CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF TRICORN INTERACTING FACTOR 3 IN *TRYPANOSOMA BRUCEI BRUCEI* AS POTENTIAL DRUG TARGETS AGAINST AFRICAN TRYPANOSOMIASIS

FLORENCE ATIENO NG'ONG'A

DOCTOR OF PHILOSOPHY

(Biochemistry)

JOMO KENYATTA UNIVERSITY OF

AGRICULTURE AND TECHNOLOGY

2018

Characterization and Functional analysis of tricorn interacting factor 3 in *Trypanosoma brucei brucei* as potential drug targets against African Trypanosomiasis

Florence Atieno Ng'ong'a

A Thesis Submitted in Fulfilment for the Degree of Doctor of Philosophy in Biochemistry in the Jomo Kenyatta University of Agriculture and Technology

2018

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature …….. Date ……..

Florence Atieno Ng'ong'a

This thesis has been submitted for examination with our approval as University Supervisors.

Signature …….. Date ……..

Dr. Steven Rueben Ger Nyanjom, PhD

JKUAT, Kenya

Signature …….. Date ……..

Prof. Fred Wamunyokoli, PhD

JKUAT, Kenya

DEDICATION

I dedicate this thesis to God, my source of inspiration, knowledge and understanding. I also dedicate this thesis to my uncle, Mr. Samuel Nyapola Ochanda whose moral and financial support, made me who I am today. To my husband Mark Maosa, whose encouragement has been a great source of strength. To my children Janet Kerubo, Alfred Maosa and Emmanuel Myles who have been affected in every way possible by this quest. Last but not least, to my siblings for moral support and prayers. God bless you all.

ACKNOWLEDGEMENTS

I would like to give thanks to God for giving me the strength and ability to complete this research work. I wish to acknowledge the support of several people and organizations for making this thesis possible. Indeed, it is a pleasant aspect that I have the opportunity to express my gratitude to all of them. Special thanks to my supervisors Dr. Steven Ger Nyanjom and Prof. Fred Wamunyokoli for their expert counsel in the design, execution and thesis write-up. My deepest gratitude to Dr. Dan Masiga for giving me a chance to work in the Molecular Biology and Bioinformatics Unit (MBBU) in ICIPE, Kasarani, Kenya. I would also like to thank the MBBU staff especially Dr. Joel Bargul and Collins Omogo for their support during the molecular work. Many thanks to the Department of Biochemistry staff members for their support during the course of undertaking this research work. I would also like to thank the Jomo Kenyatta University of Agriculture and Technology, Research Production and Extension division (JKU/2/4/RP/139), National Commission for Science, Technology and Innovation (NACOSTI/RCD/ST and I 5th CALL PhD/145) and Africa-ai-Japan Project for funds. My immeasurable gratitude goes to my uncle Mr. Samuel Nyapola Ochanda for his selfless sacrifice, unrelenting support and his belief in my abilities. Last but not least, my husband, Mr. Mark Maosa for his unconditional love and I will forever be grateful to my children (Janet Kerubo, Alfred Maosa and Emmanuel Myles) for their patience during the course of this work despite being too young to understand. May God bless you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

AAT	Animal African Trypanosomiasis
APOL-1	Apolipoprotein L-1
PSI-BLAST	Position Specific Iterated Basic Local Alignment Search Tool
FBS	Fetal Bovine Serum
BSF	Bloodstream form
Bam HI	Type II restriction endonuclease from *Bacillus amyloli*
Bp	Base pairs
C1 and C2	Catalytic domain 1 and Catalytic domain 2
CDART	Conserved Domain Architecture Retrieval Tool
CD8	Cluster of Differentiation 8
cDNA	Complementary Deoxyribonucleic acid
CSF	Cerebrospinal fluid
DNA	Deoxyribonucleic acid
ds RNA	Double-stranded Ribonucleic acid
EcoRI	Type II restriction endonuclease from *Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ESPs	Expressed/Secreted proteins
GPI	Glycosylphosphatidylinositol
HA	Haemagglutination assays
HAT	Human African Trypanosomiasis
HMI-9	Hirumi Medium-9
HpHbR	Haptoglobin-hemoglobin receptor
IC	Intermediate chromosomes

IFA	Immunofluorescent assays
IFN-gamma	Interferon gamma
IPTG	Isopropyl β-D-1-thiogalactopyranoside,
LB	Luria-Bertani
Mb	Megabase
MC	Megabase Chromosome
MEME	Multiple Expectation maximization for Motif Elicitation
NCBI	National Center for Biotechnology Information
NR	Non-redundant
PBS	Phosphate Buffered Saline
***Pf*A-M1**	*Plasmodium falciparum* Aminopeptidase, M1 family
Qpcr	Quantitative real time polymerase chain reaction
RBC	Red Blood Cell
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RT-Qpcr	Reverse transcription real time polymerase chain reaction
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide gel electrophoresis
SRA	Serum resistance-associated protein
TAE	Tris acetate EDTA
TERT	Telomerase Reverse Transcriptase
TLFs	Trypanolytic factors
TLTF	T lymphocyte triggering factor
VSG	Variable Surface Glycoprotein
X-GAL	5-Bromo-4-chloro-3-indolyl beta-D-galactoside

XhoI	Type II restriction endonuclease from *Xanthomonas holcicola*
YT	Yeast-Tryptone
YTA	Yeast-Tryptone Ampicillin

ABSTRACT

Trypanosomes are protozoans causing African trypanosomiasis, a neglected tropical disease in Africa affecting humans and animals. Despite advancement in African Trypanosomiasis research, the disease continues to threaten millions of people and animals in Sub-Saharan Africa. Control methods have focused on the use of drugs which have adverse effects and vector control methods which have proved to be ecologically unsustainable while vaccines are still not available due to antigenic variation. Studies have shown that the pathogenesis of *Trypananosoma brucei brucei* infections is parasite driven and various classes of *Trypananosoma brucei brucei* peptidases have been studied and implicated as virulence factors. Proteasomal degradative pathway is the core non-lysosomal protein degradation machinery in *Trypanosoma brucei brucei* and together with its accessory proteins, they are responsible for cell quality control. Tricorn protease and its interacting factors, F1, F2, and F3 act *in vivo* downstream of the proteasome and have been implicated in cell maintenance, defense, growth and development in many pathogenic organisms. The aim of this study was to determine the presence and investigate role of tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei* endocytic system. A combination of bioinformatics programs including PSI-BLAST, Hidden Markov Models and orthology clustering were used to identify tricorn interacting factor 3 orthologs. Molecular characterization involved generation of transcripts through PCR and mRNA detection by reverse transcription with gene specific primers. Real time PCR using SYBR green detection method was used to determine the expression levels of genes encoding tricorn interacting factor 3 orthologs in the bloodstream form of *Trypanosoma brucei brucei*. Cloning and expression of tricorn interacting factor 3 orthologs was done and expressed proteins were analyzed through SDS-PAGE. Functional analysis through RNAi involved ligation of the dsRNA targeting Tb927.11.3570 and Tb927.3.4750/90 between the two T7 promoters of linearized p2T7-177 vector to generate recombinant plasmids for transfection. Transfection was done through electroporation and RNAi was induced by tetracycline disodium. Relative expression levels of Tb927.11.3570 and Tb927.3.4750/90 in transgenic parasites was determined through qRT-PCR and parasite densities were determined by Neubuaer chamber. A total of 24 hits were identified and were categorized as proteins with isolated tricorn-like domains (21) and tricorn interacting factor 3 orthologs (3). Two (2) of the tricorn interacting factor 3 orthologs appeared to represent a gene duplication process. The 21 proteins contained isolated tricorn protease domains mainly; tricorn protease N-terminal domain, tricorn protease domain 2 and PDZ domains which are known to have no catalytic activity. Further analysis of the tricorn interacting factor 3 orthologs revealed conservation of the zinc binding motif ({HEXXH (18X) E}, the GXMEN motif specific for M1 family as well as the N-terminal substrate anchoring residue, glutamate and the proton acceptor, tyrosine. The protein-protein interactions network as predicted by STRING database revealed interactions with proteasome core complex. The two (2) monomer models built through SWISS model showed 91.1% and 88.1% of residues in most favoured regions

for Tb927.11.3570 and Tb927.3.4750/90 respectively indicating good models. Polymerase chain reaction generated amplicons approximately 285 bp and 325 bp for Tb927.11.3570 and Tb927.3.4750/90 respectively. Real time PCR revealed significant expression of both Tb927.11.3570 and Tb927.3.2750/90 in the bloodstream form of *Trypanosoma brucei brucei* with p-value of 0.009 and 0.012 respectively at 95% confidence interval. SDS-PAGE of the expressed proteins indicated presence of the expressed proteins. Through paired samples t-test, the qRT-PCR results revealed significant differential expression of Tb927.11.3570 and Tb927.3.4750/90 in tetracycline induced RNAi mutants and control *T. b. brucei* with p-values of 0.007 and 0.003 respectively at 95% confidence interval. However, the expression of the reference gene (actin) in tetracycline induced RNAi mutants and control *T. b. brucei* was insignificant with $p > 0.05$ confirming the effect of RNAi on the test genes. Using Livak method, the tetracycline induced RNAi mutants down regulated the expression of Tb927.11.3570 by 2.97 and Tb927.3.4750/90 by 2.79 fold. This represented 57% for Tb927.11.3570 and 48% knockdown for Tb927.3.4750/90. The reduction in expression of Tb927.11.3570 and Tb927.3.4750/90 resulted in decreased growth rate suggesting a possible role in parasite growth. Therefore, the study recommends design of molecules that may inhibit the activity of tricorn interacting factor 3 orthologs and further investigate their potential as drug target molecules against African trypanosomiasis.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

African Trypanosomiasis is a vector-borne disease caused by protozoan of the genus *Trypanosoma,* transmitted mostly by bites from infected tsetse flies (*Glossina sp*.) (Steverding, 2008) (Stuart *et al.,* 2008). There are two forms of the disease; Human African Trypanosomiasis (HAT) or sleeping sickness and Animal African Trypanosomiasis (AAT) or nagana (Steverding, 2008). Human African Trypanosomiasis is majorly found in Sub-Saharan Africa and is caused by two sub-species of *Trypanosoma brucei; T. b. gambiense* and *T. b. rhodesiense* (Brun *et al.,* 2010). These two species are considered different since they show different epidemiological and clinical patterns and thus are managed through different therapeutic agents (Kato *et al.,* 2016). *Trypanosoma b. rhodesiense* is found in East and Southern Africa and causes acute, rapidly progressive disease known as rhodesiense HAT (Brun *et al.,* 2010). *Trypanosoma b. gambiense* is found in West and Central Africa and causes chronic form of the disease known as gambiense HAT, an anthroponotic disease responsible for the majority of the reported HAT cases (Simarro *et al.,* 2012) (Franco *et al.,* 2014). Gambiense HAT has also been reported in Uganda (Fèvre et al. 2005).

Animal African Trypanosomiasis is caused by various trypanosome species including *T. b. brucei, T. vivax, T. congolense,* which have a wide host range within the domestic animals while *T. godfreyi, T. simiae* and *T. suis* seem to occur in pigs (Murilla *et al.,* 2014). *Trypanosoma b. brucei* is susceptible to lysis by the human Trypanosome Lysis Factor-1 but is genotypically similar to the human pathogenic forms; *T. b. gambiense* and *T. b. rhodensiense* thus is a good experimental model for both human and animal infection studies (Franco *et al.,* 2014). Lysis of the parasites occurs after uptake of the TLFs into the parasite when the lytic component Apolipoprotein L-1 (APOL-1) is taken

up into endosomal and lysosomal membranes, causing osmotic swelling and lysis by forming cation selective pores in membranes (Molina-Portela *et al*., 2005). The human infective subspecies show resistance against lysis by human TLFs. In *T. b. gambiense*, there is reduced TLF binding or uptake due to reduced expression of HpHbR (Bullard *et al.*, 2012), while in *T. b. rhodesiense* resistance of APOL1 is achieved by expression of a serum resistance-associated protein (SRA) (Kieft *et al.*, 2010).

Control of African trypanosomiasis through eradication of the insect vector (*Glossina sp*.) is covering just about 1.3% of the tsetse infested area hence tsetse infestation is still significant (Meyer *et al.,* 2016). Various approaches have been used to control tsetse population. These include sequential aerosol technique, stationary attractive devices, live bait technique and sterile insect techniques (Vreysen *et al.,* 2012). Tsetse population has also been controlled through localized (field-to-field) or area-wide approaches where the latter aims to eradicate the entire tsetse population in a geographical area (Vreysen *et al.,* 2012). Other control measures include use of trypanocidal compounds, chemotherapy and chemoprophylactic agents (Steverding, 2010). Unfortunately, the efficacy of many of these compounds is being impeded by the inherent ability of trypanosomes to undergo antigenic variation *(*Anene *et al.,* 2001) (Brun *et al.,* 2001) (Horn 2014). There is also the exploitation of trypanotolerant livestock such as the Djallonke sheep, N'Dama cattle and West African Dwarf goats, which withstand a trypanosomiasis challenge and still remain productive without any form of therapy (Yaro *et al.,* 2016). Vaccine development still remains a challenge in trypanosomiasis due to the parasite's efficient immune-evasion mechanism (Field *et al.,* 2017) (Aksoy *et al.,* 2017).

Since trypanosomes are extracellular parasites, the success of the infection process depend on the interactions between parasites' secreted proteins and host immune cells which play a key role in the disease course and immunopathology (Bossard *et al.,* 2013). Peptidase activities have also been implicated in many aspects of the physiology of the parasite as well as in pathogen–host interface and pathogenesis thus are

considered good drug targets (Klemba and Goldberg 2002). Studies have shown that the pathogenesis of *T. b. brucei* infections are parasite driven (Morrison, 2011). Different classes of *T. b. brucei* proteases have been implicated in pathogenesis (Troeberg *et al.,* 1996, Morty *et al.,* 2001, Bastos *et al.,* 2010). Aminopeptidases have also been shown to be key regulators in many intracellular and extracellular events and thus are attractive pharmaceutical targets (Drag *et al.,* 2010).

Parasite growth, differentiation, dissemination through host tissues and infection of the mammalian cells are highly dependent on proteolytic activities (Klemba and Goldberg, 2002). The genome of *T. b. brucei* contains many genes encoding peptidases which could be considered as virulence factors. Some of the peptidases that have been studied include Cathepsin B and L which are associated with parasite development, host protein degradation and iron acquisition thus aid in disease progression (Abdulla *et al.,* 2008) (O'Brien *et al.,* 2008). Others include oligopeptidase B and prolyl oligopeptidase which play a role in parasite entry into mammalian cells (Morty *et al.,* 2001), (Bastos *et al.,* 2010) (Canning *et al.,* 2013). Metallocarboxypeptidase, M32 family (TbMCP-1) also acts on bradykinin and angiotensin I thus may have a role in the pathogenesis of African Trypanosomiasis (Frasch *et al.,* 2012).

Trypanosoma b. brucei cell cycle progression has also been shown to be dependent on proteasome activity (Mutomba *et al.,* 1997) (Muñoz *et al.,* 2015). Tricorn protease and its interacting factors, F1, F2, and F3 degrade proteasomal products to free amino acids for other metabolic processes (Tamura *et al.,* 1998). Tricorn protease acts as a carboxypeptidase with di- and tripeptidase activity while its interacting factors are aminopeptidases with different substrate specificities (Groll *et al.,* 2005). Tricorn interacting factor 1 (F1) is a proline iminopeptidase belonging to the α/β hydrolase superfamily and hydrolyzes a wide spectrum of substrates mostly hydrophobic residues such as alanine, proline, phenylalanine, leucine, tyrosine and glycine in the P1 position (Goettig *et al.,* 2002). Tricorn interacting factor 1 (F1) has a molecular weight of 33.5 kDa and exhibits 14 % sequence homology to the beta propeller catalytic domain of

prolyl oligopeptidase and thus based on its structure, it docks into the six-bladed beta-propeller domain of tricorn protease (Brandstetter *et al.,* 2001) (Goettig *et al.,* 2002). However, Tricorn interacting factor 2 (F2) and Tricorn interacting factor 3 (F3) are closely related zinc aminopeptidases with a molecular weight of 89 kDa and share 56% sequence identity (Kyrieleis *et al.,* 2005). Tricorn interacting factor 2 (F2) displays a broader specificity against neutral, hydrophobic and basic residues while F3 displays a much narrower substrate specificity with strong preference for glutamate (Kyrieleis *et al.,* 2005). Tricorn interacting factor 2 and F3 belong to clan MA of M1 family of aminopeptidases, which have been shown to be essential to the survival of a range of pathogenic organisms thus are good targets for drug discovery (Drinkwater *et al.,* 2017). Therefore, identification and characterization of tricorn interacting factors would present putative chemotherapeutic targets which can be exploited in the control of African Trypanosomiasis.

1.2 Statement of the Problem

African trypanosomiasis remains a constant and persistent threat to the health of humans and livestock (Brun *et al.,* 2010). This situation is aggravated by the neglect African trypanosomiasis has experienced among the donor community (Rutto *et al.,* 2013). The haemo-lymphatic phase is usually difficult to diagnose since it is usually long, relatively asymptomatic thus active screening of the population at risk is required for early stage identification thus reduce transmission (Simo *et al.,* 2014). Screening for potential infection involves using serological tests (only available for *T. b. gambiense*), checking for clinical signs (swollen cervical glands) and staging which can be done to determine the state of disease progression. Diagnosis must be made as early as possible before the neurological stage in order to avoid complicated, difficult and risky treatment procedures (Simo *et al.,* 2014).

Exhaustive screenings require a major investment in human/animal and material resources. In Africa, such resources are often scarce, particularly in remote areas where

the disease is mostly found. As a result, many infected individuals may die before they can ever be diagnosed and treated (Lomo *et al.,* 1997) (Murilla *et al.,* 2014). In the second stage (neurological phase), the parasites cross the blood-brain barrier to infect the central nervous system and without treatment, African Trypanosomiasis is fatal (Steverding, 2008).

Current drugs used for treatment of humans include pentamidine, suramin, melarsoprol, eflornithine and nifurtimox-eflornithine combination therapy (NECT) while drugs of choices for trypanosomiasis prevention and treatment in animals are Berenil, Samorin and Homidium. These drugs are not always successful for HAT and AAT treatment because of limitations from drug efficacy, resistance, and high toxicity (Simarro *et al.,* 2011) (Simarro *et al.,* 2012). Suramin provokes certain undesirable effects, in the urinary tract and allergic reactions, Melarsoprol causes encephalopathic syndrome (Fggfdut and Boujqbsbtjujd, 2011). Eflornithine is less toxic than melarsoprol, however, the application regime is strict and difficult (Chappuis *et al.,* 2005), (Brun *et al.,* 2010). A combination treatment of nifurtimox and eflornithine which was introduced 2009 is not effective for *T. b. rhodesiense* (Brun *et al.,* 2010) (Kappagoda *et al.,* 2011). There has also been an increase in drug resistance in domestic animals especially in the Central African foci (Steverding, 2008). It is also difficult to control movement to wild animal reservoirs, which could transmit to both livestock and humans (Simarro *et al.,* 2012).

The disability adjusted life years (DALY) lost due to HAT by the year 2000 was approximately 1.5 million with estimated cost of US$150 per DALY averted (Fèvre *et al.,* 2008). While out of the 178 million cattle in Africa, 44.7 million are at risk of Animal African Trypanosomiasis. This translates to considerable losses in livestock production estimated to over US $ 4.5 billion per annum (Yaro *et al.,* 2016). Thus, nagana is a great threat to food security in endemic regions (Holt *et al.,* 2016). According to Kenya Tsetse and Trypanosomiasis Eradication Council (KENTTEC), the total tsetse infested area in Kenya is about 138,000 Km2, which translates to 38 out of

47 counties exposing eleven million people in endemic areas (Lake Victoria basin and the Mara- Serengenti ecosystem) at risk of sleeping sickness infection (KENTTEC Report, 2015). Tourism accounts for 21% of foreign exchange earnings and this sector is negatively affected by tsetse and trypanosomiasis (KENTTEC Report, 2015).

Several classes of *T. b. brucei* peptidases have been studied and their role in the parasite endocytic system determined. For example, cysteine peptidases have been detected by immunoprecipitation assays as antigens in trypanosome infections thus are involved in pathogenesis (Turk, 2006). Knockdown cysteine proteases expression by RNA interference is lethal in *T. brucei,* causing phenotypic defects (O'Brien *et al.,* 2008). Other peptidases essential in survival of *T. b. brucei* include oligopeptidase B which has been shown to inactivate atrial natriuretic factor, a peptide hormone present in plasma of trypanosome infected hosts (Morty *et al.,* 2001). Extensive studies have also been done on several classes of aminopeptidases in *T. b. brucei* exception of M1 family aminopeptidases, a family consisting of tricorn interacting factor 3. Leucine aminopeptidase (LAP), a member of M17 family of metallopeptidases is implicated in protein degradation associated with nutrient depletion (Timm *et al.,* 2017) and is also required for correct segregation of Kinetoplast (k) DNA (Peña-Diaz *et al.,* 2017). Surface metalloprotease (GP63) are involved in the developmental forms of *T. brucei* and possible host infection (Bangs *et al.,* 2001). Therefore, this study characterized tricorn interacting factor 3 in *T. b. brucei* as potential drug targets against African Trypanosomiasis.

1.3 Justification

Peptides for antigen presentation are generated through the degradation of proteasomal products (Muñoz *et al.,* 2015), thus parasite secreted oligopeptidases targeting such peptides would limit host immune response thus aid in immune evasion (Muñoz *et al.,* 2015). Tricorn interacting factor 3 are members of a diverse family of metalloenzymes, M1 aminopetidase, characterized by conserved structure and reaction specificity

(Drinkwater *et al.,* 2017). Their activity is important to parasite survival since the accumulation of un-degraded peptides may be cytotoxic (Dalal and Klemba 2007). Intracellular parasites are also known to utilize exogenously supplied amino acids for protein synthesis thus, are dependent on aminopeptidase activities to acquire amino acids (Skinner-Adams *et al.*, 2010) (Drinkwater *et al.*, 2017). Approaches targeting the disease causing factors of the parasite rather than the parasite itself could be key to overcoming antigenic variation in trypanosomes (Antoine-Moussiaux *et al.,* 2009). The structure and catalytic mechanism M1 aminopeptidases make them ideal candidates for the design of small-molecule inhibitors thus their utility as therapeutic targets for infectious human diseases (Drinkwater *et al.,* 2017).

Some drug discovery programs targeting M1 aminopeptidases have shown promise for the treatment of malaria (Drinkwater *et al.,* 2016). In *Plasmodium falciparum,* M1 aminopeptidases (*Pf*A-M1) are found in the food vacuole where they catalyze the release of amino acids from the highly sequence diverse peptides produced from α- and β- globin (Dalal and Klemba 2007). They are also involved in the final stages of the digestion of human haemoglobin (Dalal and Klemba, 2007). *Pf*A-M1 is thus essential for cell viability and knockout of *Pf*A-M1 activity resulted into cell death (McGowan *et al.,* 2009) (Mistry *et al.,* 2014). Hence, *Pf*A-M1 have been described as potential targets for new anti-malarials (Skinner-Adams *et al.,* 2010) (Drinkwater *et al.,* 2016) (Drinkwater *et al.,* 2017). In *Eimeria tenella,* aminopeptidase N, a homolog of tricorn interacting factor 3, aids in parasite development within the host (Gras *et al.,* 2014). *Neisseria meningitides* also have aminopeptidase N with structural resemblance to tricorn interacting factor 3 (Nocek, Mulligan, and Bargassa, 2008) and have been described as potential drug target for inhibition and development of novel therapeutic agents against *Neisseria meningitidis* (Węglarz-Tomczak *et al.,* 2013). RNAi of M1 aminopeptidases *Schistosoma mansoni* reduces parasite viability also attributed to reduced free amino acids necessary for novel protein synthesis (Mucha *et al.,* 2010). This study characterized tricorn interacting factor 3 orthologs in *T. b. brucei* and

recommends design of structure based inhibitors to further investigate their potential as drug targets against African Trypanosomiasis.

1.4 Objectives

1.4.1 General Objective

To determine the presence and carry out functional characterization of tricorn interacting factor 3 in *Trypanosoma brucei brucei* as potential drug targets against African Trypanosomiasis.

