzi*. *ACS Chemical Biology* **16**, 283–292 (2021).

33. Mallery, D. L. *et al.* IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *Elife* **7**, (2018).

34. Sowd, G. A. & Aiken, C. Inositol phosphates promote HIV-1 assembly and maturation to facilitate viral spread in human CD4+ T cells. *PLOS Pathogens* **17**, e1009190 (2021).

35. Livermore, T. M., Azevedo, C., Kolozsvari, B., Wilson, M. S. C. & Saiardi, A. Phosphate, inositol and polyphosphates. *Biochemical Society Transactions* **44**, 253–259 (2016).

36. Shamsuddin, A. M. Metabolism and cellular functions of IP6: a review. *Anticancer Res* **19**, 3733–6.

37. Vucenik, I. & Shamsuddin, A. M. Protection Against Cancer by Dietary IP $_6$ and Inositol. *Nutrition and Cancer* **55**, 109–125 (2006).

38. Okura, M. *et al.* A Lipid-modified Phosphoinositide-specific Phospholipase C (TcPI-PLC) Is Involved in Differentiation of Trypomastigotes to Amastigotes of Trypanosoma cruzi. *Journal of Biological Chemistry* **280**, 16235–16243 (2005).

39. Wilson, M. P. & Majerus, P. W. Isolation of Inositol 1,3,4-Trisphosphate 5/6-Kinase, cDNA Cloning, and Expression of the Recombinant Enzyme. *Journal of Biological Chemistry* **271**, 11904–11910 (1996).

40. Ho, M. W. Y. *et al.* Regulation of Ins(3,4,5,6)P4 Signaling by a Reversible Kinase/Phosphatase. *Current Biology* **12**, 477–482 (2002).

41. Bayat, A. Science, medicine, and the future: Bioinformatics. *BMJ* **324**, 1018–1022 (2002).

42. Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I. & Watson, M. A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Frontiers in Genetics* **8**, (2017).

43. Li, P.-E. *et al.* Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Research* **45**, 67–80 (2017).

44. Aslett, M. *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Research* **38**, D457–D462 (2010).

45. Amos, B. *et al.* VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research* **50**, D898–D911 (2022).

46. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**, W29–W37 (2011).

47. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539 (2011).

48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).

49. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636–W641 (2019).

50. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **35**, 1547–1549 (2018).

51. Geneious Prime 2022. (https://www.geneious.com).

52. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

53. Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research* **33**, W557–W559 (2005).

54. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

55. Zhang, C., Freddolino, P. L. & Zhang, Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Research* **45**, W291–W299 (2017).

56. Roy, A., Yang, J. & Zhang, Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research* **40**, W471–W477 (2012).

57. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research* **13**, 2178–2189 (2003).

58. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**, 1160–1166 (2019).

59. Kuraku, S., Zmasek, C. M., Nishimura, O. & Katoh, K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research* **41**, W22–W28 (2013).

60. Wang, J. *et al.* iCn3D: From Web-Based 3D Viewer to Structural Analysis Tool in Batch Mode. *Frontiers in Molecular Biosciences* **9**, (2022).

61. Bliven, S. E., Bourne, P. E. & Prli, A. Detection of circular permutations within protein structures using CE-CP. *Bioinformatics* **31**, 1316–1318 (2015).

62. Li, Z., Jaroszewski, L., Iyer, M., Sedova, M. & Godzik, A. FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research* **48**, W60–W64 (2020).

63. Ma, J. & Wang, S. Algorithms, Applications, and Challenges of Protein Structure Alignment. in 121–175 (2014). doi:10.1016/B978-0-12-800168-4.00005-6.

64. Shindyalov, I. N. & Bourne, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering Design and Selection* **11**, 739–747 (1998).

65. Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *Journal of Molecular Biology* **147**, 195–197 (1981).