1.4.2 Specific Objectives

1. To determine the presence of tricorn interacting factor 3 in *Trypanosoma brucei brucei* genome through computational approaches
2. To characterize tricorn interacting factor 3 in the bloodstream form of *Trypanosoma brucei brucei*
3. To determine the effect of knockdown of the genes encoding tricorn interacting factor 3 in *Trypanosoma brucei brucei*

CHAPTER TWO

LITERATURE REVIEW

2.1 Biology of African Trypanosomes

2.1.1 Classification of Trypanosomes

Trypanosomes are unicellular parasitic protozoa belonging to the order Kinetoplastida, and genus *Trypanosoma* (Uilenberg, 1998). Trypanosomes have been further classified by modes of transmission into Stercoraria (American trypanosomes; *T. cruzi, T. theileri, T. lewisi*) and Salivaria (African trypanosomes). African trypanosome species belong to three different sub-genera: *Trypanozoon* (*T. brucei, T. evansi, T. equiperdum*), *Dutonella* (*T. vivax, T. uniforme*) and *Nanomonas* (*T. congolense, T. simiae, T. godfreyi*) (Uilenberg, 1998) (Antoine-Moussiaux *et al.,* 2009). The *T. brucei* group of trypanosomes mostly invade tissues (humoral) whereas, *T. congolense* and to a lesser extent *T. vivax* predominantly restrict themselves to the blood circulation. The species *T. equiperdum,* which causes dourine in equines, is thought to be closely related to *T. evansi*, a causative agent of surra in equines (Sykes *et al.,* 2014).

2.1.2 Life cycle of *Trypanosoma brucei brucei*

Trypanosoma brucei has a two host life cycle; an insect host (tsetse fly) and a mammalian host (Figure 2.1). *Trypanosoma brucei* metacyclic trypomastigotes present in the saliva of an infected tsetse fly get transmitted during a blood meal to the mammalian host (Steverding, 2008). In the mammalian host, metacyclic trypomastigotes differentiate into bloodstream trypomastigotes (long, slender form) and proliferate at the site of infection for a few days, forming a trypanosomal chancre. From there, the fast dividing bloodstream form trypomastigotes spread to the lymph nodes and the bloodstream where they can also infect multiple organs (Steverding, 2008). The long slender bloodstream form rapidly multiplies increasing the parasitemia in the blood

of the host. The parasites then differentiate into non-dividing short stumpy trypomastigotes which can be taken up by the tsetse fly during a blood meal. In the midgut of the tsetse fly the parasites differentiate into procyclic trypomastigotes and proliferate into epimastigotes which leave the midgut (Steverding, 2008). The epimastigote can differentiate asymmetrically to form new epimastigotes and infective metacyclic trypomastigotes, which migrate to the salivary gland where they can be transmitted to other mammalian hosts during a blood meal thus allowing tsetse flies to remain infective over their relatively long life span (Steverding, 2008) (Rotureau *et al.*, 2012).

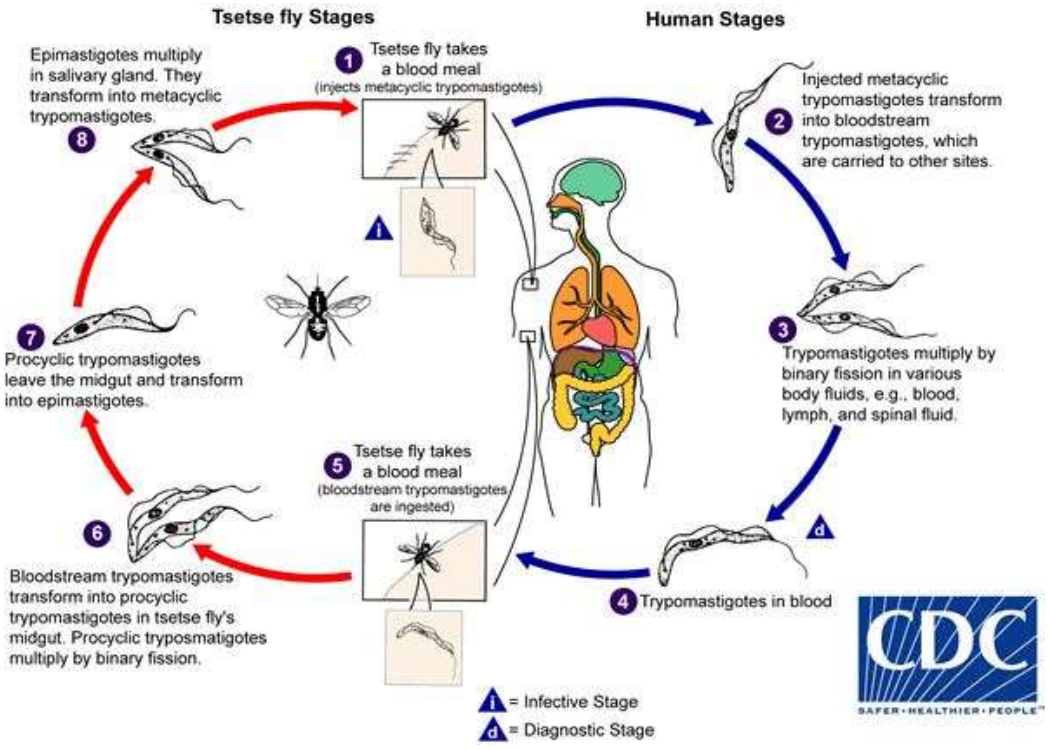


Figure 2.1: Representation of the two host life cycle of *Trypanosoma brucei* in the tsetse fly and human host (WHO, 2016).

Different strategies have been put in place for control of African Trypanosomiasis. They include the removal of susceptible populations from risk areas, vector control methods, mass chemoprophylaxis programmes, active surveillance and treatment

(Fevre *et al.*, 2008) (Franco *et al.*, 2014). Since animals are reservoirs thus are integral to the spread of the disease to humans, limiting their movement could be essential in preventing the spread and new outbreaks of the disease (Franco *et al.*, 2014).

2.1.3 Morphology, cellular and molecular biology of *Trypanosoma brucei brucei*

Trypanosoma b. brucei is approximately 20 micrometers in length with a single central nucleus and a single flagellum originating at the kinetoplast and joined to the body by an undulating membrane (Figure 2.2 and Figure 2.3). The outer surface of the organism is densely coated with a layer of glycoprotein, the variable surface glycoprotein (VSG) (Rudenko, 2012). They have the typical cell components of a eukaryote: nucleus, microtubules (cytoskeleton and flagellum), endoplasmic reticulum, golgi apparatus and a single mitochondrion. They also have a kinetoplast, which is an extension of the single mitochondrion and contains the mitochondrial genome (Steverding, 2008). The African trypanosomes show two morphologic forms: trypomastigote and epimastigote (Steverding, 2008).

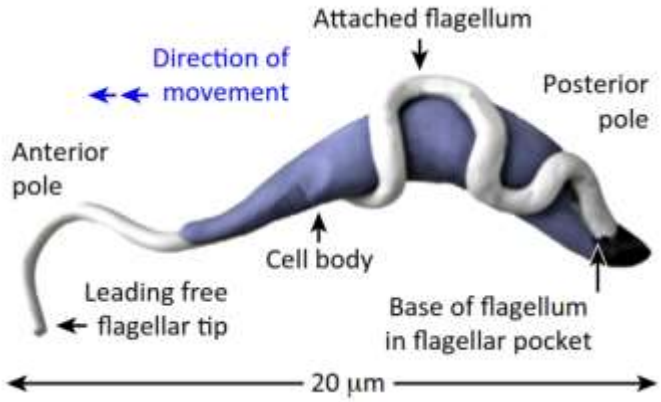


Figure 2.2: An illustration of *Trypanosoma brucei brucei* cell architecture (Bentivoglio and Kristensson, 2014).

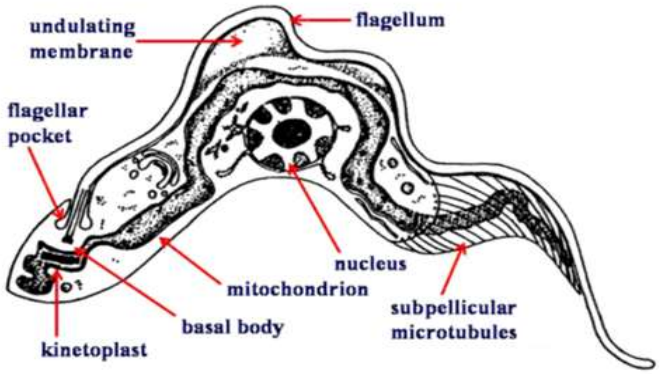


Figure 2.3: A cross-sectional view of *Trypanosoma brucei brucei* (Bentivoglio and Kristensson, 2014).

The trypanosome are typically elongated in shape and have a highly polarized structure defined by sub-pellicular microtubule cytoskeleton, with all microtubules running in an anterior to posterior direction apart from a specific microtubule quartet which runs in opposite direction (Farr and Gull, 2012). The flagellum, containing a 9+2 axoneme and a paraflagellar rod exits the cell with the basal body which is attached to the kinetoplast. The flagellar pocket is the sole point of endo- and exocytosis (Farr and Gull, 2012).

2.1.4 Genome of *Trypanosoma brucei brucei*

The *T. b. brucei* genome contains a nuclear and a kinetoplast genome totalling to 35 megabases (Mb) per haploid genome (Berriman *et al.*, 2005). The genome is composed of three classes of chromosomes, which are distinguished by size. It contains 11 pairs of diploid megabase chromsosomes (MCs), which have been sequenced and published (Berriman *et al.*, 2005), a variable number of intermediate chromosomes (ICs) and between 50-100 mini-chromosomes (Melville *et al., 2000). Trypanosome b. brucei* has approximately 1,700 specific genes (Berriman *et al.*, 2005) and is predominantly diploid, though meiotic events do occur indicating a haploid life stage at some point in the tsetse fly (Peacock *et al.*, 2011).

2.2 *Trypanosoma brucei brucei* virulent factors

Trypanosomes are highly adapted for life in the hostile environment of the mammalian host and therefore have developed various adaptations to their cell biology to facilitate their survival. These include variant surface glycoprotein (VSG) switching, B-cell mitogen production, expression of T Lymphocyte triggering factor and enzymes (Bezie *et al.,* 2014). African trypanosomes have been shown to express certain proteases during infective stages of the parasite's life cycle and are suspected to act as pathogenic factors (Bezie *et al.,* 2014).

2.2.1 Expressed / Secreted proteins

Proteases have a number of functions in the African trypanosomes including post-translational processing of proteins, degradation of the proteins involved in the morphological changes occurring during their life cycle and hydrolysis of the host proteins for nutritional and invasion purposes thus are essential for survival of the parasite in their mammalian host (Troeberg *et al.,* 1996). Trypanosomes have very distinctive features and organelles and a streamlined metabolism consistent to their parasitic life style (Smith and Bütikofer, 2010). The trypanosome secretome constitutes an important class of proteins that control and regulate a multitude of biological and physiological processes, thus making it a clinically relevant source for biomarkers and therapeutic target discoveries (Ranganathan and Garg, 2009). A class of particularly interesting proteins are the expressed/secreted proteins (ESPs) which are specifically secreted by parasites. They are involved in various aspects of pathogenesis with studies showing that the secretome of *T. b. brucei* inhibits maturation of dendritic cells and their ability to induce lymphocytic allogenic responses (Geiger *et al.,* 2010).

Trypanosoma b. brucei express different enzymes such as sialidases which are glycosylphosphatidylinositol (GPI) membrane anchored protein. In the bloodstream forms, sialidases cause red blood cell (RBC) surface alterations thus, leading to their subsequent phagocytosis (Buratai *et al.,* 2006). Cysteine proteases which include

papain and caspases are expressed during the infective stages of the parasites' life cycle thus may act as pathogenic factors in the mammalian host, where they also trigger prominent immune responses (Troeberg *et al.,* 1999). Serine oligopeptidase catalyses the degradation of several peptide hormones, such as neurotensin or atrial natriuretic peptide thus the release of oligopeptides into the bloodstream during parasite lysis phases (Hedstrom, 2002) (Morty *et al.,* 2005). Threonine proteases which include the catalytic subunits of the proteasome (Groll *et al.,* 1997), aspartic acid proteases (Szecsi, 1992) and metalloproteases (Neil, 1993) are involved in many biological processes.

2.2.2 B-Cell mitogen production

Trypanosoma b. brucei also produce a B-cell mitogen which allows generalized differentiation and expansion of clones of B-cells thus pre-empts the subsequent specific responses to new antigens (Vincendeau *et al.,* 1996). Serum-immunoglobulins, especially IgM, increase and contain not only specific anti-parasite antibodies but also antibodies reacting with other non-parasite antigenic determinants, including those of host tissue. The stimulation of non- specific and disordered immunoglobulin response could help the parasite to survive (Vincendeau *et al.,* 1996). Trypanosomes also express T lymphocyte triggering factor (TLTF), which triggers CD8 (+) T lymphocytes to proliferate and to secrete interferon (IFN)-gamma. T lymphocyte triggering factor (TLTF) is localized to small vesicles that are found primarily at or near the flagellar pocket, the site of secretion in trypanosomes. T lymphocyte triggering factor (TLTF) is secreted by trypanosomes to modulate the cytokine network of the host immune system for the benefit of the parasite (Vaidya *et al.,* 1997). Other biological processes described in trypanosome involved in virulence include antigenic variation (Horn, 2014), glycosylphosphatidylinositol (GPI) anchoring for variable surface glycoproteins (VSGs) (Ferguson and Williams, 1988), trans-splicing (Sutton and Boothroyd, 1986) and RNA editing (Benne *et al.,* 1986).

2.3 *Trypanosoma brucei* diagnosis

Trypanosoma brucei infection is determined by clinical signs and laboratory diagnosis. The available laboratory diagnosis of trypanosomes is divided into three main methods: parasitological, serological, and molecular (Wastling and Welburn, 2011) and are used for different objectives. Parasitological diagnosis is a direct method for simple, inexpensive and quick trypanosome identification in blood, lymph and cerebrospinal fluid (CSF) using microscopy (Moody and Chiodini, 2000). Fresh blood from a finger prick, chancre fluid or lymph node aspiration from humans and blood from ears or tails of animals are collected for detection of trypanosome motility for clinical investigations using the wet blood film method (Moody and Chiodini, 2000). Parasitological diagnosis is limited to morphological characteristics thus cannot distinguish all species (Moody and Chiodini, 2000).

Serological diagnostic methods are indirect methods based on detection of antibodies to specific antigens to detect the presence of circulating parasites in blood, serum and cerebrospinal fluid. Serological tests are used as tools for research, monitoring, control and surveys (Chappuis *et al.,* 2005). Many different types of serological tests have been in used, including: haemagglutination assays (HA), immunofluorescent assays (IFA), and indirect enzyme linked immunosorbent assays (ELISA). Trypanosome diagnosis based on serological method is more sensitive and can distinguish species of pathogenic trypanosomes present (Chappuis *et al.,* 2005). Molecular diagnostic methods have very high sensitivity since they are based on detection of DNA sequences that are specific for trypanosome subgenus, species, subspecies, type or strain. Specificity of the tests depend on the availability of primers that are general enough to amplify all variants of a given species but are specific enough to allow distinguishing between species (Cunningham *et al.,* 2016).

2.4 Degradation of proteasomal products by tricorn protease and its interacting factors

Proteolysis plays a significant role in proliferation and cell cycle transformations in *T. b. brucei* (Mutomba and Wang, 1998). Proteasomal degradation is critical for cell survival and in eukaryotes, the 26 S proteasome is essential for the rapid destruction of key regulatory proteins, such as cell cycle regulators and transcription factors, whose fast and tuned elimination is necessary for the proper control of the fundamental cell processes they regulate (Cromm and Crews 2017). The 26 S proteasome is also responsible for cell quality control by eliminating defective proteins such as misfolded proteins, nascent prematurely terminated polypeptides, or proteins that fail to assemble into complexes (Cromm and Crews 2017). Peptides generated by proteasomes and related systems range from 8-15 amino acids long (Baumeister *et al.*, 1998) (Rockel and Baumeister, 2008) hence must be degraded to single amino acids for further use in cell metabolism and for the synthesis of new proteins cell (Clausen and Groll, 2003) (Groll *et al.*, 2005).

Studies have also shown that proteasomal products get rapidly cleared since their accumulation in the cytosol could interfere with important protein-protein interactions essential for cell viability (Paugam *et al.*, 2003). The proteasomal degradative pathway is the core non-lysosomal protein degradation machinery in protozoan parasites including trypanosomes (Paugam *et al.*, 2003). Tricorn protease and its interacting factors has been described as the core enzymes in proteolytic system in many bacterial cells (Pallen *et al.*, 2001). In *Thermoplasma acidophilum*, tricorn protease and its interacting factors (F1, F2, and F3) act *in vivo* downstream of the proteasome degrading proteasomal products to free amino acids for other metabolic processes (Figure 2.4) (Tamura *et al.*, 1998).

Tricorn protease is a hexameric protease of 720 kDa and can assemble into a giant icosahedral capsid, which might serve as the organizing center of a multiproteolytic

complex (Tamura *et al.,* 1996; (Walz *et al.,* 1997). The monomeric 120 kDa subunit has a mosaic structure with two open Velcro beta-propeller structures of 6 and 7 blades, a helical bundle, a PDZ domain and an alpha-beta sandwich structure. These five domains combine to form one of 6 sub-units, which further assemble to form 3-2 symmetric core protein (Figure 2.5 a) (Brandstetter *et al.,* 2001; Kim *et al.,* 2002). Tricorn protease N-terminal domain folds has a six-bladed β-propeller (β6) followed by a seven-bladed β-propeller (β7). The β6 represents a gated exit from while β7 represents a passage route into the catalytic chamber and the PDZ domain is inter-sparsed between two carboxyl-terminal mixed α-β domains C1 and C2 (Figure 2.5 b) (Kim *et al.,* 2002).

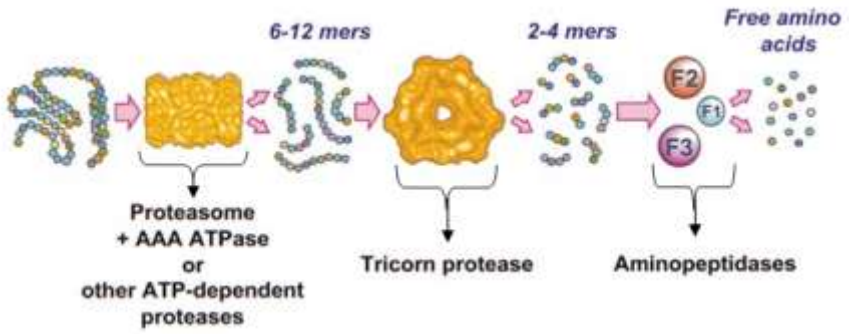


Figure 2.4: The Proteolytic Pathway in *Thermoplasma acidophilum* (Tamura *et al.,* 1998).

The nucleophilic serine is positioned at a helix entrance within subdomain C2 and the arrangement of the tetrad active site molecules suggests peptide bond hydrolysis following the classical trypsin-like serine proteases (Figure 2.6) (Kim *et al.,* 2002). Inside the cells, 20 copies of tricorn protease hexamer can assemble further to form an icosahedral capsid structure with a molecular mass of 14.6 MDa, which might serve as the organizing center of a multi-proteolytic complex (Walz *et al.,* 1997).

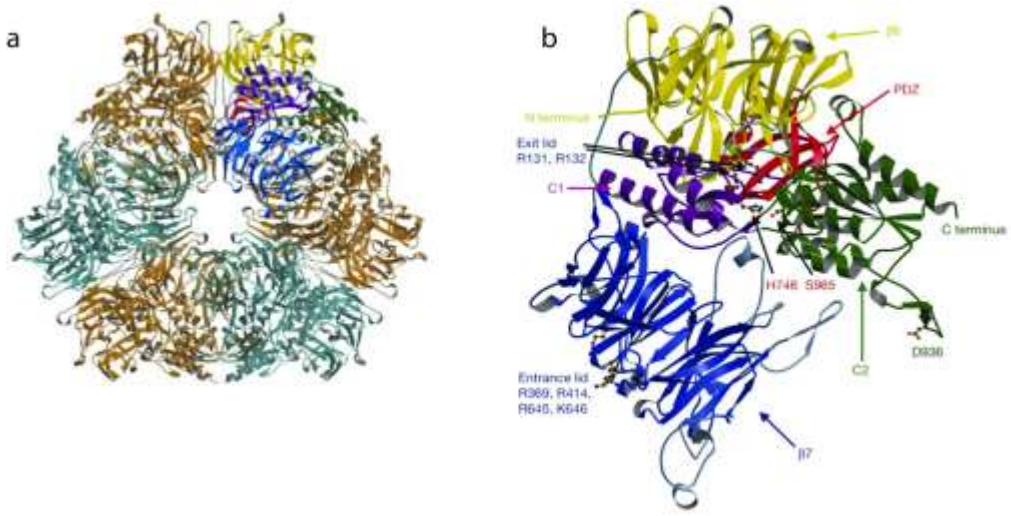


Figure 2.5: Structure of tricorn protease from *Thermoplasma acidophilum*, a; Ribbon representation of the hexameric tricorn protease with tricorn protease subunits distinguished by colour. b; Representation of tricorn protease monomer. The residues forming the propeller lids are displayed as well as the catalytic residues. D936, positioned on the C2 domain, confers specificity for basic substrate residues in the active site (Brandstetter *et al.*, 2001).

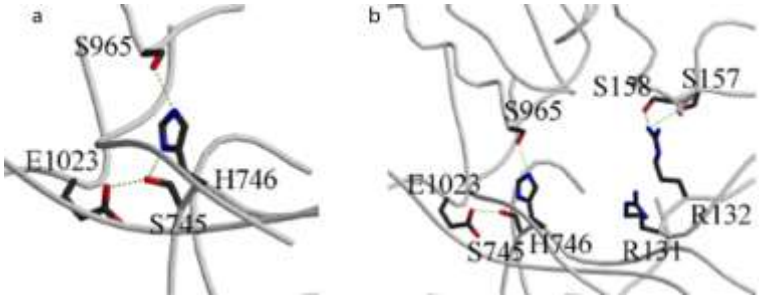


Figure 2.6: Representation of tricorn protease catalytic residues as stick models with tricorn protease main chain shown as grey coils and some residues around the active site as sticks. a: The catalytic Ser965 is activated by His746, which is oriented properly by the sequential hydrogen-bond networks of Ser745 and Glu1023. b: Broad range of active site including the putative substrate binding site (Arg131-Arg132) and its hydrogen-bond partners (Ser157-Ser158) (Kim *et al.*, 2002).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study site

This study was conducted at the Molecular Biology and Bioinformatics Unit in ICIPE, Duduville Campus, Kasarani, Nairobi.

3.2 Experimental animals and stock *Trypanosoma brucei brucei*

The Swiss mice used in this study were obtained from ICIPE's animal house and the *Trypanosoma brucei brucei* (ILtat 1.4) were obtained from ICIPE's cryobank. These parasites originated from ILRI cryobank.

3.3: Computational identification of tricorn-like proteins in *Trypanosoma brucei brucei*

3.3.1 BLAST Homology searches

3.3.1.1 Homology search using tricorn protease

Position specific iterated basic local alignment search tool (PSI-BLAST) (Altschul *et al.,* 1997) search was initiated using the *Thermoplasma acidophillum* tricorn protease (AAC 44621.1) protein sequences against the NCBI's non-redundant (NR) protein database. The search was performed to determine patterns of conservation which aid in the recognition of distant similarities with a default cut-off e-value of 1e-5 for significant matches.

A domain search in Superfamily 1.75 HMM library and genome assignments server based on a library of Hidden Markov Models (Gough *et al.,* 2001) (Oates *et al.,* 2015) was initiated using the *Thermoplasma acidophilum* tricorn protease (AAC 44621.1) to identify tricorn domains in various combinations in *T. b. brucei* genome. A search was also initiated using tricorn protease (AAC 44621.1) in Conserved Domain Architecture

Retrieval Tool (CDART) (Geer *et al.,* 2002) to identify *T. b. brucei* proteins with significant structural similarities to the structure of tricorn protease.

3.3.1.2 Homology search using tricorn interacting factors

Position specific iterated basic local alignment search tool (PSI-BLAST) (Altschul et al. 1997) search was initiated using the *Thermoplasma acidophillum* tricorn protease interacting factors, proline iminopeptidase, F1 (CAC11959.1), tricorn protease interacting factor F2 (CAC11446.1) and tricorn protease interacting factor 3 (CAC11944.1) protein sequences against the NCBI's non-redundant (NR) protein database. The search was performed to determine patterns of conservation which aid in the recognition of distant similarities with a default cut-off e-value of 1e-5 for significant matches.

3.3.1.3 Tricorn interacting factor 3 orthology search in OrthoMCL

Identification of orthologous groups was done through OrthoMCL database (Chen, 2006) which employs Markov Cluster algorithm to group (putative) orthologs for sequences which had PSI-BLAST significant matches. The algorithms involved BLASTp comparisons of the query sequences (CAC11944.1) and sequences from the trypanosome genome and other eukaryotic genomes. The between species and within species relationships of the putative orthologs were identified by reciprocal best similarity pairs with a default cut-off e-value of 1e-5.

3.3.1.4 Identification of tricorn interacting factor 3 conserved domains

Identification of conserved domains was also done through NCBI CDD (Marchler-Bauer *et al.,* 2015) which employs the reversed position specific BLAST (RPS-BLAST) generating either specific or non-specific hits. Transmembrane helices, O-, N-glycosylation and phosphorylation sites were determined using TMHMM, NetOGlyc and NetPhos softwares, respectively.

3.3.2 Sequence Alignment

Both pairwise and multiple sequence alignments of tricorn interacting factor 3 orthologs were performed by MAFFT v7 (Multiple alignment program for amino acid sequences) (Katoh and Standley, 2013) which uses the BLOSUM62 matrix to evaluate the significance of the alignments. A default gap penalty of 1.53 and a cut-off e-value of 1e-5 were used to obtain more reliable results.

3.3.3 Motif occurrence and analysis

The tricorn interacting factor 3 conserved motifs were identified through Multiple Expectation maximization for Motif Elicitation suite (MEME v.4.11.2) (Bailey *et al.,* 2009). The sequences were uploaded in MEME suite's motif discovery tool which searched for statistically significant motifs in the sequences ranging from 6-50 amino acids wide with a defualt cut-off e-value of 1e-5. The identified motifs were analyzed through motif scanning and comparison tool of MEME suite (FIMO) to identify their occurrence in other trypanosome proteins in any other possible combinations while motif enrichment analysis was done by MEME suite's MAST tool.

3.3.4 Protein-Protein Interactions

Protein-protein interactions of the tricorn interacting factor 3 orthologs with other *T. b. brucei* proteins were analyzed through STRING database v.10.0 (Search Tool for Retrieval of Interacting Proteins) (Szklarczyk *et al.,* 2010) to determine their associations in terms of co-expression, text mining, protein homology and gene neighborhood. The searches were performed at high confidence levels (≥ 0.7) and included both first and second shell interactions. Protein-protein interaction enrichment p-value of ≤ 0.05 was considered significant.

3.3.5 Protein structure prediction and validation

The structures of the putative tricorn interacting factor 3 orthologs in *T. b. brucei* were

predicted through (PS)v2 structure prediction server (Huang *et al.,* 2015) and SWISS model (Arnold *et al.,* 2006) (Biasini *et al.,* 2014). The sequences were used to search and select a suitable template which was used to build the models. The predicted models were saved in pdb files through PyMOL Molecular Graphics System, v.1.8 Schrödinger, LLC (Delano, 2002). The pdb files were uploaded to PDBsum (Laskowski *et al.,* 1997) for validation.

3.3.6 Phylogenic tree of tricorn interacting factors 3 orthologs in *Trypanosoma brucei brucei*

The amino acid sequences of tricorn interacting factors 3 orthologs were aligned using MAFFT version 7.0 (Katoh and Standley, 2013). Phylogenetic analysis of aligned amino acid sequences was performed using the MEGA version 7.0 packages (Kumar, Stecher and Tamura, 2016). Phylogenetic tree was constructed using Maximum Likelihood method (Whelan and Goldman, 1995) and bootstrap resampling (1000 replicates) was used to assess the robustness of the groupings obtained.

3.4 Molecular characterization of tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei*

3.4.1 Genomic DNA preparation

3.4.1.1 Mice maintenance and infection

Three adult male Swiss mice (six weeks old) weighing approximately 20g were maintained in ICIPE animal house on commercial mice pellets. The mice were housed in the same cage which was clearly labelled for identification purposes. Mice infections were done through intraperitoneal injections with 2 x 10^3 *T. b. brucei* ILTat1.4 /ml in 0.2 ml cold phosphate buffered saline (PBS) (pH 7.4) supplemented with 1% w/v glucose. The parasites were obtained from International Livestock Research Institute (ILRI), Kenya. To maintain cell viability in PBS, parasites were kept on ice prior to

injection into mice and parasitemia was monitored daily through microscopy. Blood films were prepared by placing a drop of blood from mouse tail on one end of a slide and a spreader slide was used to disperse the blood over the slide's length. The slide was air dried after which the slide was then stained using giemsa stain and covered with a cover slip. The monolayer was then viewed under a light microscope. The mice were handled in adherence to experimental guidelines and procedures approved by the Institutional Animal Care and Use Committee (IACUC, Ref: C/TR/4/325/125) and by the Trypanosomiasis Research Centre of the Kenya Agricultural Research Institute (KARI-TRC). These IACUC regulations conformed to national guidelines provided by the Kenya Veterinary Association.

3.4.1.2 Parasite harvesting

The mice were exsanguinated three (3) days post infection when the parasitemia levels reached 1 x 10^6 parasites / ml. The blood was collected in a heparinized syringe at a ratio of 1:40. The blood was transferred into a sterile eppendorf tube and mixed with PBS, pH 7.4 in a ratio of 1:1. The parasites were harvested from mice blood by centrifuging for 5 minutes at 200 xg on a desktop centrifuge. The pellet containing mainly the blood cells was discarded while the supernatant was transferred into a sterile eppendorf tube. This was centrifuged for 10 minutes at 1400 xg to pellet the parasites.

3.4.1.3 DNA extraction and quantification

Genomic DNA was isolated using Qiagen's DNeasy DNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Quantitation and purity analysis was done using NanoDrop 2000c Spectrophotometer (Thermo Scientific, Waltham, MA, USA). The NanoDrop measurements were based on the optical density readings at 260 and 280 nm.

3.4.1.4 DNA electrophoresis

The extracted genomic DNA was resolved in 1% agarose gel, ethidium bromide stained and viewed under UV light. Briefly, one (1) gram of agarose was weighed and transferred into an Erlenmeyer flask. 100 ml 1x TAE was added, this was mixed by swirling gently. The mixture was heat in a microwave for three (3) minutes to melt the agarose completely. The molten agarose was cooled to about 50°C prior to addition of ethidium bromide (0.5µg/ml). It was then poured into the casting tray with gel combs in place and allowed to set at room temperature for approximately 30 minutes. The combs were removed and the gel was placed in the gel tank and enough running buffer (1x TAE) was added to cover the surface of the gel. A loading dye was added to 3µl of the extracted DNA sample. Electrophoresis was done at 70 volts for one (1) hour and the gel was exposed to UV light in a gel documentation system.

3.4.2 Polymerase Chain Reaction (PCR)

Primers for PCR were designed based on the two putative tricorn interacting factor 3 sequences using Primer 3 plus (Untergasser *et al.,* 2007). The primers were optimized.

Gene	**Primer**	**Sequence**	**Product (bp)**
Tb927.3.4750/90	Forward	5'-GGCTCTTGGCGCAGTATAAC-3'	325
	Reverse	5'-GAGCACAGTCATCCTGTGGA-3'	
Tb927.11.3570	Forward	5'-TGGTAATGCAACAACGAAGG-3'	285
	Reverse	5'-GGTATTGGGTTTTCCCGTTT-3'	

Polymerase chain reaction was carried out using Dream Taq green PCR master mix (2x), (Thermo Scientific, Waltham, MA, USA) on a Arktic Gradient 96 thermal cycler (Thermo Scientific, Waltham, MA, USA) with the following thermal cycling conditions:

Steps	Conditions	
Initial denaturation	95°C for 3 minutes	1 cycle
Denaturation	95°C for 30 seconds	35 cycles
Annealing	Tb927.11.3570; 56.9°C, Tb927.3.4750; 57.6°C for 30 seconds	
Extension	72°C for 30 seconds	
Final extension	72°C for 10 minutes	1 cycle

The reaction mixtures for the 20 µl PCR reaction constituted 10 µl Dream Taq green PCR master mix (2x), 2µl forward primer, 10 pMol, 2µl reverse primer, 10pmol, 2 µl template 50 ng/µl genomic DNA and 4µl nuclease free water to a final volume of 20µl. Amplicons were analyzed on 1.2% EtBr stained agarose gel electrophoresis as previously described in section 3.4.1.4.

3.4.3 RNA preparation

3.4.3.1 Total RNA extraction, quantification and quality estimation

Trypanosoma b. brucei (ILTat 1.4) was cultured as described in section 3.4.1.1 and harvested as described in section 3.4.1.2. Total RNA was extracted using Direct-zol™ RNA Mini-Prep kit (Zymo Research, Irvine, CA, USA) according to manufacturer's instructions. The quantity and purity of the extracted RNA was estimated using NanoDrop 2000c Spectrophotometer (Thermo Scientific, Waltham, MA, USA). The NanoDrop measurements were based on the optical density readings at 260 and 280 nm.

3.4.3.2 RNA electrophoresis

To evaluate the integrity of the RNA, both non-denaturing and denaturing gel electrophoresis were carried out. For non-denaturing gel electrophoresis, 1.2% EtBr

stained agarose gel electrophoresis on 1x TAE buffer was prepared. The samples were then prepared by adding 2 µl of the loading dye to 2 µl of the extracted RNA solution; this was then heated at 70°C for 1 min and then chilled back in ice before loading on the already precast gel. The gel was run at 65 volts for 1hr and visualized under UV trans-illuminator. For the denaturing gel electrophoresis, 1.2% agarose was prepared by melting 0.5 g agarose in 36 ml distilled water. The molten agarose was allowed to cool to about 60°C, after which 5 ml 10x MOPs buffer (200 mM MOPS free acid, 10 mM EDTA, 50 mM sodium acetate pH adjusted to 7.0 by NaOH or by glacial acetic acid) was added. The solution was mixed by swirling gently. Then, in a fume hood, 9 ml of 37% formaldehyde solution was added and swirled gently to mix. This was then poured into the gel tray and allowed to set. The samples were prepared for loading by mixing 2 µl of the extracted RNA, 8 µl sterile distilled water and 10 µl RNA sample buffer (20 mM MOPS, 1mM EDTA, 5 mM NaAcetate, 50% (v/v formamide, 2.2 M formaldehyde). The RNA was then denatured by heating at 65^{o}C for 5 minutes, then chilled on ice for 5 minutes after which, 2 ul RNA sample dye was added, vortexed for 5 seconds and spun for 5 seconds to pull the sample to the bottom of the tube before loading into the already precast gel. The gel was run at 70 volts for 45 minutes before visualizing under UV trans-illuminator.

3.4.4 First strand cDNA preparation

3.4.4.1 First strand cDNA synthesis

The extracted total RNA was used to synthesize first strand cDNA using the RevertAid H minus First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA). 1 µl of the extracted total RNA (1.412 µg) was added into a sterile nuclease free tube on ice. This was followed by addition of 1µl oligo $(dT)_{18}$ primer. The mixture was topped up to 12 µl by nuclease free water. The contents were mixed gently, briefly centrifuged and incubated at 65°C for 5 minutes. The tube was chilled back on ice and the following were added; 4 µl 5X reaction buffer, 1 µl RiboLock RNase Inhibitor (20 U/µl) and 1 µl

RevertAid H minus M-MuLV Reverse Transcriptase (200 U/µl). The contents were mixed gently and briefly centrifuged before incubation at 42°C for 1 hour. The reaction was terminated by heating at 70°C for 5 minutes.

3.4.4.2 First strand cDNA analysis

The synthesized first strand cDNA was analysed through PCR amplification carried out with gene specific primers outlined in section 3.4.2 The *T. b. brucei* actin and telomerase reverse transcriptase (TERT) genes were used as control. Polymerase chain reaction was done following the conditions described in section 3.4.2. Amplicons were analysed through agarose gel electrophoresis as described in section 3.4.1.4.

Gene	**Primer**	**Sequence**	**Product (bp)**
Actin	Forward	5'-GTACCACTGGCATTGTTCTCG-3'	133 bp
	Reverse	5'-CTTCATGAGATATTCCGTCAGGTC-3'	
TERT	Forward	5'-GAGCGTGTGACTTCCGAAGG-3'	108 bp
	Reverse	5'-AGGAACTGTCACGGAGTTTGC-3'	

3.4.5 Real time Quantitative PCR (qPCR)

Real time quantitative PCR was conducted with Stratagen Mx 3005P Q- PCR system (La Jolla, CA, USA) applying the SYBR green detection method using *T. b. brucei* actin and telomerase reverse transcriptase (TERT) genes as internal control genes as outlined in section 3.4.4.2. The qPCR primers were designed to amplify shorter fragments as shown below;

Gene	Primer	Sequence	Product (bp)
Tb927.3.4750/90	Forward	5’-GGAAGTACGATGGGAAGGAA-3’	134 bp
	Reverse	5’-GAGCACAGTCATCCTGTGGA-3’	
Tb927.11.3570	Forward	5’-CCACCTTGCTGATGAAGCTA-3’	113 bp
	Reverse	5’-CCCACACACCTCGTGAATAG-3’	

Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo Scientific, Waltham, MA, USA) was used and 10 µl reaction was constituted by adding the following: 5µl Maxima SYBR Green/ROX qPCR Master Mix (2X), 1µl forward primer, 10 pmol, 1µl reverse primer, 10 pmol, 0.5 µl of template cDNA (20 ng/ul) and topped-up to 10 µl by nuclease free water. The thermal cycling conditions constituted;

Steps	Conditions	
Initial activation	95°C for 10 minutes	1 cycle
Denaturation	95°C for 15 seconds	40 cycles
Annealing	57.0°C for 30 seconds	
Extension	72°C for 30 seconds	
Melting step	95°C for 1 minute, 55.0°C for 30 seconds, 95°C for 30 seconds	1 cycle

3.4.6 Cloning

3.4.6.1 Cloning of tricorn interacting factor 3 orthologs

Using a PCR based strategy, the tricorn interacting factor 3 orthologs were amplified using specific primers incorporating restriction sites as shown below;

Gene	Primer	Sequence	Product (bp)
Tb927.3.4750/90	Forward	5’TTTg^gatccGGCTCTTGGCG CAGTATAAC3’	325 bp
	Reverse	5’TGGc^tcgagGAGCACAGTCA TCCTGTGGA3’	
Tb927.11.3570	Forward	5’TTTg^gatccTGGTAATGCAA CAACGAAGG3’	285 bp
	Reverse	5’GTGg^aattcGGTATTGGGTT TTCCCGTTT3’	

For both genes, the forward primer incorporated the BamHI site. However, the reverse primer incorporated XhoI site for Tb927.3.4750/90 and EcoRI site forTb927.11.3570. The restriction sites were shown in lower case. Dream Taq DNA polymerase was used for amplification since it introduces the adenosine overhangs for compatibility with p-GEM-T vector system which employs the A/T tail cloning technology (it contains the thymidine overhangs). The PCR products were purified and sequenced before ligation into the linearized vector using the p-GEM-T- Easy vector systems cloning kit (Promega, Woods Hollow Road, Madison, USA). The ligation reaction (10 µl) was set up by adding 5 µl of 2X Rapid Ligation Buffer T4 DNA Ligase, 1 µl of pGEM®-T Easy Vector, 1 µl PCR product, 1 µl T4 DNA Ligase (3 Weiss units/µl) and 2 µl deionized water. The reaction contents were mixed well and incubated overnight at 4°C.

3.4.6.2 Transformation

Transformation of *E. coli* JM109 High Efficiency Competent Cells was done by adding 2 µl of the ligation reaction to a sterile 1.5ml tube on ice. The *E. coli* JM109 High Efficiency Competent Cells were placed in an ice bath until just thawed (5 minutes) then mixed gently flicking the tube. 50µl of cells were transferred to the ligation reaction and the tube was gently flicked and incubated on ice for 20 minutes. The cells

were heat shocked for 45 seconds in a water bath at exactly 42°C and returned immediately to ice for 2 minutes. 950 µl room temperature SOC medium was added to the ligation reaction and then incubated for 1.5 hours at 37°C with shaking (~150rpm).

3.4.6.3 Bacterial culture

100 µl of transformed bacteria were plated onto LB/ampicillin/IPTG/ X-Gal plates for blue-white screening. The plates were incubated overnight at 37°C and white colonies selected for further screening through colony PCR. Positive colonies were expanded in LB/ampicillin broth overnight at 37°C. Plasmid purification was done using Qiagen's plasmid purification kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. The recombinant plasmid was digested using BamHI and XhoI; for Tb927.3.4750/90, and BamHI and ECoRI for Tb927.11.3570 for 12 hours in water bath set at 37°C. The genes of interest from both digestions were gel purified for expression.

3.4.7 Expression

3.4.7.1 Ligation

pGEX-5x-3 plasmid DNA (glutathione S-transferase tag) was linearized by digesting with two different sets of restriction enzymes (BamHI and XhoI; BamHI and ECoRI) for 12 hours in water bath set at 37°C and purified using the Bioline PCR / gel extraction kit (Bioline, Humber Rd, London, UK). The purified gene of interest was ligated into the linearized vector using Promega T4 DNA ligase according to manufacturer's instructions and transformed into competent *E. coli* BL-21 (DE3) pLysS cells.

3.4.7.2 Protein expression

The transformed cells were cultured as described in section 3.4.6.3 on LB/Ampicillin agar plates. For recombinant protein expression, positive clones were grown in YT (Yeast-Tryptone) media containing ampicillin (100ug/ml) overnight in a shaker

incubator (150 rpm) at 37°C. The overnight culture was diluted at 1:100 using pre-warmed YTA media and allowed to grow at 37°C to an OD_{600} of 0.6-0.8 prior to induction of protein expression with isopropylthio-β-galactoside (IPTG, 0.5-1mM final concentration). Bacterial cells were picked at an interval of 1 hour for 3 hours after induction. The cells were lysed through sonication and transferred to protein sample buffer (Glycerol, Tris—HCL, β-ME, SDS, Bromophenol blue). Boiling was done for 10 minutes at 95-100°C before loading.

3.4.8 Sodium dodecyl sulfate polyacrylamide gel electrophoresis

The expressed proteins were analyzed through Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) adopted from Laemmli, 1970. 10 % SDS-PAGE gel was prepared by mixing distilled water (4.1 mL), bis-Acrylamide (30% 37.5:1; Bio-Rad) (3.3 mL), Tris–HCl (1.5 m, pH 8.8) (2.5 mL), SDS (100 µL), N, N, N′, N′-tetramethylethylene-diamine (TEMED) (10 µL) and Ammonium persulfate (APS), 10% (32 µL) in that order. The separating gel was poured into the bottom of the tank and the top of the gel was layered with isopropanol. This helped to remove bubbles at the top of the gel and helped to keep the polymerized gel from drying out. The gel was left to polymerize for about 30 minutes. After polymerization, isopropanol was removed and the remaining traces were washed with distilled water. Then 4% stacking gel was prepared by adding distilled water (6.1 mL), bis-Acrylamide (30%, 37.5:1) (1.3 mL), Tris–HCl (0.5 M, pH 6.8) (2.5 mL), SDS, 10% (100 µL), TEMED (10 µL), Ammonium persulfate (APS), 10% (100 µL). The stacking gel was poured on top of the separating gel, the combs were added to make wells and allowed to polymerize in about 30 minutes. The gel was clamped into the apparatus and both chambers were filled with buffers and samples were loaded into the wells for separation by electrophoresis.

3.4.9 RNAi studies

3.4.9.1 Preparation of recombinant plasmid for transfection

cDNA was prepared through a single PCR reaction with primers incorporating the BamHI and HindIII restriction sites (shown in lower case) to the 5' and 3' ends respectively as shown below;

Gene	**Primer**	**Sequence**	**Product (bp)**
Tb927.3.4750/90	Forward	5’TTTg^gatccGGCTCTTGGCGCAGTATAAC3’	325 bp
	Reverse	5’TGGa^agcttGAGCACAGTCATCCTGTGGA3’	
Tb927.11.3570	Forward	5’TTTg^gatccATCCTGGTAATGCAACAACGAAGG3’	285 bp
	Reverse	5’GTGa^agcttGGTATTGGGTTTTCCCGTTT3’	

PCR amplification was carried out using Dream Taq green PCR master mix (2x), (Thermo Scientific, Waltham, MA, USA) as described in section 3.4.2. However, a gradient annealing from 55°C -77°C was done. The reaction mixtures for the 20 µl PCR reaction constituted 10 µl Dream Taq green PCR master mix (2x), 2 µl forward primer, 10 pMol, 2 µl reverse primer, 10 pmol, 2 µl 50 ng/µl cDNA (template) and 4 µl nuclease free water to a final volume of 20 µl. The amplicons were analyzed through agarose gel electrophoresis as described in section 3.4.1.4. The products were purified using QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) according to manufacturer’s instructions prior to use in transcription reaction.

The cDNA incorporated the BamHI and HindIII restriction sites to the 5' and 3' ends respectively. The amplified fragments were then ligated into the two opposing T7

promoters bearing tetracycline operators of linearized p2T7-177 vector. Transfection of *E. coli* JM109 High Efficiency Competent cells was done through heat shock method in a water bath. The bacteria were cultured overnight at 37°C in LB-ampicillin plates. Colony expansion was done in LB- ampicillin broth and plasmid purification was done using Qiagen's plasmid purification kit (Qiagen, Hilden, Germany) according to manufacturer's instructions and was stored at -20°C. Restriction digestion was done screen for recombinant plasmids.

3.4.9.2 Trypanosome culture

Trypanosoma b. brucei bloodstream form expressing both T7 RNA polymerase and tetracycline repressor were cultured in HMI 9 medium (Hirumi and Hirumi, 1989) supplemented with 10% heat inactivated (56ºC, 30 min) fetal bovine serum (Gibco, ThermoFischer Scientific), L-glutamine, penicillin/streptomycin (50 μg ml^{-1}) and β-mercaptoethanol (14 μl / litre) in vented non-adherent culture flasks. The parasites were incubated at 37ºC with 5% CO_2 supply. The parasites were cultured in the presence of G418 (15 mg/ml) and hygromycin (50 mg/ml) to maintain the T7 RNA polymerase and the tetracycline repressor constructs, respectively. The parasites were grown for 14 days prior to transfection.

3.4.9.3 Transfection

For transfections, 3 x 10^6 bloodstream *T. b. brucei* / ml were harvested by centrifuging at 200 xg for 5 minutes and washed in PBS. The parasites were re-suspended in 1 ml of gene pulser electroporation buffer (Bio-Rad, Hercules, California, United States). Then 10 μl of 50 ng/μl of the transformed p2T7-177 plasmid was added to the parasites which were then electroporated with a single pulse of 400 V, 500 μ F with a time constant of 5 milliseconds using a Bio-Rad Gene Pulser X-cell system (Bio-Rad, Hercules, California, United States) in 0.4-cm-gap cuvettes. The parasites were immediately transferred into 15 ml fresh HMI-9 medium and incubated at 37°C in humidified CO_2 Galaxy® 48 R incubator (Hamburg, Germany). RNAi induction was

done 24 hours after transfection by adding 0.5 µg/ml tetracycline disodium dissolved in PBS. Selection was done in presence of 2.5 µg/ ml bleomycin for a period of five (5) days to prevent development of RNAi revertants. Cell densities were determined using Neubauer chamber. The empty plasmid was used for mock transfections of control parasites.

3.4.9.4 RNAi validation

To validate RNAi and investigate persistence of its effect, total RNA was extracted from tetracycline induced RNAi mutants using Direct-zol™ RNA Mini-Prep kit (Zymo Research, Irvine, CA, USA) five days post transfection Zymo Research RNA purification handbook. One-step qRT-PCR was done using QIAGEN One step qRT-PCR kit (Qiagen, Hilden, Germany). The master mix (25 µl) was prepared in ice by adding 5 µl of 5x Qiagen onestep qRT-PCR buffer, 1 µl dNTP (10 mM of each dNTP), 1 µl each of forward and reverse primers (10 µM), 1 µl Qiagen one step enzyme mix, 1 µl RNase inhibitor, 1 µg total RNA and topped up to 25 µl with nuclease free water (14 µl) in PCR tubes. The tubes were kept in ice until the thermal cycler was at 50°C. Reverse transcription at 50°C for 30 minutes, followed by conventional PCR steps which consisted of initial PCR activation step of 15 min 95°C, then 35 cycles of denaturation 30 seconds at 94°C, annealing 30 seconds at 57°C, extension 1 min 72°C and additional step was set at 94°C for 1 minute, 55.0°C for 30 seconds, then 94°C for 30 seconds for the dissociation curve analysis. The relative amount of gene transcripts were calculated using Livak Method (Livak and Schmittgen, 2001).