66. Ye, Y. & Godzik, A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19**, ii246–ii255 (2003).

67. Zhang, Y. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research* **33**, 2302–2309 (2005).

68. Busam, R.D., Arrowsmith, C., Berglund, H., Collins, R., Edwards, A., Ericsson, U.B., Flodin, S., Flores, A., Hammarstrom, M., Holmberg, S.L., Johansson, I., Karlberg, T., Kotenyova, T., Moche, M., Nilsson, M.E., Nordlund, P., Nyman, T., Ogg, D., Sagemark, J., Sundstrom, M., Uppenberg, J., Van Den Berg, S., Weigelt, J., Persson, C., Thorsell, A.G., Hallberg, B.M., Structural Genomics Consortium (SGC). (2007) Structure of human Inositol 1,3,4-trisphosphate 5/6-kinase. To be published. .

69. Gräslund, S. *et al.* The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expression and Purification* **58**, 210–221 (2008).

70. Kelly, S., Ivens, A., Manna, P. T., Gibson, W. & Field, M. C. A draft genome for the African crocodilian trypanosome Trypanosoma grayi. *Scientific Data* **1**, 140024 (2014).

71. Kelly, S. *et al.* An Alternative Strategy for Trypanosome Survival in the Mammalian Bloodstream Revealed through Genome and Transcriptome Analysis of the Ubiquitous

Bovine Parasite Trypanosoma (Megatrypanum) theileri. *Genome Biology and Evolution* **9**, 2093–2109 (2017).

72. Omasits, U., Ahrens, C. H., Müller, S. & Wollscheid, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30**, 884–886 (2014).

73. Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research* **47**, D351–D360 (2019).

74. Laha, D., Portela-Torres, P., Desfougères, Y. & Saiardi, A. Inositol phosphate kinases in the eukaryote landscape. *Advances in Biological Regulation* **79**, 100782 (2021).

75. Adl, S. M. *et al.* Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology* **66**, 4–119 (2019).

76. Hashimoto, M. *et al.* Inositol 1,4,5-trisphosphate receptor regulates replication, differentiation, infectivity and virulence of the parasitic protist *Trypanosoma cruzi*. *Molecular Microbiology* **87**, 1133–1150 (2013).

77. Centers for Disease Control and Prevention. Parasites - American Trypanosomiasis: Antiparasitic Treatment. (2021).

78. Hasslocher-Moreno, A. M. *et al.* Safety of benznidazole use in the treatment of chronic Chagas' disease. *Journal of Antimicrobial Chemotherapy* **67**, 1261–1266 (2012).

79. Jackson, Y., Wyssa, B. & Chappuis, F. Tolerance to nifurtimox and benznidazole in adult patients with chronic Chagas' disease. *Journal of Antimicrobial Chemotherapy* **75**, 690–696 (2020).

80. Rodríguez Arce, E. *et al.* Aromatic amine N-oxide organometallic compounds: searching for prospective agents against infectious diseases. *Dalton Transactions* **44**, 14453–14464 (2015).

81. Mosquillo, M. F. *et al. Trypanosoma cruzi* biochemical changes and cell death induced by an organometallic platinum-based compound. *Chemical Biology & Drug Design* **92**, 1657–1669 (2018).

82. Mosquillo, M. F. *et al.* Effect of a new anti-T. cruzi metallic compound based on palladium. *BioMetals* **31**, 961–974 (2018).

83. Mosquillo, M. F. *et al.* Comparative high-throughput analysis of the *Trypanosoma cruzi* response to organometallic compounds. *Metallomics* **12**, 813–828 (2020).

84. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

85. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

86. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).

87. SRA-Toolkit 2.9.0-centos_linux64. BEAR Applications. https://bear-apps.bham.ac.uk/applications/2018b/SRA-Toolkit/2.9.0-centos_linux64/ .

88. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525–527 (2016).

89. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* **14**, 687–690 (2017).

90. Yao, P. *et al.* Evidence for a direct cross-talk between malic enzyme and the pentose phosphate pathway via structural interactions. *Journal of Biological Chemistry* **292**, 17113–17120 (2017).