CHAPTER FOUR

RESULTS

4.1: Computational identification of tricorn-like proteins in *Trypanosoma brucei brucei*

4.1.1 BLAST Homology searches

4.1.1.1 Identification of tricorn-like proteins in *Trypanosoma brucei brucei*

The PSI-BLAST sequence homology search using tricorn protease identified a total of four (4) proteins in *T. b. brucei,* however; all of them had very high E-values with low scores showing that the hits were not significant (Table 4.1). A domain search in Superfamily 1.75 HMM library and genome assignments server identified a total of 21 *T. b. brucei* proteins with tricorn protease-like domains of which 10 had tricorn protease domain 2, 10 had PDZ domains and one (1) had tricorn protease N-terminal domain (Table 4.2). Tricorn protease domain 2 occurred in seven (7) different domain combinations with a total of two (2) different partner domains (WD 40 repeat and Quinoprotein alcohol dehydrogenase). Tricorn protease N-terminal domain occurred in one (1) domain combination with three (3) different partner domains (WD40 domain, MIF4G domain and Tetratricopeptide domain). PDZ domain occurred in five (5) domain combinations with a total of four (4) different partner domains (WD40 repeat, ARM repeat, Rapamycin binding and Protein kinase-like) (Figure 4.1). Similar results were obtained with CDART search.

Table 4.1: PSI-BLAST homologies between tricorn protease and its interacting factors in *Trypanosoma brucei brucei* genome

Query sequence	Description	Max score	Total score	Query cover	E value	Identity	Accession
Tricorn protease (AAC 44621.1)	prolyl oligopeptidase	31.6	31.6	5%	1.1	29%	XP_823036.1
	calpain-like protein,	31.6	31.6	11%	1.5	26%	XP_828202.1
	Hypothetical protein,	29.6	29.6	5%	3.8	38%	XP_828877.1
	Actin-like	29.6	29.6	3%	4.8	33%	XP_822455.1
Proline iminopeptidase F1 (CAC11959.1)	Bem46-like serine peptidase	41.6	41.6	39%	1.00E-04	27%	XP_827374.1
	Hydrolase	36.6	36.6	52%	5.00E-03	22%	XP_844912.1
	Kinesin	32.3	32.3	75%	1.30E-01	21%	XP_828410.1
	Monoglyceride lipase	31.2	31.2	27%	0.26	26%	XP_844586.1
	Hypothetical protein	30	30	33%	0.73	28%	XP_847161.1
	Hypothetical protein	29.6	29.6	16%	0.94	35%	XP_847354.1
	Hypothetical protein	29.6	29.6	25%	1	26%	XP_828720.1
	Hypothetical protein	29.3	29.3	39%	1.1	23%	XP_845302.1
Tricorn protease interacting factor F2 (CAC11446.1)	Aminopeptidase	378	378	79%	6E-119	34%	XP_828438.1
	Aminopeptidase	257	257	67%	3E-74	30%	XP_844067.1
	Aminopeptidase	32.7	32.7	6%	0.39	32%	XP_847581.1
Tricorn protease interacting factor F3 (CAC11944.1)	Aminopeptidase	369	369	79%	7.00E-116	33%	XP_828438.1
	Aminopeptidase	273	273	67%	6.00E-80	30%	XP_844067.1
	Aminopeptidase	38.5	38.5	8%	6.00E-03	35%	XP_847581.1

The hits were obtained from searches initiated by the archaeon tricorn protease and its interacting factors running against NCBI's non-redundant protein database. Significant matches were obtained with F2 and F3 query sequences which gave similar hits in *T. b. brucei.* E value is a parameter that describes the number of hits one can expect to see by chance when searching a database of a particular size. It decreases exponentially as the Score (S) of the match increases.

Table 4.2: Domain combinations for tricorn-like proteins in superfamily HMM server and CDART

Gene ID	Description	Superfamily	Family	E-value
AAC 44621.1	Tricorn protease, *Thermoplasma acidophilum*	Tricorn protease N-terminal domain	Tricorn protease N-terminal domain	5.1E-11
		Tricorn protease domain 2	Tricorn protease domain 2	1.8E-12
		ClpP/crotonase	Tail specific protease, catalytic domain	5.4E-11
		PDZ domain-like	PDZ domain	7E-06
Tb927.10.14470	Intraflageller transport protein	Tricorn protease N-terminal domain	Tricorn protease N-terminal domain	2.36E-06
Tb927.11.260	Predicted WD 40 repeat	Tricorn protease domain 2	Tricorn protease domain 2	3.53E-16
Tb927.8.6940	Hypothetical protein	Tricorn protease domain 2	Tricorn protease domain 2	1.60E-13
Tb927.4.5380	Hypothetical protein	Tricorn protease domain 2	Tricorn protease domain 2	1.83E-07
Tb927.11.14040	Hypothetical protein	Tricorn protease domain 2	Tricorn protease domain 2	3.01E-12
Tb927. 5. 2570	Translation initiation factor	Tricorn protease domain 2	Tricorn protease domain 2	9.55E-21
Tb927.7.5720	Hypothetical protein	Tricorn protease domain 2	Tricorn protease domain 2	4.18E-19
Tb927.7.3980	Immunodomant antigen	Tricorn protease domain 2	Tricorn protease domain 2	3.53E-09
Tb927.6.3450	Hypothetical protein	Tricorn protease domain 2	Tricorn protease domain 2	2.44E-33
Tb927.1.2980	Hypothetical protein	Tricorn protease domain 2	Tricorn protease domain 2	7.19E-15
Tb927. 4.450	Coatomer alpha subunit	Tricorn protease domain 2	Tricorn protease domain 2	2.75E-08
Tb927.10.13330	Paraflagellar rod component 20	PDZ Domain	PDZ Domain	9.30E-06
Tb927.1.4280	Hypothetical protein, conserved	PDZ Domain	PDZ Domain	1.41E-05
Tb927.11.2660	Golgi reassembly stacking protein	PDZ Domain	PDZ Domain	6.100E-11
Tb927.8.5740	Proteasome 26S non-ATPase subunit 9	PDZ Domain	PDZ Domain	3.52E-09
Tb927.10.15080	Hypothetical protein	PDZ Domain	PDZ Domain	2.12E-08
Tb927.10.15930	pdz domain containing protein	PDZ_serine_protease	PDZ_serine_protease	4.26E-08
Tb927.4.800	TOR-like 1	PDZ Domain	PDZ Domain	5.60E-06
Tb927.11.6320	MRB1-associated protein	PDZ Domain	PDZ Domain	4.86E-05

Gene ID	Description	Superfamily	Family	E-value
Tb927.6.2090	pdz domain containing protein	PDZ_serine_protease	PDZ_serine_protease	8.71E-09
Tb927.10.6700	Hypothetical protein	PDZ Domain	PDZ Domain	1.19E-05
Tb927.6.2090	pdz domain containing protein	PDZ_serine_protease	PDZ_serine_protease	8.71E-09

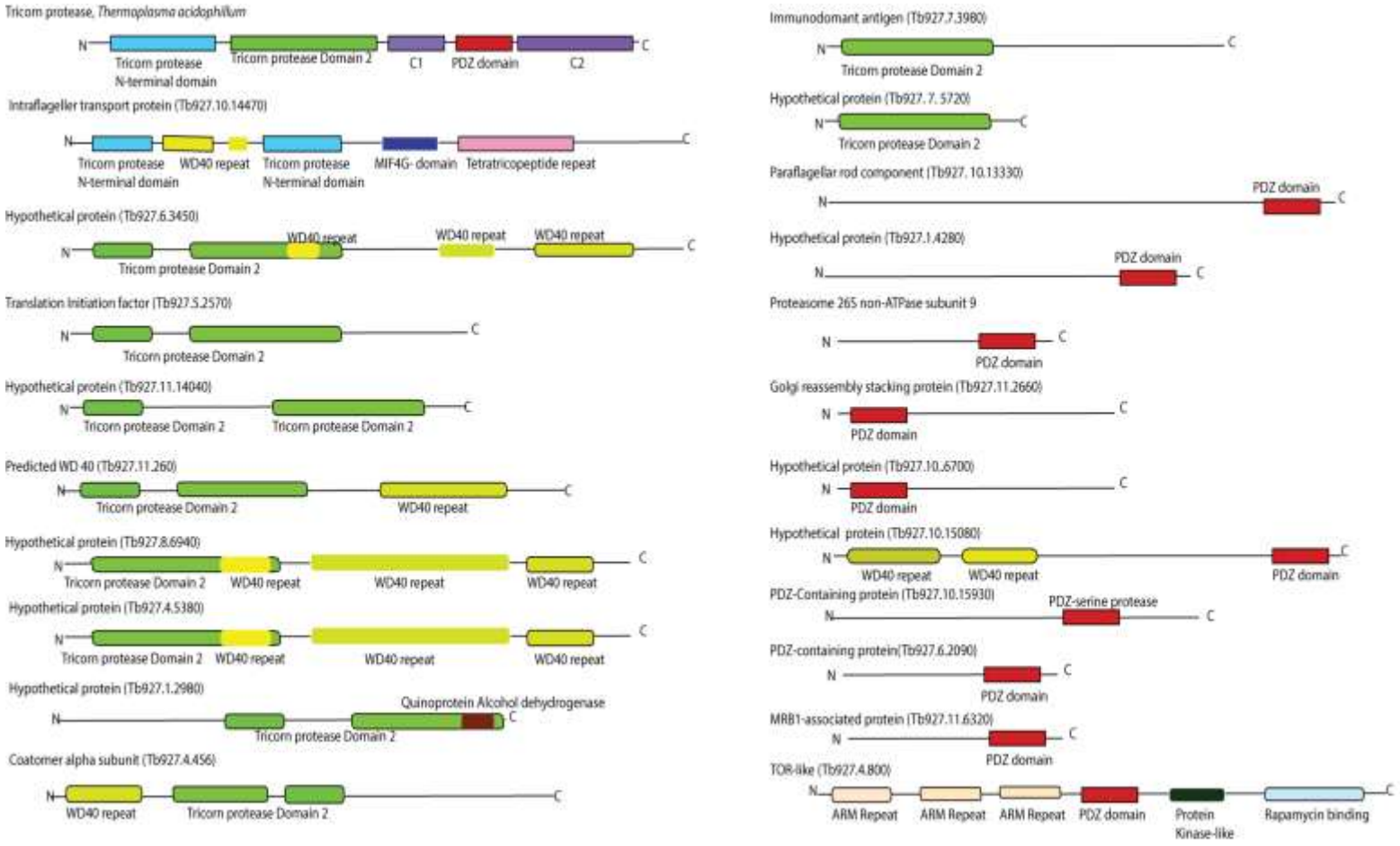


Figure 4.1: Domain combination of *Trypanosoma brucei brucei* proteins with tricorn –like domains in Superfamily HMM server and CDART. The domain combination occurrence is in N-to C-terminal order and the proteins were drawn to scale. Key: C1: Catalytic domain 1, C2: Catalytic domain 2.

4.1.1.2 Identification of tricorn interacting factors orthologs

Tricorn interacting factor 1 identified no significant hits (Table 4.1). However, tricorn interacting factors; F2 and F3 identified three (3) similar putative aminopeptidases of which two (2) were significant hits (Table 4.1). A PSI-BLAST search initiated by the two *T. b. brucei* proteins (XP_828438.1 and XP_844067.1) identified a total of 46 proteins including 20 in trypanosomes. The 20 trypanosome proteins were mainly annotated as aminopeptidases, puromycin-sensitive aminopeptidase-like proteins, putative aminopeptidases and cytosol alanyl aminopeptidase. Other significant matches occurred in *Bodo saltans, Angomonas deanei, Leishmania major, Leptomonas seymouri, Trichomonas vaginalis, Encephalitozoon intestinalis Toxoplasma gondii, Plasmodium falciparum, Babesia bovis, Theileria parva, Cryptosporidium parvum* and *Entamoeba histolytica* (Table 4.3).

Table 4.3: BLAST homologies between *T. b. brucei* aminopeptidases (XP_828438.1 and XP_844067.1) and other eukaryotes

Description	Max score	Total score	Query cover	E value	Identity	Accession
Aminopeptidase,*Trypanosoma brucei brucei*	1810	1810	100%	0.00E+00	100%	XP_828438.1
Aminopeptidase, putative, *Trypanosoma b. gambiense*	1808	1808	100%	0.00E+00	99%	XP_011779552.1
Putative aminopeptidase, *Trypanosoma congolense*	1409	1409	99%	0.00E+00	78%	CCC94954.1
Putative aminopeptidase, *Trypanosoma vivax*	1256	1256	99%	0.00E+00	68%	CCC52886.1
Aminopeptidase, *Trypanosoma cruzi*	1268	1268	99%	0.00E+00	70%	XP_8096100.1
Aminopeptidase, *Trypanosoma cruzi*	760	760	56%	0.00E+00	72%	XP_815344.1
Aminopeptidase, putative, *Bodo saltans*	882	882	100%	0.00E+00	52%	CUG87639.1
Cytosol alanyl aminopeptidase, *Angomonas deanei*	926	926	98%	0.00E+00	53%	EPY35348.1
Cytosol alanyl aminopeptidase, *Angomonas deanei*	924	924	98%	0.00E+00	53%	EPY37412.1
Aminopeptidase-like protein, *Leishmania major*	733	733	98%	0.00E+00	43%	XP_001684037.1
Aminopeptidase, putative, *Entamoeba histolytica*	380	380	98%	4.00E-118	30%	XP_652558.1
Aminopeptidase N, *Encephalitozoon intestinalis*	337	337	68%	5.00E-103	35%	XP_003072227.1
Glutamyl aminopeptidase, *Encephalitozoon intestinalis*	337	337	68%	5.00E-103	35%	XP_003072361.1
Aminopeptidase, putative, *Bodo saltans*	288	288	59%	2.00E-84	34%	CUF06443.1
Unnamed protein product, *Trypanosoma congolense*	276	276	68%	2.00E-80	32%	CCD13518.1
Aminopeptidase, *Trypanosoma brucei brucei*	275	275	69%	2.00E-79	32%	XP_844067.1
Aminopeptidase, putative, *Trypanosoma b. gambiense*	274	274	69%	3.00E-79	32%	XP_011772479.1
Cytosol alanyl aminopeptidase, *Trypanosoma b. brucei*	273	273	69%	5.00E-79	32%	AAS67871.1
Putative aminopeptidase, *Leishmania major*	269	269	63%	7.00E-78	32%	XP_003722307.1
aminopeptidase N-like, *Trichomonas vaginalis*	268	268	67%	4.00E-77	31%	XP_001308807.1
Putative aminopeptidase, *Leptomonas seymouri*	266	266	58%	2.00E-76	32%	KPI86776.1
Aminopeptidase, *Trypanosoma cruzi*	260	260	58%	2.00E-74	33%	XP_822037.1
Aminopeptidase, *Trypanosoma cruzi*	219	219	39%	2.00E-63	37%	XP_806986.1
Aminopeptidase N-like, *Trichomonas vaginalis*	229	229	59%	2.00E-63	31%	XP_001319276.1

Description	Max score	Total score	Query cover	E value	Identity	Accession
Metallo-peptidase, *Angomonas deanei*	225	225	65%	3.00E-62	29%	EPY25844.1
Puromycin-sensitive aminopeptidase, *Leishmania major*	186	186	49%	7.00E-49	26%	XP_001681734.1
Puromycin-sensitive aminopeptidase, *Leptomonas seymouri*	186	186	49%	2.00E-48	27%	KPI87559.1
Metallopeptidase, putative, *Bodo saltans*	177	221	44%	1.00E-45	30%	CUG86927.1
Aminopeptidase, *Trypanosoma cruzi*	149	149	12%	2.00E-42	63%	XP_808981.1
Putative aminopeptidase, *Trypanosoma vivax*	147	147	34%	2.00E-39	31%	CCC47239.1
Puromycin-sensitive aminopeptidase, *Trypanosoma cruzi*	151	199	42%	2.00E-37	28%	XP_815719.1
Puromycin-sensitive aminopeptidase, *Trypanosoma cruzi*	149	199	53%	8.00E-37	28%	XP_821929.1
Putative aminopeptidase N, *Toxoplasma gondii*	124	124	43%	4.00E-28	29%	KYF42161.1
M1-family aminopeptidase, *Plasmodium falciparum*	119	119	56%	1.00E-26	26%	XP_001349846.1
Aminopeptidase N protein, *Toxoplasma gondii*	115	115	0.43	1.00E-25	0.28	KYF42688.1
Alpha-aminoacylpeptide hydrolase, *Theileria parva*	112	112	48%	7.00E-25	27%	XP_765924.1
Putative aminopeptidase N, *Babesia bovis*	103	103	47%	2.00E-22	25%	ABC25604.1
Aminopeptidase N-like, *Trichomonas vaginalis*	90.9	90.9	39%	1.00E-18	24%	XP_001306552.1
Aminopeptidase N, *Cryptosporidium parvum*	87	87	40%	2.00E-17	26%	AAK53986.1
Leukotriene A-4 homolog, putative, *Bodo saltans*	81.3	81.3	19%	4.00E-16	32%	CUG66200.1
Leukotriene A-4 homolog, putative, *Bodo saltans*	81.3	81.3	19%	5.00E-16	32%	CUG65549.1
Putative metallo-peptidase, *Trypanosoma vivax*	57.8	57.8	37%	8.00E-09	23%	CCC50087.1
Aminopeptidase, *Trypanosoma brucei brucei*	47	47	37%	4.00E-05	22%	XP_847581.1
Aminopeptidase, *Trypanosoma brucei gambiense*	46.6	46.6	37%	5.00E-05	23%	XP_011776109.1
Putative aminopeptidase, *Leptomonas seymouri*	43.1	43.1	27%	6.00E-04	23%	KPI87164.1

The hits were obtained from PSI-BLAST searches initiated by the *T. b. brucei* aminopeptidases (XP_828438.1 and XP_844067.1) and running against NCBI's non-redundant protein database. Each of the *T. b. brucei* sequences was queried separately and they identified similar hits other organisms.

4.1.1.3 Identification of tricorn interacting factor 3 orthologs in OrthoMCL

The OrthoMCL database search identified three (3) putative aminopeptidases in *T. b. brucei* as orthologs of tricorn interacting factor 3 (Table 4.4). Ortholgs were also identified in other trypanosome species including *T. b. gambiense, T. cruzi* and *T. congolense.* Other organisms which had significant matches included *Leishmania major, Trichomonas vaginalis* and *Acanthamoeba castellanii* (Table 4.4). The orthologs were clustered into three (3) different ortholog groups; OG5_127177, OG5_145268, and OG5_166686 with the gene product of Tb927.11.3570 clustering in OG5_127177, a group that also contained the archaeon tricorn interacting factors 2 and 3. The gene products of Tb927.3.4750 and Tb927.3.4590 clustered in ortholog group OG_145268, a group that included 10 other putative aminopeptidases, M1 family from other *Trypanosoma* and *Leishmania* species. The third ortholog group OG_166686 included the puromycin-sensitive aminopeptidase-like protein, M1 family in *Leishmania* and *T. cruzi* (Table 4.4).

Table 4.4: Identified ortholog groups from OrthoMCL database

Organism	Accession number	Group	Product	E-value	Mw (Da)
Thermoplasma volcanium	vol|NP_111319	OG5_127177	aminopeptidase N	0.00E+00	88070
Thermoplasma volcanium	tvol|NP_111818	OG5_127177	aminopeptidase N	0.00E+00	88981
Sulfolobus solfataricus	ssol|NP_343540	OG5_127177	tricorn interacting factor 2	0.00E+00	90516
Sulfolobus solfataricus	ssol|NP_344003	OG5_127177	tricorn interacting factor 3	2.00E-179	90144
Acanthamoeba castellanii	aca1_220200	OG5_127177	hypothetical protein	7.00E-126	93795
Trichomonas vaginalis	Tvag_008880	OG5_127177	aminopeptidase N	7.00E-123	94661
Saccharomyces cerevisiae	scer|scer_s288c__YHR047C	OG5_127177	aminopeptidase	6.00E-118	100663
Trypanosoma vivax	tviv|TvY486_1103700	OG5_127177	aminopeptidase,	1.00E-116	98619
Trypanosoma b. gambiense	tbrg|Tbg1002.11.4060	OG5_127177	aminopeptidase,	2.00E-112	98072
Trypanosoma b. brucei	tbru|Tb927.11.3570	OG5_127177	aminopeptidase,	2.00E-112	98098
Trypanosoma congolense	tcon|TcIL3000.11.3540	OG5_127177	aminopeptidase,	2.00E-111	100293
Trypanosoma cruzi	tcru|Tc00.1047053511051.70	OG5_127177	aminopeptidase,	1.00E-106	98383
Mus musculus	mmus|ENSMUSP00000001480	OG5_127177	unnamed product	1.00E-109	103326
Homo sapiens	hsap|ENSP00000320324	OG5_127177	Puromycin sensitive aminopeptidase	8.00E-109	103277
Trypanosoma cruzi	tcru|Tc00.1047053509683.130	OG5_127177	aminopeptidase,	7.00E-86	55645
Leishmania major	lmaj|LmjF.29.2240	OG5_145268	aminopeptidase,	6.00E-79	96390
Trypanosoma congolense	tcon|TcIL3000.0.42640	OG5_145268	aminopeptidase,	4.00E-78	96780
Trypanosoma b. brucei	tbru|Tb927.3.4750	OG5_145268	aminopeptidase	7.00E-77	100250
Trypanosoma b. brucei	tbru|Tb927.3.4790	OG5_145268	aminopeptidase	7.00E-77	100250
Trypanosoma b. gambiense	tbrg|Tbg1002.3.5310	OG5_145268	aminopeptidase, putative	2.00E-75	100254

Organism	Accession number	Group	Product	E-value	Mw (Da)
Trypanosoma b. gambiense	tbrg|Tbg1002.3.5350	OG5_145268	aminopeptidase, putative	2.00E-75	100254
Leishmania major	lmaj|LmjF.26.0300	OG5_127177	aminopeptidase-like protein	2.00E-72	98094
Leishmania major	lmaj|LmjF.12.1250	OG5_166686	puromycin-sensitive aminopeptidase	6.00E-45	149574
Trypanosoma cruzi	tcru|Tc00.1047053506529.90	OG5_166686	puromycin-sensitive aminopeptidase	3.00E-35	133799
Trypanosoma cruzi	tcru|Tc00.1047053508593.150	OG5_166686	puromycin-sensitive aminopeptidase	3.00E-35	133587

The identified *T. b. brucei* tricorn interacting factor 3 orthologs grouped in different ortholog groups. The table is sorted by increasing e-value.

4.1.1.4 Identification of tricorn interacting factor 3 conserved domains

Conserved domain analysis revealed a total of three (3) specific conserved domains, two (2) of which were similar to the conserved domains in the archaeon tricorn interacting factor 2 and 3 (Table 4.5). These conserved domain included peptidase M1 Aminopeptidase N family (M1_APN_2) domain and ERAP1-like C-terminal domain (ERAP1_C) (Table 4.5).

Table 4.5: Identified conserved domains from NCBI CDD database

Gene Identity	Product description	From	To	E-value	Accession	Short name
CAC11944.1	Tricorn interacting factor 3, *Thermoplasma acidophilum*	6	417	0	cd09601	M1_APN_2
		474	730	4.31E-17	pfam11838	ERAP1_C
		2	777	0	COG0308	PepN
CAC11446.1	Tricorn interacting factor 2, *Thermoplasma acidophilum*	8	423	0	cd09601	M1_APN_2
		479	772	5.39E-24	pfam11838	ERAP1_C
		3	783	0	COG0308	PepN
Tb927.11.3570	Putative aminopeptidase, *Trypanosoma brucei brucei*	17	467	0	cd09601	M1_APN_2
		535	847	1.46E-39	pfam11838	ERAP1_C
		9	404	4.53E-151	Pfam01433	Peptidase_M1
Tb927.3.4750	Putative aminopeptidase, *Trypanosoma brucei brucei*	21	454	2.70E-173	cd09601	M1_APN_2
		524	847	8.31E-22	pfam11838	ERAP1_C
		14	394	7.05E-122	Pfam01433	Peptidase_M1
Tb927.3.4790	Putative aminopeptidase, *Trypanosoma brucei brucei*	21	454	2.7E-173	cd09601	M1_APN_2
		524	847	8.31E-22	pfam11838	ERAP1_C
		14	394	7.05E-122	Pfam01433	Peptidase_M1

Two (2) O,-N glycosylation sites were predicted in Tb927.11.35 70 at frame/strand 379 and frame/strand 63 with a score of 0.592 and 0.576 respectively, and Tb927.3.4750/90 at frame/strand 3 and frame/strand 6 with a score of 0.734 and 0.694 respectively. Several phosphorylation sites were predicted in both proteins with no transmembrane domain detected (Appendix I and Appendix II). Serine phosphorylation sites accounted for 51.6% of all sites, followed by threonine (37.1%) and tyrosine (11.3%) for Tb927.3.4750/90 which had a total of 62 phophorylation sites. A similar scenario was observed in Tb927.11.3570 (74 phosphoryalation sites) with 45.9%, 39.2% and 14.9% representing phosphoserine, phosphothreonine and phosphotyrosine sites respectively (Appendix III and Table 4.6).

Table 4.6: Distribution of phosphorylation sites

	Phosphoserine	**Phosphothreonine**	**Phosphotyrosine**
Tb927.3.4750/90	32 (51.6%)	23 (37.1%)	7 (11.3)
Tb927.11.3570	34 (45.9%)	29 (39.2%)	11 (14.9%)

The ratio of phosphoserine, phosphothreonine and phosphotyrosine sites in *Trypanosoma brucei brucei* tricorn interacting factor 3 orthologs

4.1.2 Sequence Alignment

A multiple sequence alignment of nine (9) sequences from Table 4.4, including the *T. b. brucei* putative aminopeptidases revealed conserved amino acid residues, exopeptidase GXMEN motif specific for M1 aminopeptidase family, zinc-binding motif {HEXXH (18X) E} where X is any residue. This conservation pattern was also observed in orthologs of other trypanosomes (Appendix V). A pairwise alignment of the gene products of Tb927.11.3570 and Tb927.3.4750/90 revealed 32% sequence identity and 69% sequence similarity to the gene product of Tb927.11.3570 (Appendix VI).

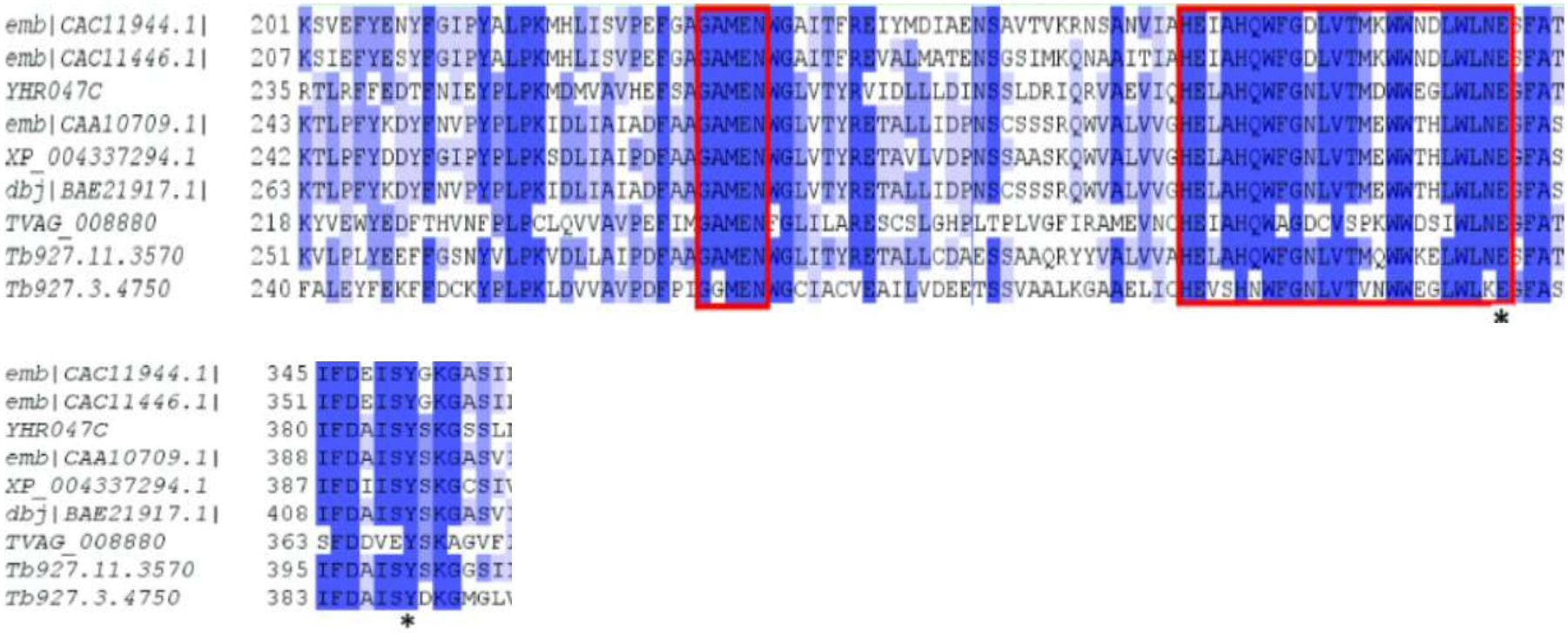


Figure 4.2: A section of multiple sequence alignment of tricorn interacting factors 3 and its orthologs The peptides sequences were retrieved from GenBank and TriTrypsDB. The GXMEN and {HEXXH (18X) E} motifs are boxed. The zinc binding ligand glutamate and the proton acceptor tyrosine are marked by *. The alignment was generated by MAFFT and edited by Jalview 2.9. Key: emb|CAA11944.1 and emb|CAA11446.1: F3 and F2, respectively, *T. acidophilum,* YHR047C: Arginine/alanine amino peptidase, *S. cerevisiae,* emb|CAA10709.1: Puromycin sensitive aminopeptidase, partial, *H. sapiens*, TVAG _008880: Aminopeptidase, *T. vaginalis,* XP_004337294.1: Hypothetical protein, *A. castellanii,* dbj|BAE21917.1: Unnamed protein product, *M. musculus,* and Tb927.11.3570, putative aminopeptidase and Tb927.3.4750, M1 family, in *T. b. brucei.* The residues were coloured by residue identity.

4.1.3 Motif occurrence and analysis

Tricorn interacting factor 3 and orthologs conserved motif analysis using MEME suite revealed 5 statistically significant motifs also identified through multiple sequence alignment. The motifs ranged from 29-50 amino acids wide with a similarity matrix of ≤ 0.3 (Figure 4.3). Motif occurrence analysis in trypanosome genome identified 16 other proteins with zinc binding (motif 1) but with relatively higher q-values.The q-value for motif 1 in Tb927.11.3570 was 1.43e-30 and 2.02e-28 for both Tb927.3.4750 and Tb927.3.4790 while the best hit (AAZ13253) among the other 16 proteins had a q-value of 0.266 (Appendix VII). A similar scenario was observed with the other motifs; motif 2 (exopeptidase motif specific for M1 family) seemed to occur in 43 other proteins but with less significant q-values relative to the tricorn interacting factor 3 orthologs. The q-value for motif 2 in Tb927.11.3570 was 3.58e-45 and 4.46e-44 for both Tb927.3.4750 and Tb927.3.4790 while the best hit (AAZ10952) among the other 43 proteins had a q-value of 0.377 (Appendix VII). The motif 3 occurred in 35 other proteins with q-value for Tb927.11.3570 was 1.79e-44 and both Tb927.3.4750 and Tb927.3.4790 had a q-value of 5.10e-43 while the best hit (EAN80621) among the 35 hits had a q-value of 0.609 (Appendix VII). The motif 4 occurred in 36 other proteins with q-value for Tb927.11.3570 was 1.71e-27 and both Tb927.3.4750 and Tb927.3.4790 being 1.01e-43 while the best hit (AAZ13515) among the 35 hits had a q-value of 0.000579 (Appendix VII). The motif 5 occurred in 35 other proteins with q-value for Tb927.11.3570 was 1.71e-47 and both Tb927.3.4750 and Tb927.3.4790 being 2.02e-15 while the best hit (EAN80603) among the 35 hits had a q-value of 0.493 (Appendix VII). However, no protein was identified with a combination of the 5 motifs identified in the tricorn interacting factor 3 orthologs.

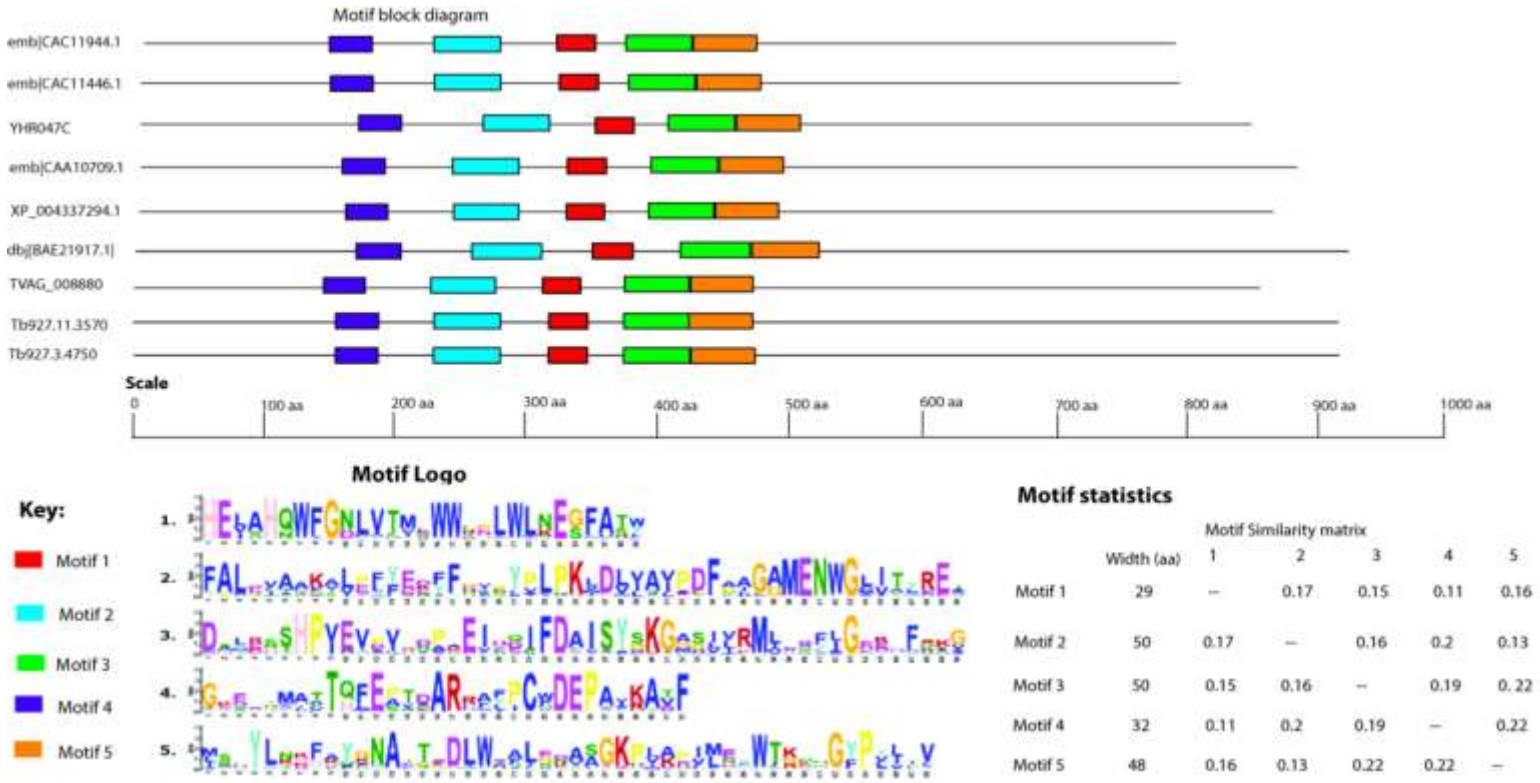


Figure 4.3: Conserved motif analysis of the tricorn interacting factor 3 and its orthologs using MEME suite. Motif 1: zinc binding motif, Motif 2: catalytic region of M1 family aminopeptidase, motif 3 contains the tyrosine proton acceptor, motif 4 and motif 5 are uncharacterized. In the motif logo, the height of the letter indicates relative frequency at a given position (x-axis) in the motif. Species key: emb|CAA11944.1 and emb|CAA11446.1: F3 and F2, respectively, *T. acidophilum*, YHR047C: Arginine/alanine amino peptidase, *S. cerevisiae*, emb|CAA10709.1: Puromycin sensitive aminopeptidase, partial, *H. sapiens*, dbj|BAE21917.1: Unnamed protein product, *M. musculus*, TVAG_008880: Aminopeptidase N-like, *T. vaginalis*, XP_004337294.1: Hypothetical protein, *A. castellanii*, putative aminopeptidases in *T. b. brucei*. (Tb927.11.3570 and Tb927.3.4750).

4.1.4 Protein-Protein Interactions

The protein-protein interaction networks of the putative tricorn interacting factor 3 orthologs in *T. b. brucei* through STRING database revealed 33 nodes with 114 edges and average node degree of 6.91. The clustering coefficient was 0.722 with a protein-protein interaction enrichment p-value of 0. Based on gene ontology annotations, the network's major functional enrichments was proteolysis (GO: 0051603) as biological process with e-value of 4.54e-14 with cellular component GO: 0005737) being cytoplasm. The major KEGG pathway for the network was proteasome (03050) however Fructose-1, 6-Bisphosphatase (FBPase), sedoheptulose-1,7-bisphosphatase (SBPase) and glycine dehydrogenase (AAZ12256) were predicted to be involved in carbon metabolism (01200). The network revealed three (3) clusters where all the tricorn interacting factor 3 orthologs interacted with aminopeptidase P1, putative (AAZ10244) which has experimentally determined interaction with polyubiquitin, putative (EAN79944) linking them to the proteasome core complex. EAN79944 also had experimentally determined interactions with other proteins including FBPase, SBPase and AAZ12256 (Figure 4.4).

Other proteins in the network included aminopeptidase, putative (EAN79326), aminopeptidase, putative (caa), aminopeptidase, putative (AAZ10508), aspartyl aminopeptidase, putative (AAZ10375), thimet oligopeptidase A, putative (opa), proteasome alpha 5 subunit, putative (EAN77455), proteasome beta 5 subunit, (EAN78022), proteasome 26 S non-ATPase subunit 9 (AAZ13338), DNA topoisomerase II (TOP2 beta), DNA topoisomerase II (TOP2), apurinic/ apyrimidinic endonuclease, putative (APE), elongation factor 1 gamma (EAN80264), cytosolic alanine aminopeptidase (lap), aminopeptidase P putative (EAN77242), aminopeptidase, putative (EAN79218), proteasome beta 3 subunit, PSB3 (EAN79693), proteasome beta 2 subunit, putative, PSB4 (EAN77890), proteasome beta 7 subunit (AAZ10634), proteasome beta 6 subunit (AAZ12543), proteasome alpha 2 subunit (EAN77461), proteasome alpha 3 subunit, putative (putative, AAZ12506 (20SPA3), proteasome beta-

1 subunit (AAZ11689), proteasome alpha 7 subunit (putative, AAZ10115 (psa7), proteasome alpha 1 subunit (EAN78018), proteasome alpha 7 subunit, (EAN79662 (PSA4), and aminopeptidase, putative (EAN79621).

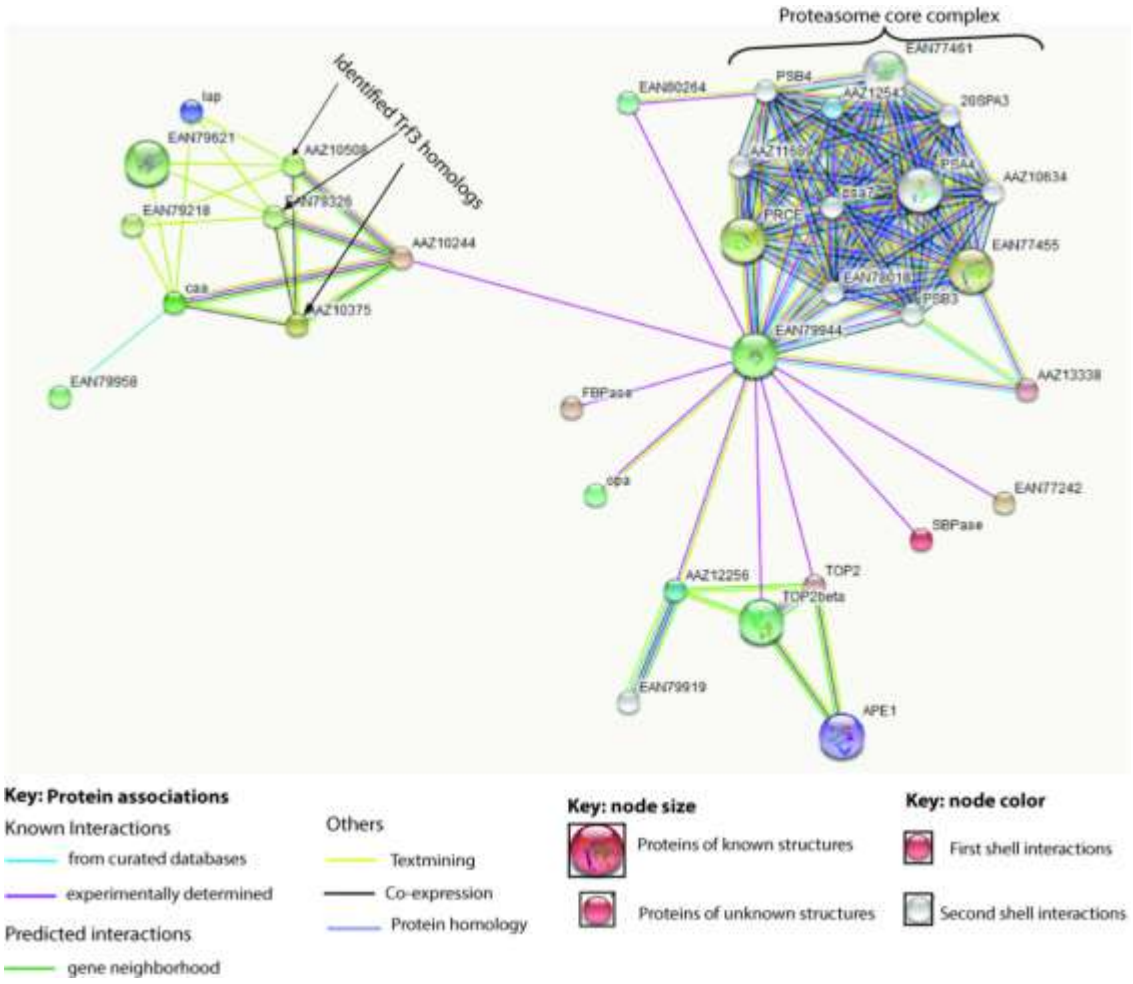


Figure 4.4: Protein-protein interaction network of the putative tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei* in STRING database. The putative orthologs are shown with the arrow mark. The interacting proteins are shown by their STRING database identity.

4.1.5 Structure prediction and validation

A monomer model was built for the tricorn interacting factor 3 orthologs in *T. b. brucei* since the oligo state of the template (1z5hA) used was a monomer. Two models were built since Tb927.3.4750 and Tb927.3.4790 were identical. The predicted models validation through PROCHECK analysis in PDBsum was based on 118 structures of resolution of atleast 2.0 Å and R-factor ≤ 20% thus a good quality model was expected to have > 90% of the residues in most favoured regions. For, Tb927.11.3570 predicted model, non-glycine and non-proline residues in the most favoured regions (A, B, L) were 707 representing 91.1%, those in additionally allowed regions (a, b, l, p), were 57 representing 7.5%, those in generously allowed regions (~a, ~b, ~l, ~p) were 9 representing 1.1 % while those in disallowed regions were 4 representing 0.5% of the residues. The number of end residues (excluding Gly and Pro) were 2, the number of glycine residues were 55 and number of proline residues were 37 (Figure 4.5). For, Tb927.3.4750 predicted model, non-glycine and non-proline residues in the most favoured regions (A, B, L) were 681 representing 88.1%, those in additionally allowed regions (a, b, l, p) were 77 representing 10.0%, those in generously allowed regions (~a, ~b, ~l, ~p) were 11 representing 1.4 %, while those in disallowed regions were 4 representing 0.5 % of the residues. The number of end residues (excluding Gly and Pro) were 2, number of glycine residues were 58 and number of proline residues were 38 (Figure 4.5).

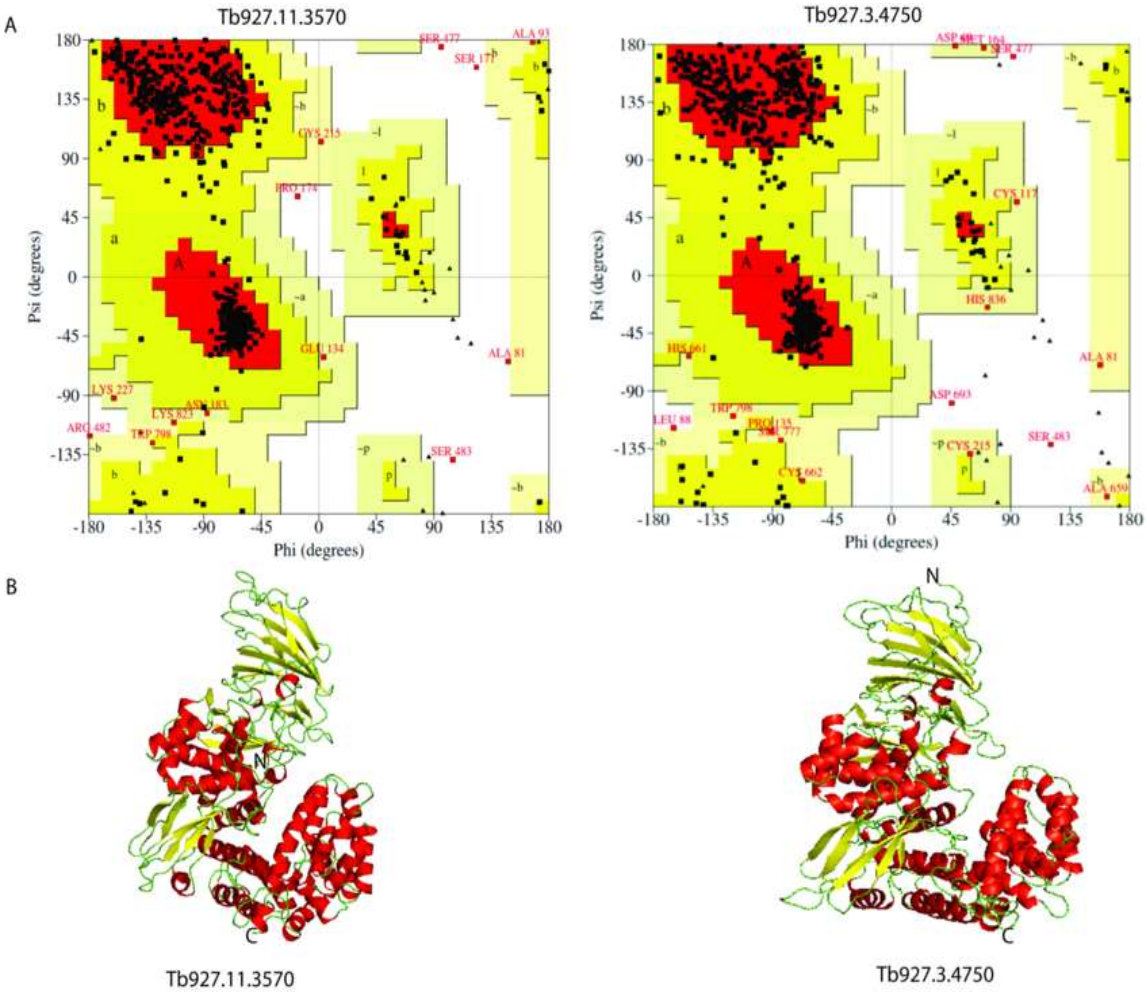


Figure 4.5: (A): Ramanchandaran plot analysis of the predicted models. The black squares (■) represent all residues in in the most favoured regions and in additionally allowed regions while the red squares (■) represent residues in generously allowed regions and in disallowed regions. The triangles (▲) represent proline residues. For Tb927.11.3570 predicted model, a total of 13 residues were found in generously allowed regions and in disallowed regions representing about 1.5% of total residues. For, Tb927.3.4750 predicted model, a total of 27 residues were found in generously allowed regions and in disallowed regions representing about 3.1% of total residues. Figure 4.7 (B) Predicted 3-D structures of the tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei* colored according to the secondary structures; yellow

represents the β-sheets , α-helices are shown in red, while loops are shown in green colour.