91. Martin, K. L. & Smith, T. K. The myo-inositol-1-phosphate synthase gene is essential in Trypanosoma brucei. *Biochemical Society Transactions* **33**, 983 (2005).

92. Ullu, E., Tschudi, C. & Chakraborty, T. RNA interference in protozoan parasites. *Cellular Microbiology* **6**, 509–519 (2004).

93. Taylor, M. C. & Kelly, J. M. pTcINDEX: a stable tetracycline-regulated expression vector for Trypanosoma cruzi. *BMC Biotechnology* **6**, 32 (2006).

94. Gonzalez, S. N., Valsecchi, W. M., Maugeri, D., Delfino, J. M. & Cazzulo, J. J. Structure, kinetic characterization and subcellular localization of the two ribulose 5-phosphate epimerase isoenzymes from Trypanosoma cruzi. *PLOS ONE* **12**, e0172405 (2017).

95. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281–2308 (2013).

96. Lander, N., Cruz-Bustos, T. & Docampo, R. A CRISPR/Cas9-riboswitch-Based Method for Downregulation of Gene Expression in Trypanosoma cruzi. *Frontiers in Cellular and Infection Microbiology* **10**, (2020).

97. Lander, N., Chiurillo, M., Vercesi, A. & Docampo, R. Endogenous C-terminal Tagging by CRISPR/Cas9 in Trypanosoma cruzi. *BIO-PROTOCOL* **7**, (2017).

98. Lander, N., Li, Z.-H., Niyogi, S. & Docampo, R. CRISPR/Cas9-Induced Disruption of Paraflagellar Rod Protein 1 and 2 Genes in Trypanosoma cruzi Reveals Their Role in Flagellar Attachment. *mBio* **6**, (2015).

99. Docampo, R. Molecular parasitology in the 21st Century. *Essays in Biochemistry* **51**, 1–13 (2011).

100. Taylor, M. C., Huang, H. & Kelly, J. M. Genetic Techniques in Trypanosoma cruzi. in 231–250 (2011). doi:10.1016/B978-0-12-385863-4.00011-3.

101. Cruz-Bustos, T., Ramakrishnan, S., Cordeiro, C. D., Ahmed, M. A. & Docampo, R. A Riboswitch-based Inducible Gene Expression System for *Trypanosoma brucei*. *Journal of Eukaryotic Microbiology* **65**, 412–421 (2018).

102. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology* **305**, 567–580 (2001).

103. Minning, T. A., Weatherly, D. B., Atwood, J., Orlando, R. & Tarleton, R. L. The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. *BMC Genomics* **10**, 370 (2009).

104. Norman, K. L. *et al.* Inositol polyphosphates regulate and predict yeast pseudohyphal growth phenotypes. *PLOS Genetics* **14**, e1007493 (2018).

105. Payne, W. E. & Fitzgerald-Hayes, M. A mutation in PLC1, a candidate phosphoinositide-specific phospholipase C gene from Saccharomyces cerevisiae, causes aberrant mitotic chromosome segregation. *Molecular and Cellular Biology* **13**, 4351–4364 (1993).

106. Bosch, D. & Saiardi, A. Arginine Transcriptional Response Does Not Require Inositol Phosphate Synthesis. *Journal of Biological Chemistry* **287**, 38347–38355 (2012).

107. Resnick, A. C. *et al.* Inositol polyphosphate multikinase is a nuclear PI3-kinase with transcriptional regulatory activity. *Proceedings of the National Academy of Sciences* **102**, 12783–12788 (2005).

108. Odom, A. R., Stahlberg, A., Wente, S. R. & York, J. D. A Role for Nuclear Inositol 1,4,5-Trisphosphate Kinase in Transcriptional Control. *Science (1979)* **287**, 2026–2029 (2000).

109. Azevedo, C. & Saiardi, A. Extraction and analysis of soluble inositol polyphosphates from yeast. *Nature Protocols* **1**, 2416–2422 (2006).