The overall structure of the tricorn interacting factor 3 orthologs in *T. b. brucei* was shown to have four domains; N-terminal domain I predominantly composed of β-sheets, catalytic domain II composed of mixed α/β structure, domain III composed of β-sheets and C-terminal domain IV composed of α-helices, organized into a super helix. For Tb927.11.3570 gene product, domain 1 composed of 13 β-sheets and 2 α-helices, domain II; 6 β- sheets and 11-α helices, domain III; 6 β-sheets, and domain IV; 18 α-helices. For Tb927.3.4750 gene product, domain I composed of 12 β-sheets and 2 α-helices, domain II; 7 β- sheets and 12-α helices, domain III; 7 β-sheets, and domain IV; 16 α-helices (Figure 4.6).

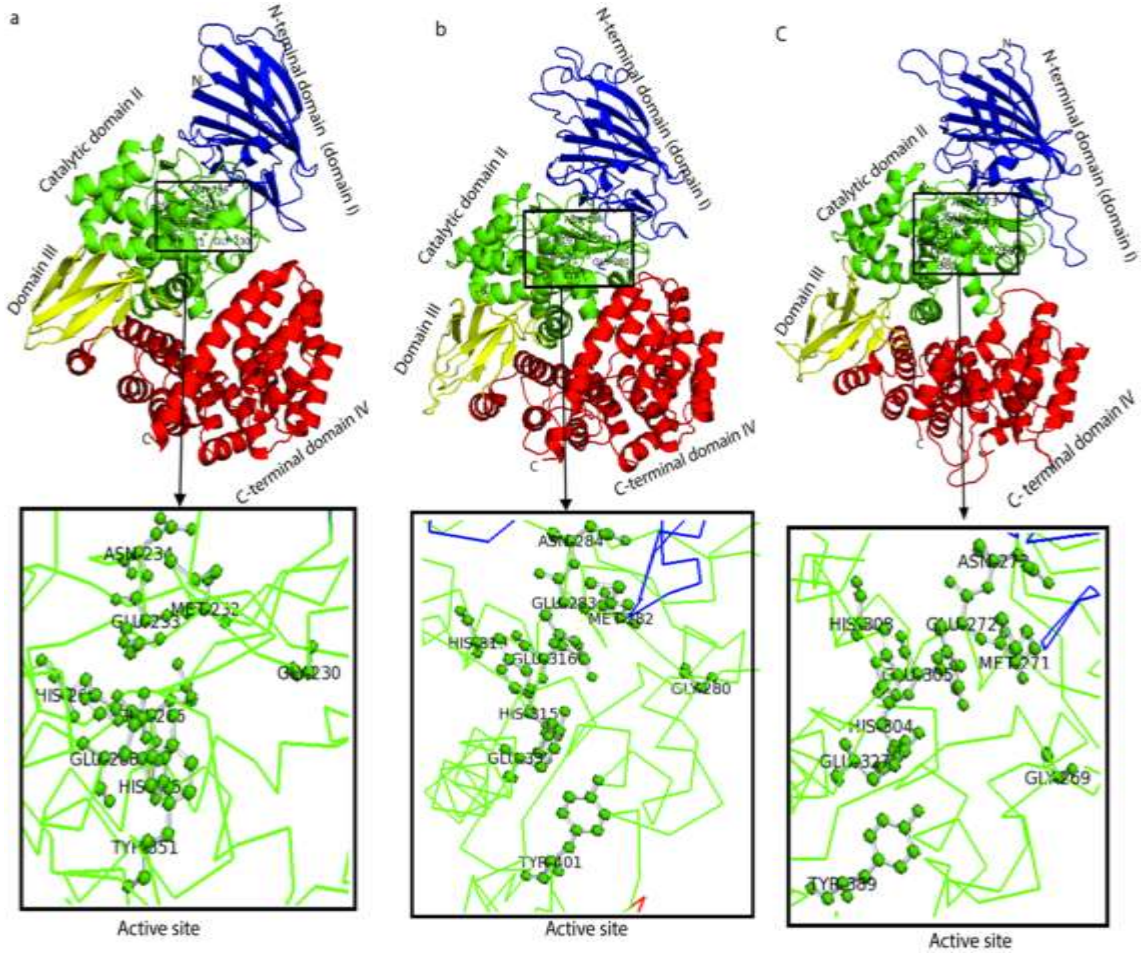


Figure 4.6: A cartoon representation of the overall structure and domain organization the F3 and *Trypanosoma brucei brucei* orthologs. The proteins were coloured according to domains; blue; N-terminal domain (domain I), green; catalytic domain II, yellow; domain III and red; C-terminal domain IV. (a): F3, *Thermoplasma acidophilum*, (b); aminopeptidase, putative (Tb927.11.3570) and aminopeptidase, M1 family (Tb927.3.4750), *Trypanosoma brucei brucei*. The conserved active site is shown as ball and stick while the protein backbone is shown in ribbon representation.

4.1.6 Phylogenic tree of tricorn interacting factors 3 orthologs in *Trypanosoma*

The phylogenetic tree calculated from the alignment of tricorn interacting factor 3 orthologs sequences indicated different evolutionary pathways with two main branches with Tb927.3.4750 and Tb927.3.4790, representing possible gene duplication, tended to cluster together on the same branch. Other species where gene duplication could have also occurred include *T. b. gambiense* and *T. evensi* (Figure 4.7). *Trypanosome congolense* and *T. vivax* seemed to have two (2) orthologs each which clustered in different branches (Figure 4.7). The American trypanosome, *T. cruzi* orthologs had a similar clustering pattern with *T. b. brucei* orthologs (Figure 4.7).

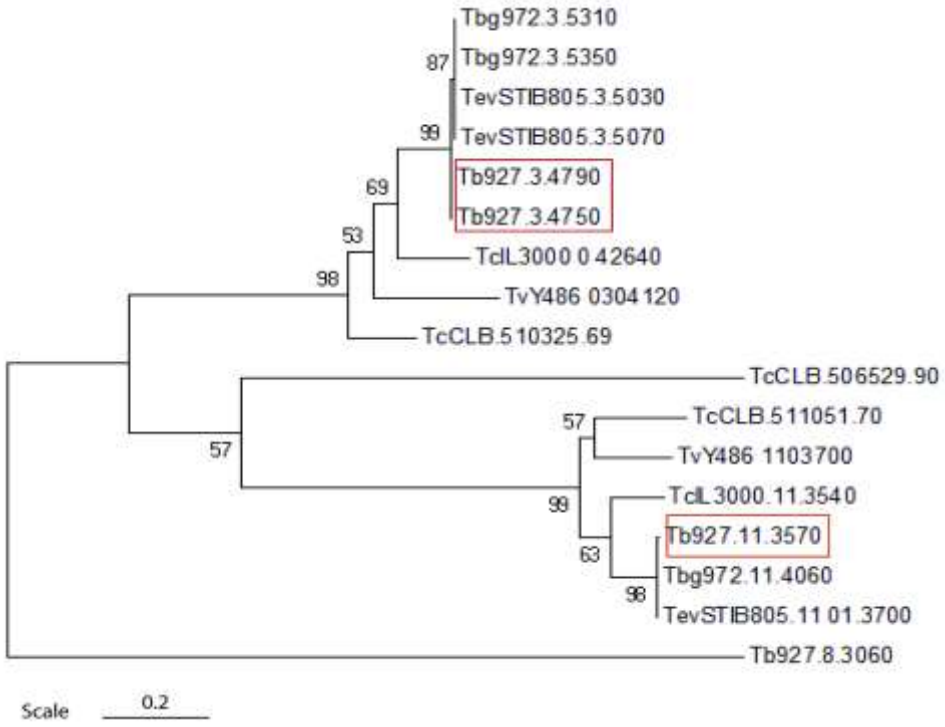


Figure 4.7: A phylogenetic tree of tricorn interacting factor 3 orthologs in trypanosomes built by Maximum (ML) Likelihood method based on the JTT matrix-based model. The phylogram was constructed from the alignment of amino acid sequences of tricorn interacting factor 3 and their orthologs in other trypanosomes. The red boxes indicate the *T. b. brucei* tricorn interacting factor 3 orthologs. The node value given as percentages represent the measure of support for the node where 100 represents maximal support, that is, sequences to the right node cluster together to the exclusion of any other. A cytosolic leucyl aminopeptidase from *Trypanosoma b. brucei* (Tb927.8.3060) was used as an outgroup. The tree was drawn to scale with bootstrap confidence levels of 1000 replicas.

4.2 Molecular characterization of tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei*

4.2.1 DNA Electrophoresis

The presence and purity of the extracted genomic DNA was assessed using a NanoDrop 2000c Spectrophotometer (Thermo Scientific, Waltham, MA, USA). The OD at A_{260} was 24.657 while A_{280} was 13.857 for the extracted genomic DNA thus $A_{260/280}$ ratio was 1.78 hence considerably pure DNA and the concentation was 1232.8ng/µl (Appendix VIII). The integrity of the DNA was assessed through agarose gel electrophoresis (Figure 4.8 a). Molecular characterization through PCR amplification produced an approximately 285 and 325 bp amplicons representing (1299 to 1583 base and 1093 to 1415 base) of Tb927.11.3570 and Tb927.3.4750/90 respectively (Figure 4.8 b). The open reading frame (ORF) for both the genes was 2616 bp encoding 871 amino acid long peptides.

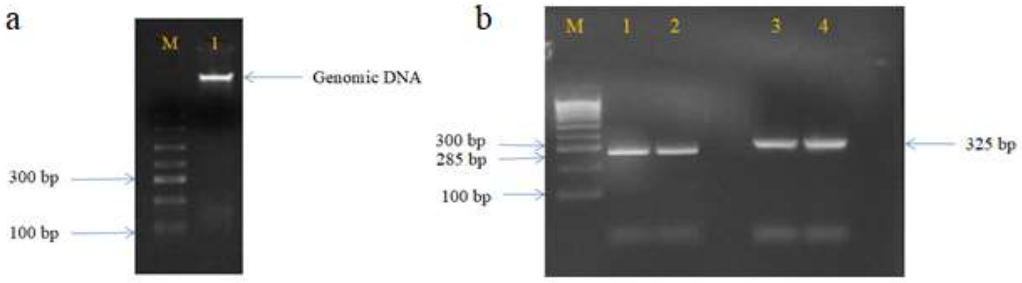


Figure 4.8: a: Agarose gel electrophoresis of *T. b. brucei* genomic DNA. 1% ethidium bromide stained agarose gel, 65 volts for 1 hour 30 minutes. M: Molecular weight marker; HyperLadder 100bp (Bioline), Lane 1: *T. b. brucei* genomic DNA. b; Lanes 1 and 2; Tb927.11.3570, product ≈ 285 bp while lanes 3 and 4; Tb927.3.4750/90, product ≈ 325 bp. 5 µl samples were loaded.

4.2.2 Total RNA quantification and electrophoresis

The presence and purity of the extracted RNA was assessed using a NanoDrop 2000c Spectrophotometer (Thermo Scientific, Waltham, MA, USA). The OD at A_{260} was 35.306 while A_{280} was 17.35 for the extracted total RNA thus $A_{260/280}$ ratio was 2.03 hence considerably pure RNA. The RNA concentration was 1412.2ng/µl (Appendix VIII). The integrity of the RNA was assessed through non-denaturing and denaturing agarose gel electrophoresis. The non-denaturing gel showed a brighter distinct band around the 1kb band of the molecular marker indicating that RNA was intact while the denaturing formaldehyde agarose gel showed the 28S and 18S bands (Figure 4.9).

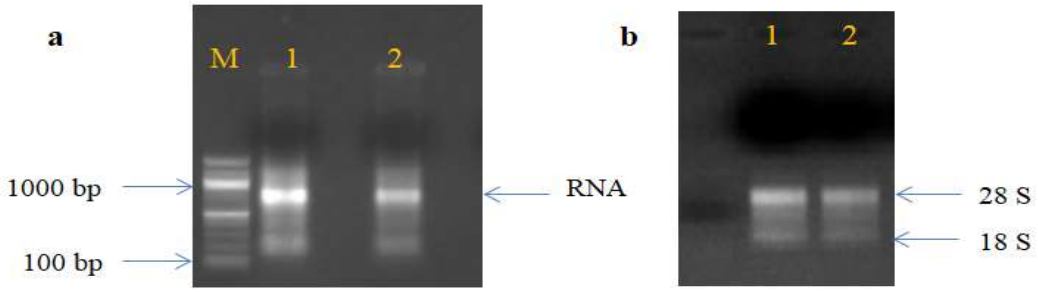


Figure 4.9: (a): Non -denaturing RNA analysis on 1.2 % ethidium bromide stained agarose gel. 2µl sample was loaded, 65 volts for 1 hour. Lanes 1 and 2; extracted total RNA, M: Quickload 100 bp ladder (NEBs). Figure 4.9 (b): Formaldehyde denaturing RNA analysis on 1.2% ethidium bromide stained agarose gel, lanes 1 and 2; denatured RNA. 5µl samples were loaded, voltage 65 for 1 hour.

4.2.3 First strand cDNA analysis

The synthesis of first strand cDNA analysed through PCR amplification and gel electrophoresis produced ≈ 285 bp and ≈ 325 bp amplicons for Tb927.11.3570 and Tb927.3.4750/90 respectively (Figure 4.10).

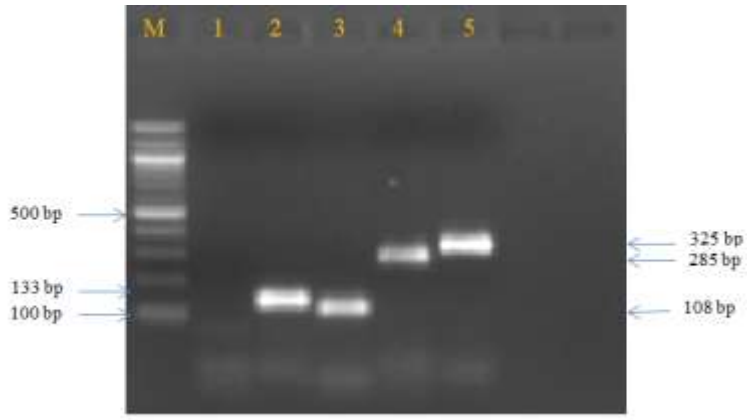


Figure 4.10: Analysis of first strand cDNA. Electrophoresis was done in 1.2 % agarose at 65 volts for 1 hr 30 min in 1X TAE buffer. The gel was stained with ethidium bromide. DNA Marker: Quickload 100 bp ladder (NEBs). Lane 1: No template control, lane 2; Actin (133 bp), lane 3; TERT (108 bp), lane 4; Tb927.11.3570 (285 bp) and lane 5; Tb927.3.4750/90 (325 bp). 10 µl samples were loaded.

4.2.4 Real time Quantitative PCR (qPCR)

The cycle threshold (Ct) values for the putative tricorn interacting factor 3 orthologs were 16.71 and 16.39 for Tb927.11.3570 and Tb927.3.4750 respectively, while the reference actin genes and TERT had a Ct value of 15.58 and 16.8 respectively with the no template control Ct value being 39.93. The cycle thresholds were <29 for the target genes (Figure 4.11 a). The dissociation curves produced single peaks with melting temperatures > 80°C (Figure 4.11 b).

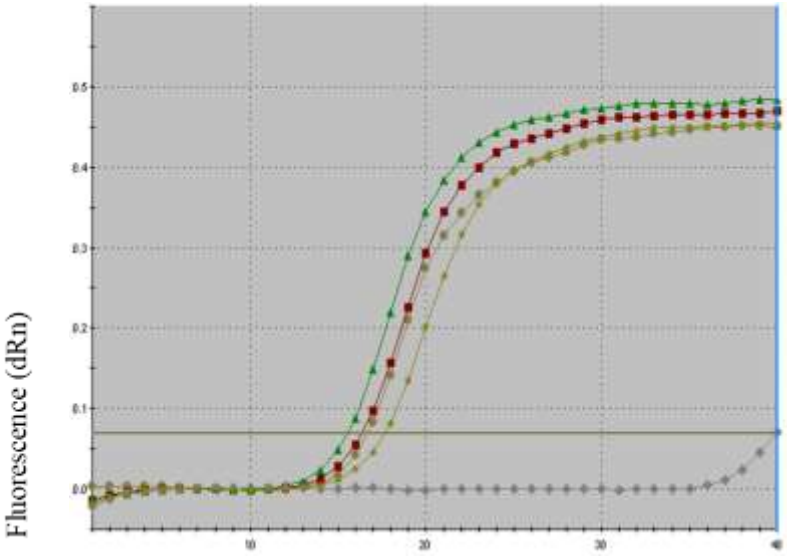


Figure 4.11 a: Amplification plots for tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei*. The reactions were set up in triplicates which were treated as one sample for plotting. NTC: No template control.

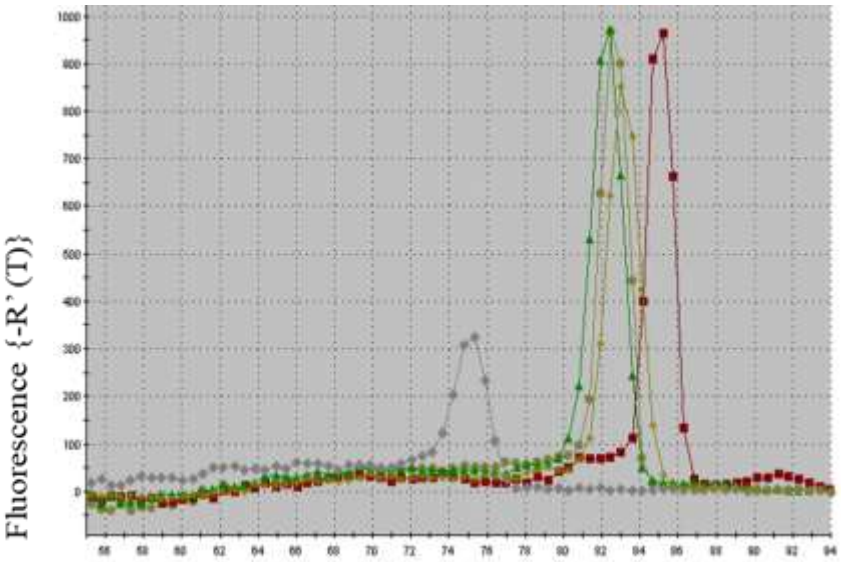


Figure 4.11 b: The dissociation curves for tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei*. The reactions were set up in triplicates which were treated as one sample for plotting. The PCR products melted at temperatures > 80°C.

4.2.5 Blue-white selection

White colonies represented the recombinant clones while blue colonies represented non-recombinant clones (Figure 4.12).

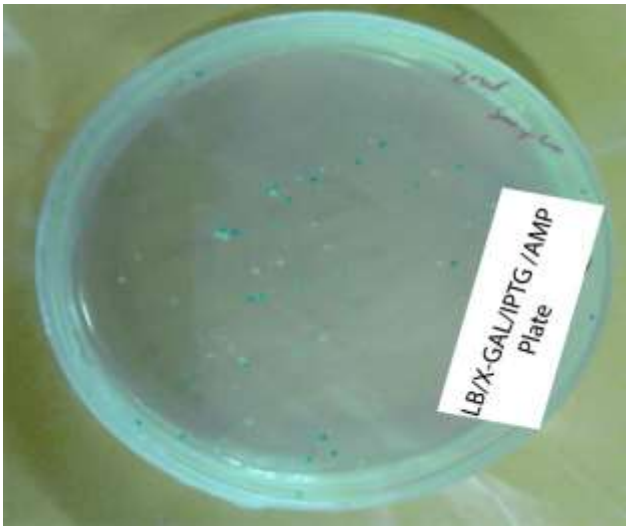


Figure 4.12: Blue-white colour selection. β-galactosidase hydrolyzes 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (X-GAL) into colorless galactose and 4-chloro-3-brom-indigo, forming an intense blue precipitate thus white colonies were recombinant.

Expansion and purification of the target genes for expression produced approximately 285 and 325 bp amplicons representing Tb927.11.3570 and Tb927.3.4750/ 90 respectively (Figure 4.13 a). Ligation of the target genes into the expression vector and screening for positive clones through colony PCR also produced an approximately 285 and 325 bp amplicons representing Tb927.11.3570 and Tb927.3.4750/90 respectively (Figure 4.13 b).

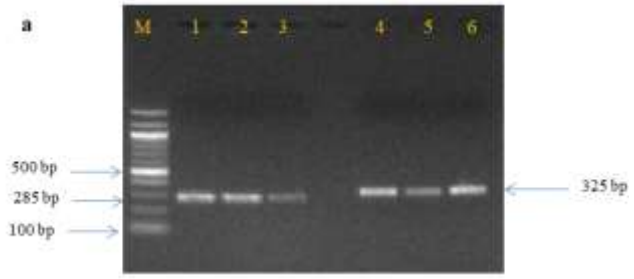


Figure 4.13 a: Electrophoresis of purified products for expression. Electrophoresis was done on 2 % agarose ethidium bromide stained at 90 volts for 1 hr in 1X TAE buffer. M: Quickload 100 bp ladder (NEBs). Lanes 1-3; Tb927.11.3570, Lanes 4 - 6; Tb927.3.4750/90.

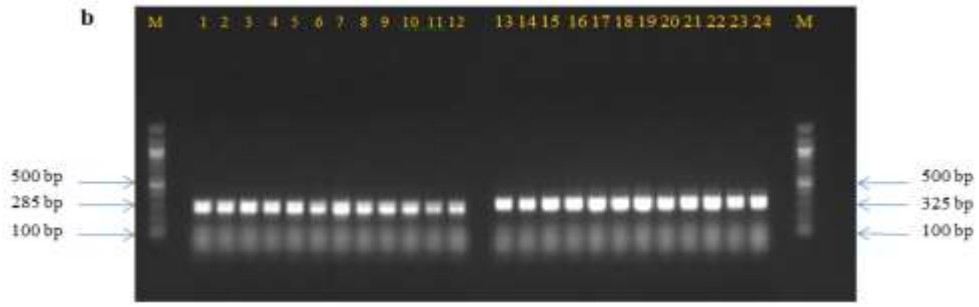


Figure 4.13 b: Electrophoresis of colony PCR products of *E. coli* BL-21 (DE3) pLysS cells transformed with recombinant pGEX-3x-5 plasmids. M: Quickload 100 bp DNA ladder (NEBs), Lanes 1-12; Tb927.11.3570, Lanes 13-24; Tb927.3.4750/90.

4.2.6 Analysis of expressed proteins through SDS-PAGE

The presence of expressed proteins of intersest were shown by the presence bands whose intensity increased with time (Figure 4.14).

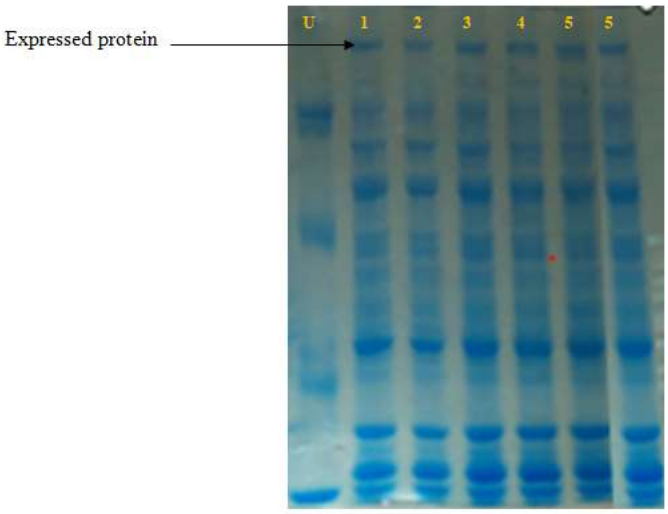


Figure 4.14: SDS-PAGE analysis of expressed products. Time course experiment, where U: Uninduced sample, Lanes 1-2: Time I; 3 and 4, Time II, 5 and 6, Time III. Lanes 1, 3 and 5 represent Tb927.11.3570 while lanes 2, 4 and 6 represent Tb927.3.4750/90.

4.3. RNAi studies

4.3.1 Preparation of cDNA

The synthesized cDNA template incorporating BamHI and HindIII restriction sites also produced approximately 325 bp amplicons for Tb927.3.4750/90 and 285 bp for Tb925.11.3570. Gradient PCR was performed with a range of 12 different annealing temperatures from 55°C for lanes 1 and 13, 55.5°C for lanes 2 and 14, 56.5°C for lanes 3 and 15, 58.4°C for lanes 4 and 16, 60.8°C for lanes 5 and 17, 63.9°C for lanes 6 and 18, 67.5°C for lanes 7 and 19, 70.8°C for lanes 8 and 20, 73.3°C for 9 and 21, 75.3°C for lanes 10 and 22, 76.4°C for lanes 11 and 23, and 77°C for 12 (Figure 4.15 a). The products were gel purified prior to transfection (Figure 4.15 b).

Figure 4.15 a: Analysis of cDNA for RNAi. Electrophoresis was done in 1.2 % agarose at 65 volts for 1hr 30 min in 1X TAE buffer. The gel was stained with ethidium bromide. 10 µl sample was loaded. DNA Marker (M): Quickload 100 bp DNA ladder (NEBs), Lanes 1-12; Tb927.3.4750/90 and lanes 13-23; Tb927.11.3570.

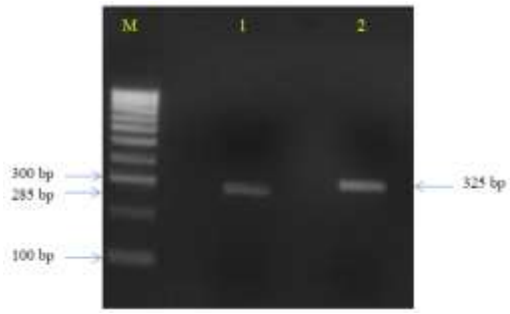


Figure 4.15 b: Analysis of purified cDNA for RNAi. Electrophoresis was done in 1.2 % agarose at 65 volts for 1 hr 30 min in 1X TAE buffer. Gel was stained with ethidium bromide. M: DNA Marker, HyperLadder 100 bp (Bioline). Lane 1; Tb927.11.3570, product ≈ 285 bp and lane 2; Tb927.3.4750, product ≈ 325 bp.

4.3.2 Restriction digestion of the recombinant plasmid

To confirm the presence of the insert cDNA targeting either Tb927.11.3570 or Tb927.3.4750/90, restriction digestion was done on the recombinant plasmid generating the insert band and the plasmid backbone (Figure 4.16).

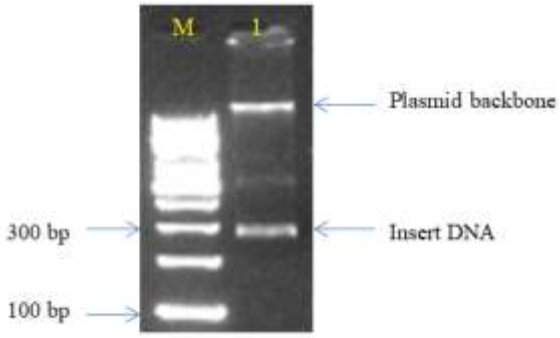


Figure 4.16: Analysis of digested recombinant p2T7-177 plasmid. Electrophoresis was done in 1% agarose at 70 volts for 1 hr 30 min in 1X TAE buffer. Gel was stained with ethidium bromide. M: DNA Marker, HyperLadder 100 bp (Bioline), 1: Digested plasmid.

4.3.3 RNAi validation

The extracted total RNA appeared as a smear on agarose gel (Figure 4.17 a). The relative expression of mRNA in control *T. b. brucei* (uninduced) produced sharper bands as compared to that of the tetracycline induced RNAi mutants (Figure 4.17 b). The ΔΔCt values for Tb927.11.3570 and Tb927.3.4750/90 was 1.57 and 1.48 respectively (Table 4.7) and through Livak method, the tetracycline induced RNAi mutants expressed tricorn interacting factor 3 orthologs genes at 2.97 and 2.79 fold lower than the control for Tb927.11.3570 and Tb927.3.4750/90 respectively (Table 4.7 and Figure 4.17). The percentage knockdown was 57% for Tb927.11.3570 and 48% for Tb927.3.4750/90 (Table 4.7). The standard deviation across the quantitative PCR replicates was ≤ 0.22 with no significant variations between the mean Ct values and Ct values of the triplicate experiments (Table 4.7). Based on paired t-test, the expression of Tb927.11.3570 and Tb927.3.4750/90 was significantly different between tetracycline induced RNAi mutants and control *T. b. brucei* with p-values of 0.007 and 0.003 respectively. However, the p-value for the reference gene (actin) was > 0.05 indicating no significant difference in expression in tetracycline induced RNAi mutants and control *T. b. brucei* (Table 4.7 and Figure 4.17). The dissociation curves also produced

single peaks with melting temperatures > 80°C (Figure 4.18). The tetracycline induced RNAi mutants showed reduced growth as compared to control (uninduced) *T. b. brucei* (Table 4.8, Figure 4.19 and Figure 4.20).

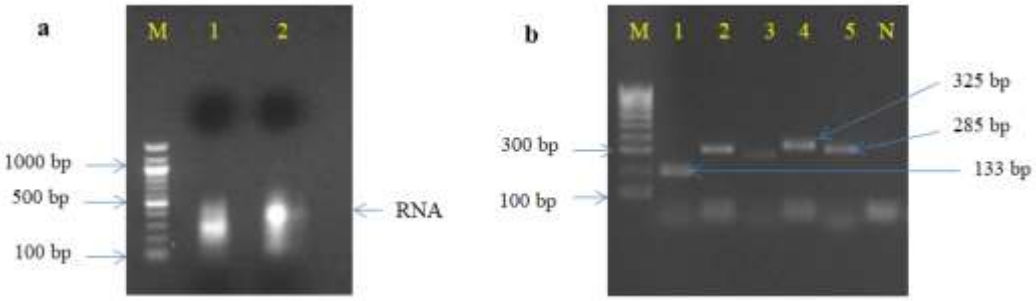


Figure 4.17: a: Non -denaturing RNA analysis on 1% EtBr stained agarose gel. 5µl sample was loaded, voltage 65 for 1 hour. M: Quickload 100 bp (NEBs); Lanes 1 and 2, extracted total RNA. Figure 4.17 b: Agarose gel electrophoresis of q RT-PCR products. Electrophoresis was done in 1.2 % agarose at 65 volts for 1 hr 30 min in 1X TAE buffer. Gel was stained with ethidium bromide. M: HyperLadder 100 bp (Bioline), Lane 1: Actin, Lanes 2 and 4; Tb927.3.4750/90 amplicons in tetracycline induced RNAi mutants and control *T. b. brucei* respectively, while lanes 3 and 5 represent Tb927.11.3570 amplicons in tetracycline induced RNAi mutants and control *T. b. brucei* respectively. 5 µl samples were loaded.

Table 4.7: Fold changes in expression and knockdown efficiency of tricorn interacting factor 3 in *Trypanosoma b. brucei*

	T	Df	p-value	Mean Ct	Std. Deviation	95% Confidence Interval of the Difference				
						Lower	Upper			
One tailed t-test										
Control parasites										
Tb927.11.3570	1896.63	2	0.009	16.727	0.015	16.689	16.765			
Tb927.3.4750/90	600.46	2	0.012	16.383	0.047	16.266	16.501			
Actin	256.34	2	0.001	14.800	0.100	14.552	15.048			
Tetracycline induced RNAi mutants								$\Delta\Delta_{Ct}$	**%KD**	$2^{-(\Delta\Delta ct)}$
Tb927.11.3570	214.94	2	0.020	18.267	0.170	17.997	18.538	1.57	57	-2.97
Tb927.3.4750/90	378.23	2	0.017	17.833	0.094	17.683	17.983	1.48	48	-2.79
Actin	313.22	2	0.002	14.768	0.094	14.618	14.918			
Paired t-test										
	T	**Df**	**p-value**	**Mean**	**Std. Deviation**	**Std. Error**	**95% Confidence Interval of the Difference**			
							Lower	**Upper**		
Tb927.11.3570-Tb927.11.3570$_{\text{tetracycline induced RNAi mutants}}$	-12.062	2	0.007	-1.540	0.221	0.128	-2.089	-0.991		
Tb927.3.4750/90-Tb927.3.4750/90$_{\text{tetracycline induced RNAi mutants}}$	-18.268	2	0.003	-1.450	0.138	0.079	-1.792	-1.109		
Actin-Actin$_{\text{tetracycline induced RNAi mutants}}$	0.378	2	0.742	0.333	0.153	0.088	-0.346	0.413		

Ct: Cycle threshold, $\Delta\Delta_{Ct} = (\Delta_{CtT} - \Delta_{CtR})$ where CtT is the mean Ct of tetracycline induced RNAi mutants while CtR is the mean Ct of Reference gene (Actin), % KD (Knockdown efficiency) = $(1-\Delta\Delta_{Ct})$, Livak method = $2^{-(\Delta\Delta ct)}$, t: calculated student's t-value, df: degrees of freedom.

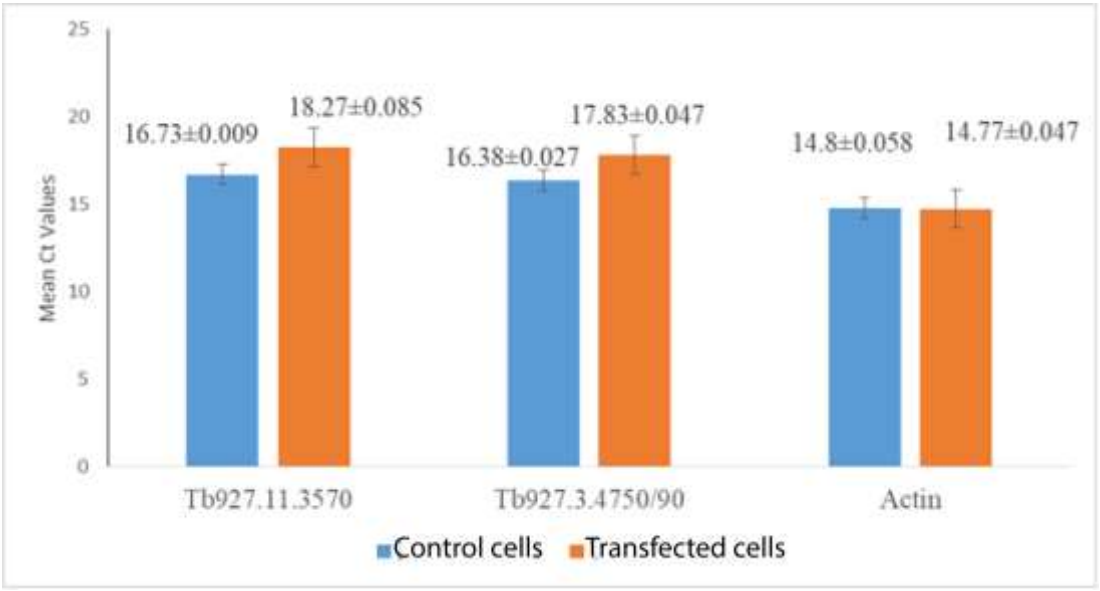


Figure 4.18: Graphical representation of relative expression of tricorn interacting factor 3 orthologs mRNA in tetracycline induced RNAi mutants (transfected parasites) and control *Trypanosoma brucei brucei*. Higher the Ct values indicate lower mRNA expression.

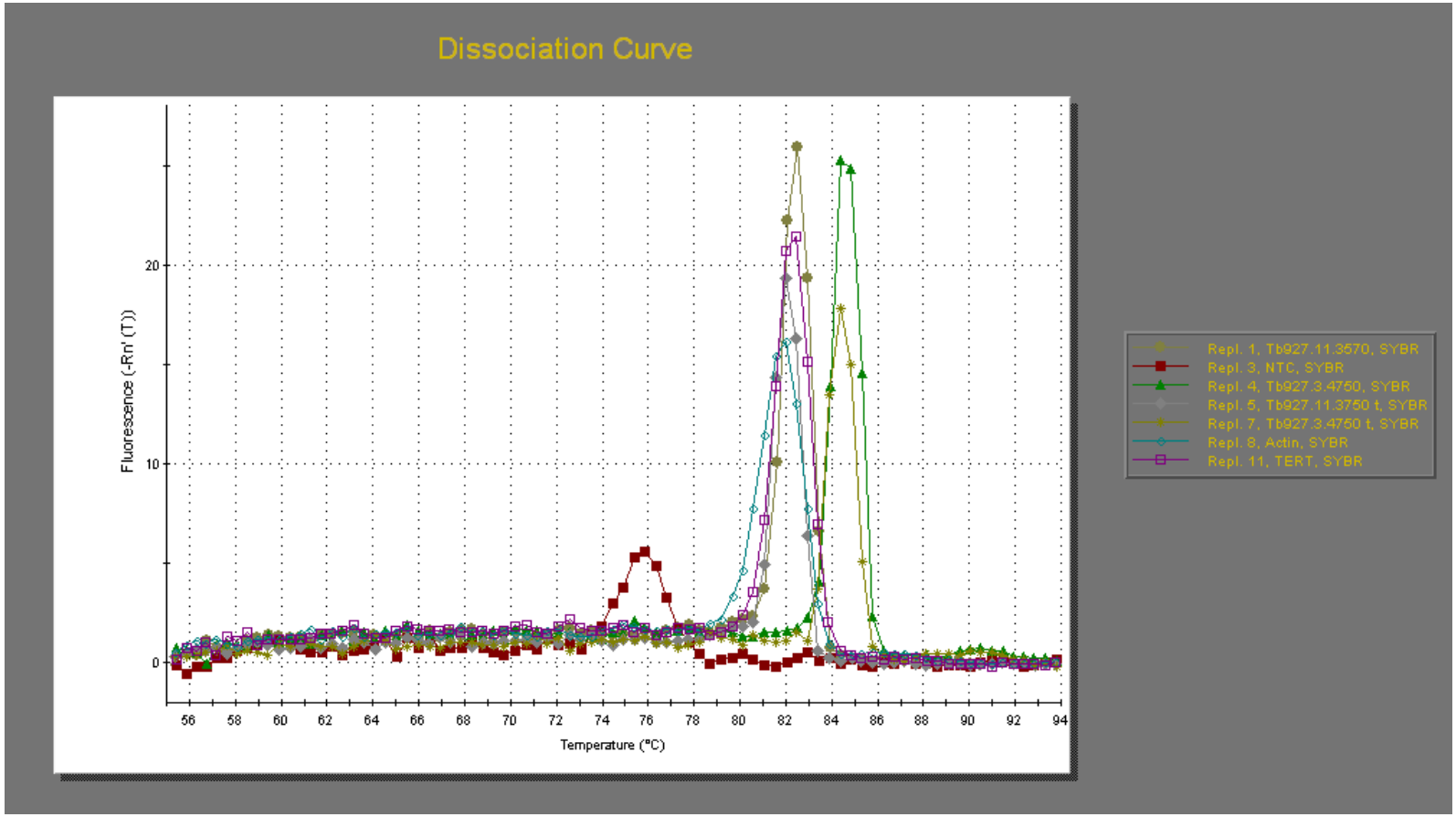


Figure 4.19: The dissociation curves for the tricorn interacting factor 3 orthologs in *T. b. brucei* in control and tetracycline induced RNAi mutants. Tb927.11.3570t and Tb927.3.4750t represent the tetracyclin induced RNAi mutants.The reactions were set up in triplicates which were treated as one sample for plotting. The PCR products melted at temperatures > 80°C.

Table 4.8: *Trypanosoma brucei brucei* densities

Time in hours	Parasite densities					
	Tb927.11. 3570$_{RNAi}$	**Log$_2$**	**Tb927.3. 4750/90$_{RNAi}$**	**Log$_2$**	**Control**	**Log$_2$**
Transfection	3 x 10^6	21.517	3 x 10^6	21.517	3 x 10^6	21.517
0 (RNAi induction)	2 x 10^3	10.966	3 x 10^3	11.551	4 x 10^3	11.659
24	5 x 10^3	11.551	6 x 10^3	12.551	5 x 10^4	15.610
48	4 x 10^4	15.288	5 x 10^4	15.610	4 x 10^6	21.932
72	3 x 10^5	18.195	4 x 10^5	18.610	6 x 10^6	22.517
96	4 x 10^5	18.610	5 x 10^5	18.932	3 x 10^7	24.839
120	5 x 10^5	18.932	6 x 10^5	19.195	5 x10^7	25.575

Cell counts were done 24 hours after RNAi induction using Neubauer chamber.

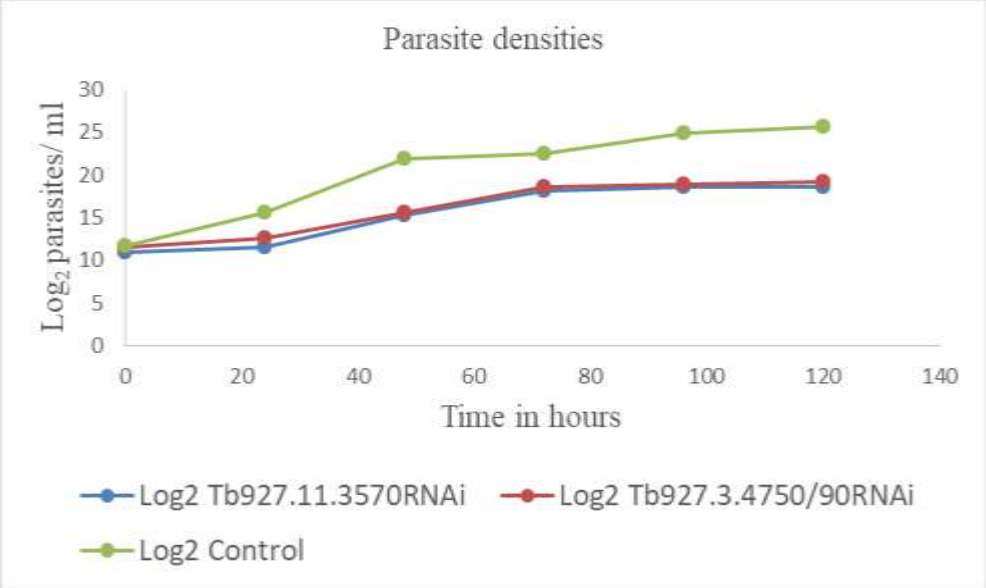


Figure 4.20: Effect of RNAi mediated decrease in Tb927.11.3570 and Tb927.3.4750/90 expression on the growth of *T. b. brucei in vitro*.

4.4 Data analysis

Sequence inference was done using BLAST algorithm which produced significant E-values ranging from 3.00E-74 to 6.00E-117. Hidden Markov Models and CDART hits E-values ranged from 1.19E-05 to 2.44E-33. Thus based on E-values, the identified hits had statistically significant matches. Domain conservation was also analysed based on the E-values while clustering coefficient was used to assess the protein-protein interactions. Ramanchandaran plot analysis was used to validate the predicted models. The clustering in phylogenetic tree was assessed by bootstrap resampling with all clusters having > 50% bootstrap values.

Polymerase chain reaction amplicons were sequenced at Macrogen and analysed using BioEdit. The sequenced products were used as query sequences to identify the tricorn interacting factor 3 orthologs in TriTrypsDB. For differential expression gene analysis, the study applied a popular and well-reviewed statistical method, the t-test. One tailed t-test was used determine the significance of the relative mRNA expression of the tricorn interacting factor 3 orthologs in normal bloodstream form of *T. b. brucei* which were shown by $p < 0.05$ at 95% confidence interval assuming binomial distribution model, indicating significant expression. To test the differential mRNA expression of Tb927.11.3570 and Tb927.3.4750/90 in tetracycline induced RNAi mutants and control *T. b. brucei*, paired t-test was used at 95% confidence interval and genes were considered to be differentially expressed if p-value was < 0.05. The test statistics showed $p = 0.007$ for Tb927.11.3570 and $p = 0.003$ for Tb927.3.4750/90 indicating that the genes were significantly differentially expressed in tetracycline induced RNAi mutants and control *T. b. brucei*. This was validated by the reference gene, actin whose differential mRNA expression was insignificant in tetracycline-induced RNAi mutants and control parasites based on p-value which was 0.742. The computed fold change values were calculated using the $2^{\wedge(\text{Log}_2\text{FoldChange})}$ equation in excel and the Log_2 fold changes obtained were plotted generating a linear graph which estimated the parasite growth over time.

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Discussion

The identification criteria used in this study failed to detect genes encoding tricorn protease in *T. b. brucei* genome which could suggest possible loss of function during evolution of kinetoplastids or a deep evolutionary divergence. Studies have shown that the C-terminal region of tricorn protease has detectable structural homology with several C-terminal processing proteases (CTP) and tail specific proteases of bacterial and eukaryotic origin (Kim *et al.,* 2002). This is consistent with findings in this study which identified several homologs in bacteria and some eukaryotes (Appendix IX). The carboxy-terminal regions of tricorn protease homologs aligned with carboxy-terminal proteases and D1 processing protease with conservation of active site residues suggesting conservation of function (Appendix X). C-terminal processing proteases contribute to virulence in some pathogenic bacteria including *Staphylococcus aureus*, *Escherichia coli* (Shaw, 2014).

The possible lack of tricorn protease in *T. b. brucei* genome as predicted by this study is consistent with previous studies which suggested that tricorn protease may be limited to archaea and eubacteria and proposed existence of functional analogous (Borissenko and Groll, 2005). Tetrahedral aminopeptidase (TET) is a complementary protein degradation machinery to tricorn protease in *Pyrococcus horikoshii* (Borissenko and Groll, 2005). In eukaryotes, tripeptidyl peptidase II (TPP II) acts downstream of the proteasome and has been described as tricorn protease functional analog (Rockel *et al.,* 2002) (Tomkinson, 1999). It is implicated in peptide processing of MHC class I antigens (Reits *et al.,* 2004) (Glas *et al.,* 2010). The patchy distribution of tricorn protease suggests existence of variant pathways in which functional homologs replace the function of tricorn (Tamura *et al.,* 2001). Other studies suggested that detecting tricorn protease analogues in eukaryotes might be difficult since tricorn is built from

five folding domains, thus it is possible that tricorn analogues might assemble non covalently from different gene products and could have additional functionalities (Brandstetter *et al.,* 2001).

Trypanosoma b. brucei proteasome has been characterized and shown to have both chymotrypsin-like, caspase and trypsin-like activities generating a pool of oligopeptides in the cytosol that are further degraded free amino acids for other metabolic processes (Wang *et al.,* 2003) (Groll *et al.,* 2005). This is unlike the *Thermoplasma acidophilum* proteasome which has been shown to have greater chyomotrypsin-like activity and reduced trypsin-like, caspase-like proteolytic activities (Groll *et al.,* 2005). Proteasomes are highly selective and at the same time, degrade an enormously diverse set of substrates. Other proteases which could possibly substitute the functions of the proteasome have the potential for self-compartmentalization (Yao and Cohen, 1999). Giant proteases may cooperate with proteasomes, and possibly even replace them (Yao and Cohen, 1999) (Glas *et al.,* 2010). Tripeptidyl peptidase II, a serine peptidase that removes amino-terminal tripeptides from unblocked oligopeptides, like tricorn protease, can assemble into a higher-order structure that has an internal channel thus has been shown to function like tricorn protease (Rockel and Baumeister, 2008). *Trypanosoma b. brucei* have genes encoding tripeptidyl peptidase, thus could substitute some functions of tricorn protease (Tomkinson, 1999), further supporting the findings of this study.

This study identified 21 *Trypanosoma b. brucei* proteins with isolated tricorn-like domains including tricorn protease N-terminal domain and tricorn protease domain 2 which are WD40 domains exhibiting β-propeller architecture. These are abundant and top interacting domains in eukaryotic genomes, though they have not been found to display enzymatic activities (Ponting and Pallen 1999). WD40 and PDZ domains have been described in many eukaryotic cells where they mediate protein interactions (Ponting and Pallen, 1999). Tricorn protease N-terminal domain was identified in Intraflageller transport protein (Tb927.10.14470), whose knockdown seems to blocks flagellum construction and results in accumulation of IFT-B proteins in the flagellum

thus is essential for intraflagellar transport (Glover *et al.,* 2015). Tricorn protease domain 2 was mainly found in hypothetical proteins as the only conserved domain. The domain was also found in translation initiation factor (Tb 927. 5. 2570), which mediates translation in *T. brucei* (Li *et al.,* 2017) and also in Coatomer alpha subunit (Tb927. 4.450), cytosolic protein complex which forms a coat around vesicles budding from the Golgi apparatus thus may play a role in intracellular transport. Some of the hits had PDZ domains which mediate protein-protein interactions by binding to the C-terminus of their target protein (Hung and Sheng, 2002). Majority of the hits had PDZ domain as the only conserved domain while others like TOR-like 1 (Tb927.4.800), the PDZ Domain also existed in combination with other different partner domains including the catalytic domain of Target of Rapamycin (TOR). TOR is a central component of the eukaryotic growth regulatory network and controls the expression of many genes transcribed by all three RNA polymerases. Therefore, tricorn-like domains in *T. b. brucei* could be involved in diverse functions which enhance parasite proliferation.

The study also identified three (3) putative tricorn interacting factor 3 orthologs, two (2) of which were identical. These proteins displayed structural homology with similar residues conservation pattern and motif occurrence with the archaeon protein. However, the low sequence similarity observed could be an indication of existence of sequence divergence among putative tricorn interacting factor 3 orthologs in different organisms. The three ortholog clusters could represent occurrence of tricorn interacting factor 3 in different organisms which may be related to the specific functions performed. The overall structure of the tricorn interacting factor 3 in *Thermoplasma acidophilum* has four domains; N-terminal domain (Domain I) which is composed predominantly of β-sheets mainly composed of hydrophobic amino acids, catalytic domain II composed of mixed α/β- structure, domain III composed of β-sheets and C-terminal domain IV composed of predominantly α-helices and is organized into a super helix (Kyrieleis *et al.,* 2005). This domain composition and organization is remarkably similar to the overall domain structure of putative tricorn interacting factor 3 orthologs in *T. b. brucei.*

The zinc binding ({HEXXH (18X) E} motif, the GAMEN motif, the zinc binding ligand; glutamate and the proton acceptor; tyrosine are located in domain II, a domain with high sequence conservation among the hits. Histidine residues coordinate zinc binding while glutamate participates in enzyme reactions. Moreover, the sequence GAMEN distinguishes M1 aminopeptidases from leukotriene A4 hydrolases in the M1 family which lack it. Structural variability were mainly observed in the barrel-like domain III which had six (6) β-sheets as compared to eight (8) β- sheets in the archaeon protein, the length of the loops where the putative tricorn interacting factor 3 orthologs in *T. b. brucei* seemed to have longer loops as compared to the archaeon tricorn interacting factor 3. These variations could have been due to insertions which led to more secondary structure features in the *T. b. brucei* orthologs but had less significant effect on the overall folding of the protein. However, the α-helices domain IV was the most variable interms of sequence conservation.

The M1 aminopeptidases which include the tricorn interacting factor 3 orthologs cleave neutral amino acids from the N-terminus of oligopeptides ensuring a pool of free amino acids for other metabolic functions. Although amino acid metabolism in trypanosomes is a complex process and varies depending on parasite environment, trypanosomes maintain an amino acid pool of alanine, glutamate and proline as precursors of metabolites and for osmoregulation functions (Darlyuk *et al.,* 2009). Efflux of these amino acids is used to prevent cell swelling upon hypotonic stress (Darlyuk *et al.,* 2009). Proline and alanine are neutral amino acids thus their levels in *T. b. brucei* could be directly dependent on the activity of tricorn interacting factor 3 orthologs. The bloodstream form relies on D-glucose in the mammalian hosts' bloodstream as the energy source which is metabolized to succinate and acetate (Mazet *et al.,* 2013) while the procyclics are dependent on L-proline in the insect vector as the energy source (Mantilla *et al.,* 2017) where L-proline is also metabolized succinate and further to alanine. Thus, the action of tricorn interacting factor 3 orthologs also could be essential to maintaining a pool of proline for the insect stage. The ability of the parasite to invade

the host tissues and cells and also to evade the host's immune system depends on their ability to hydrolyse oligopeptides and proteins (Stijlemans *et al.,* 2015). Indeed, the tricorn interacting factor 3 orthologs in *T. congolense* also identified in this study have been described as immunogenic (Pillay *et al.,* 2013).

The predicted posttranslational modifications of tricorn interacting factor 3 orthologs in *T. b. brucei* were mainly O-N glycosylation and phosphorylation, major molecular mechanisms through which protein function is regulated. Majority of the predicted phosphorylation sites in both proteins were phosphoserine sites. Phosphoserine sites aid in high affinity enzyme binding which may affect various physiological processes including intracellular localization, preference of interacting partners, enzymatic functions through conformational changes or masking of the functional amino acid residues (Inoue *et al.*, 2010). In *Trypanosoma b. brucei*, serine phosphorylation results in loop activation causing the enzyme to adopt an active conformation and promotes correct alignment of catalytic residues (Nett *et al.,* 2009). Therefore, the higher phosphoserine sites in the tricorn interacting factor 3 orthologs could be important in regulation their physiological processes.

The more interactions predicted in the protein-protein interactions network could indicate that the proteins were potentially biologically connected as a group. In *Thermoplasma acidophilum,* a functional coherence of the interacting factors with proteasome through tricorn protease has been demonstrated (Tamura *et al.,* 1998). The putative tricorn interacting factor 3 orthologs in *T. b. brucei* could have a similar functional coherence with the proteasome core complex through interaction with AAZ10244 (Tb927.3.2090), an aminopeptidase P1, synonymous to tricorn interacting factor 1 (F1) in *T. acidophilum* (Goettig *et al.,* 2002) and polyubiquitin which confers specificity to proteasomal protein degradation. The putative tricorn interacting factor 3 orthologs also had predicted interactions with cytosolic leucyl aminopeptidase and aspartyl aminopeptidase. Cytosolic leucyl aminopeptidase has been shown to play

important roles in physiological processes, such as the catabolism of endogenous and exogenous proteins, peptides, protein turnover and processing, modulation of gene expression, antigen processing and defense (Cadavid-Restrepo *et al.,* 2011). The phylogenic relationship of tricorn interacting factor 3 orthologs demonstrated similar conservation pattern in different trypanosome species including *T. cruzi* but with different evolutionary pathways.

The detection of mRNA by reverse transcription in the bloodstream form validated the presence and expression of Tb927.11.3570 and Tb927.3.4750/90 in *T. b. brucei* genome. A nucleotide BLAST analysis of the sequenced products at NCBI database picked Tb927.11.3570, Tb927.3.4750 and Tb927.3.4790 as the best hits with significant E-values further validating their presence. Through qPCR, the expression of tricorn interacting factor 3 orthologs in *T. b. brucei* was significant which could indicate that the proteins are highly abundant in the bloodstream form of the parasite. The duplication of Tb927.3.4750/90 suggests a strategy to increase protein levels further showing their importance in parasite endocytic system. The decline in growth observed in tetracycline induced RNAi mutants points out to their possible involvement in cell growth and differentiation.

Despite the reduced growth, it was observed that decreased expression of each of the tricorn interacting factor 3 ortholog had no observable phenotypic changes in *T. b. brucei* RNAi mutants. This could be due to inability to achieve complete abolishment of the gene function (Balana-Fouce and Reguera 2007). This may not necessarily depict an unimportant gene in viability (Motyka *et al.*, 2004). Therefore, it is possible that decreased expression tricorn interacting factor 3 orthologs affects growth of the parasite hence viability. The parasite growth curve was slow but steady and this could be attributed to the background expression from the repressed tetracycline induced T7 promoters, a common problem in tetracycline promoters (Wang *et al.,* 2000). The steady

growth observed in the mutants could be also as a result of steady emergence of revertants.

5.2 Conclusions

1. Proteins with tricorn-like domains are widely distributed in *Trypanosoma b. brucei* genome and they exist in various combinations with different domain partners indicating possible involvement in diverse functions.
2. *Trypanosoma b. brucei* genome also has three (3) genes coding for tricorn interacting factor 3 orthologs, aminopeptidases belonging to M1 family, enzymes known to cleave neutral amino acids from N-terminal end of peptides.
3. Tricorn interacting factor 3 othologs are significantly expressed in bloodstream form of *T. b. brucei*.
4. Knockdown of genes encoding tricorn interacting factor 3 orthologs in the bloodstream form of *T. b. brucei* slightly reduced growth. Therefore, since knockdown was not absolute, only the parasites which had sufficient knockdown could have been killed or differentiate at slower rates.
5. Tricorn interacting factor 3 othologs could be exerting their effect in a synergistic manner since independent knockdown of the genes had no effect on the morphology of the parasite and little effect on growth and differentiation.
6. The slight growth inhibition observed could have been due to reduced pool of neutral amino acids which play a critical role in osmoregulation and well as being sources of metabolites thus aid in survival of the parasites in the mammalian host.

5.3 Recommendations

The study recommends:

1. Design of molecules that may inhibit the activity of tricorn interacting factor 3 orthologs in *Trypanosoma b. brucei* and further investigate their potential as drug targets against African Trypanosomiasis.
2. Flourescent tagging assays
3. Knockout of Tb927.11.3570, and Tb927.3.4750/90 to completely abolish their functions to effectively determine their role in *Trypanosoma b. brucei* endocytic system.

REFERENCES

Abdulla, M., Brien, T. O., Mackey, Z. B., Sajid, M., Grab, D. J. and Mckerrow, J. H. (2008). RNA Interference of *Trypanosoma brucei* Cathepsin B and L Affects Disease Progression in a Mouse Model. *PLoS Neglected Tropical Diseases, 2*(9), 1 - 6.

Aksoy, S., Buscher, P., Lehane, M., Solani, P. and Abbeele, J. (2017). Human African Trypanosomiasis Control : Achievements and Challenges. *PLoS Neglected Tropical Diseases, 11*(4), 1 - 6.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research, 25*(17), 3389 - 3402.

Anene, B. M., Onah, D. N. and Nawa, Y. (2001). Drug Resistance in Pathogenic African Trypanosomes: What Hopes for the Future? *Veterinary Parasitology, 96*(2), 83 - 100.

Antoine-Moussiaux, N., Philippe, B., and Daniel, D. (2009). Host-Parasite Interactions in Trypanosomiasis: On the Way to an Antidisease Strategy. *Infection and Immunity, 77*(4), 1276 - 1284.

Arnold, K., Lorenza, B., Jürgen K., and Torsten S. (2006). The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics, 22*(2), 195 - 201.

Bailey, T. L., Boden, M. B., Fabian, A. F., Martin, G., Charles, E. C., Luca, R., Jingyuan, L., Wilfred, W. N. and William, S. (2009). MEME Suite: Tools for Motif Discovery and Searching. *Nucleic Acids Research, 37*(SUPPL. 2), 202 - 208.

Balana-Fouce, R. and Reguera, R. M. (2007). RNA interference in *Trypanosoma brucei*: a high-throughtput engine for functional genomics in trypanomastids? *Trends in Parasitology, 23*, 348 - 351

Bangs, J. D., Ransom, D. A., Nimick, M., Christie, G. and Hooper, N. M. (2001). *In Vitro* Cytocidal Effects on *Trypanosoma brucei* and Inhibition of *Leishmania major* GP63 by Peptidomimetic Metalloprotease Inhibitors. *Molecular and Biochemical Parasitology, 114*(1), 111 - 117.

Bastos, I. M. D., Motta, F. N., Grellier, P. and Santana, J. M. (2010). Prolyl Oligopeptidase of *Trypanosoma brucei* Hydrolyzes Native Collagen, Peptide Hormones and Is Active in the Plasma of Infected Mice. *Microbes and Infection, 12*(6), 457.- 466.

Baumeister, W., Jochen, W., Frank, Z. and Erika S. (1998). The Proteasome: Paradigm of a Self-Compartmentalizing Protease. *Cell, 92*(3), 367–380.

Benne, R., Van Den Burg, J., Brakenhoff, P. J., Sloof, P., Van Boom, J. H. and Tromp, M. C. (1986). Major Transcript of the Frameshifted Coxll Gene from Trypanosome Mitochondria Contains Four Nucleotides That Are Not Encoded in the DNA. *Cell, 46*(6), 819 - 826.

Bentivoglio, M. and Krister, K. (2014). Tryps and Trips: Cell Trafficking across the 100-Year-Old Blood-Brain Barrier. *Trends in Neurosciences, 37*(6), 325 - 333.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U., Arrowsmith, C. (2005). The Genome of the African Trypanosome: *Trypanosoma Brucei. Science, 309*(5733), 416 - 422.

Bezie, M., Girma M., Dagnachew S., Tadesse D. and Tadesse G (2014). African Trypanosomes : Virulence Factors, Pathogenicity and Host Responses. *Journal of Veterinary Advances, 4*(11), 732 - 745.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T., Bertoni, M., Bordoli, L. and Schwede, T. (2014). "SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information." *Nucleic Acids Research, 42*(W1), 252 - 258.

Borissenko, L. and Michael G. (2005). Crystal Structure of TET Protease Reveals Complementary Protein Degradation Pathways in Prokaryotes. *Journal of molecular biology, 346*(5), 1207 - 1219.

Bossard, G., Gérard, C., and Anne G. (2013). Secreted Proteases of *Trypanosoma brucei* g*ambiense*: Possible Targets for Sleeping Sickness Control? *BioFactors, 39*(4), 407 - 414.

Brandstetter, H., Kim, J. S., Groll, M. and Huber, R. (2001). Crystal Structure of the Tricorn Protease Reveals a Protein Disassembly Line. *Nature, 414*(6862), 466 - 470.

Brun, R., Schumacher, R., Schmid, C., Kunz, C. and Burri, C. (2001). The Phenomenon of Treatment Failures in Human African Trypanosomiasis. *Tropical Medicine and International health, 6*(11), 906 - 914.

Brun, R., Johannes, B., Francois, C., and Burri, C. (2010). Human African Trypanosomiasis. *Lancet, 375*(10009), 148 - 159.

Bullard, W., Kieft, R., Capewell, P., Veitch, N. J., Macleod, A. and Hajduk, S. L. (2012). Haptoglobin-Hemoglobin Receptor Independent Killing of African Trypanosomes by Human Serum and Trypanosome Lytic Factors. *Virulence, 3*(1), 72 - 76.

Buratai, L.B., Nok, A. J., Ibrahim, S., Umar, I. A. and Esievo, K. (2006). Characterization of Sialidase from Bloodstream Forms of *Trypanosoma vivax*. *Cell Biochemistry and Function, 24*(1), 71 - 77.

Cadavid-Restrepo, G., Gastardelo, T. S., Faudry, E., de Almeida, H., Bastos, I., Negreiros, R. S., Lima, M., Assumpção, T. C., Almeida, K. C., Ragno, M., Ebel, C., Ribeiro, B., Felix, C. R. and Santana, J. (2011). The Major Leucyl Aminopeptidase of *Trypanosoma cruzi* (LAPTc) Assembles into a Homohexamer and Belongs to the M17 Family of Metallopeptidases. *BMC Biochemistry, 12*(1), 46.

Clausen, T and Groll, M. (2003). Molecular shredders: How proteasomes fulfil their role. *Current Opinion in Structural Biology*, *13* (6), 665 - 673

Canning, P., Dean R., Rory, E. M. and Vilmos, F. (2013). Crystal Structures of *Trypanosoma brucei* Oligopeptidase B Broaden the Paradigm of Catalytic Regulation in Prolyl Oligopeptidase Family Enzymes." *PLOS ONE, 8*(11), 1 - 10.

Chappuis, F., Nitya, U., Stietenroth, K., Meussen, A. and Bovier, P. A. (2005). Eflornithine is Safer than Melarsoprol for the Treatment of Second-Stage *Trypanosoma brucei gambiense* Human African Trypanosomiasis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 41*(2),748 - 751.

Chappuis, F., Louis, L., Simarro, P., Büscher, P. and Lejon, V. (2005). Options for Field Diagnosis of Human African Trypanosomiasis. *Clinical Microbiology Reviews, 18*(1): 133 - 146.

Chen, F. (2006). OrthoMCL-DB: Querying a Comprehensive Multi-Species Collection of Ortholog Groups. *Nucleic Acids Research, 34*(90001), D363 - 368.

Cunningham, L. J., Lingley, J. K., Haines, L. R., Ndung'u, J. M., Torr, S. J. and Adams, E. R. (2016). Illuminating the Prevalence of *Trypanosoma brucei* S.l. in *Glossina* Using LAMP as a Tool for Xenomonitoring. *PLoS Neglected Tropical Diseases, 10*(2), 1 - 13.

Dalal, S. and Klemba, M. (2007). Roles for Two Aminopeptidases in Vacuolar Hemoglobin Catabolism in *Plasmodium falciparum*. *Journal of Biological Chemistry, 282*(49), 351008 - 351087.

Delano, W L. (2002). PyMOL Molecular Graphics System.

Drag, M., Matthew, B., Jonathan A. E. and Guy S. S. (2010). Aminopeptidase Fingerprints, an Integrated Approach for Identification of Good Substrates and Optimal Inhibitors. *Journal of Biological Chemistry, 285*(5), 3310 - 3318.

Drinkwater, N., Vinh, N., Mistry, S., Bamert, R., Ruggeri, C., Holleran, J., Loganathan, S., Paiardini, A., Charman, S., Powell, A., Avery, V., McGowan, S. and Scammells, P. (2016). Potent Dual Inhibitors of *Plasmodium falciparum* M1 and M17 Aminopeptidases through Optimization of S1 Pocket Interactions. *European Journal of Medicinal Chemistry, 110*, 43 - 64.

Drinkwater, N., Lee, J., Yang, W., Malcolm, T. and McGowan, S. (2017). M1 aminopeptidases as drug targets: broad applications or therapeutic niche? *FEBS Journal*, *284*(10), 1473 - 1488.

Farr, H. and Keith, G. (2012). Cytokinesis in Trypanosomes. *Cytoskeleton*, *69*(11), 931 - 941.

Ferguson, M. A., and Williams, A. (1988). Cell-Surface Anchoring of Proteins via Glycosyl-Phosphatidyl Inositol Structures. *Annual Reviews in Biochemistry*, *57*, 285 - 320.

Fevre, E. M., Beatrix V. W., Susan C. W., and Lutumba, P., (2008). The Burden of Human African Trypanosomiasis. *PLoS Neglected Tropical Diseases, 2*(12), 12-18.

Fggfdut, B. and Boujqbsbtjujd, P. G. (2011). Therapy. *Mayo Clinic Proceedings*, *86*(6), 561 - 583.

Field, M. C., Horn, D., Fairlamb, A. H., Ferguson, M., Gray, D., Read, K., Rycker, M., Torrie, L. S., Wyatt, P., Wyllie, S. and Gilbert, I. (2017). Anti-Trypanosomatid Drug Discovery : An Ongoing Challenge and a Continuing Need." *Vector-Borne Disease*, *15*, 217 - 232.

Franco, J. R., Simarro, P. P., Diarra, A. and Jannin, J. G. (2014). Epidemiology of Human African Trypanosomiasis. *Clinical epidemiology*, *6*, 257 - 275.

Frasch, P. A., Adriana K. C., Luiz-Juliano, J. C. and Gabriela, T. N. (2012). Characterization of the M32 Metallocarboxypeptidase of *Trypanosoma brucei:* Differences and Similarities with Its Orthologue in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, *182*(2), 63- 70.

Geer, L. Y., Domrachev, M., Lipman, D. J. and Bryant, S. H. (2002). CDART : Protein Homology by Domain Architecture CDART. *Genome Research*, *10*, 1619 - 1623.

Geiger, A., Hirtz, C., Bécue, T., Bellard, E., Centeno, D., Gargani, D., Rossignol, M., Cuny, G. and Peltier, J. (2010). Exocytosis and Protein Secretion in *Trypanosoma*. *BMC Microbiology, 10*, 20-25.

Glas, R., Giulio P., Rainier, K., and Riccardo, G. (2010). The Enigma of Tripeptidyl-Peptidase II: Dual Roles in Housekeeping and Stress. *Journal of Oncology* ,doi:10.1155/2010/128478.

Glover, L., Alsford, S., Baker, N., Turner, D. J., Sanchez-flores, A., Hutchinson, S., Hertz-fowler, C., Berriman, M. and Horn, D. (2015). Genome-Scale RNAi Screens for High-Throughput Phenotyping in Bloodstream-Form African Trypanosomes. *Nature Protocols*, *10*(1), 106 - 133.

Goettig, P., Groll, M., Kim, J. S., Huber, R. and Brandstetter, H. (2002). Structures of the Tricorn-Interacting Aminopeptidase F1 with Different Ligands Explain Its Catalytic Mechanism. *EMBO Journal, 21*(20), 5343 - 5352.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001). Assignment of Homology to Genome Sequences Using a Library of Hidden Markov Models. *Journal of Molecular Biology, 313*(4), 903 - 919.

Gras, S., Byzia, A., Gilbert, F., Mcgowan, S., Drag, M., Silvestre, A., Niepceron, A., Lecaille, F., and Lalmanach, G. (2014). Development. *Eukaryotic cell, 13*(7), 884 - 895.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M. and Bartunik, H. D and Huber, R. (1997). Structure of 20S Proteasome from Yeast at 2.4 A Resolution. *Nature, 386*(6624), 463 - 471.

Groll, M., Bochtler, M., Clausen, T. and and Huber, R. (2005). Molecular Machines for protein degradation. *Biological Chemistry 6*(2), 222 - 256.

Hedstrom, L. (2002). Serine Protease Mechanism and Specificity. *Chemical Reviews, 102*(12), 4501 - 4523.

Holt, H. R., Selby, R., Mumba, C., Napier, G. B. and Guitian, J. (2016). Assessment of Animal African Trypanosomiasis (AAT) Vulnerability in Cattle-Owning Communities of Sub-Saharan Africa. *Parasites and Vectors*, *5*, 1 - 12.

Horn, D. (2014). Antigenic Variation in African Trypanosomes. *Molecular and Biochemical Parasitology, 195*, 123 - 129.

Huang, T., Hwang, J. K., Chen, C. H., Chu, C. S., Lee, C. W. and Chen, C. C. (2015). (PS) 2: Protein Structure Prediction Server Version 3.0. *Nucleic Acids Research, 43*, 338 - 342.

Hung, A. Y., and Morgan, S. (2002). PDZ Domains: Structural Modules for Protein Complex Assembly. *Journal of Biological Chemistry*, *277*(8), 5699 - 5702.

Inoue, M., Yasuda, K., Uemura, H., Yasaka, N., Hiroshi, I., Sei, Y., Horikoshi, N. and Fukuma, T. (2010). Phosphorylation-Dependent Protein Interaction with *Trypanosoma* brucei 14-3-3 Proteins that Display Atypical Target Recognition. *PLOS ONE*, *5*(12), 3 - 10.

Kato, C. D., Matovu, E., Mugasa, C. M. and Nanteza, A. and Alibu, V. P. (2016). The Role of Cytokines in the Pathogenesis and Staging of *Trypanosoma brucei rhodesiense*; Sleeping Sickness. *Allergy Asthma Clinical Immunilogy*, 12(4): 1–10.

Katoh, K. and Daron M. S. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772 - 780.

Kieft, R., Capewell, P., Turner, C. M., Veitch, N. J., MacLeod, A. and Hajduk, S. (2010). Mechanism of *Trypanosoma brucei gambiense*; Resistance to Human Trypanosome Lytic Factor. *Proceedings of the National Academy of Sciences of the United States of America, 107*(37), 16137 - 16141.

Kim, J. S., Groll, M., Musiol, H. J., Behrendt, R., Kaiser, M., Moroder, L., Huber, R. and Brandstetter, H. (2002). Navigation inside a Protease: Substrate Selection and Product Exit in the Tricorn Protease from *Thermoplasma acidophilum*. *Journal of Molecular Biology, 324*(5), 1041 - 1050.

Klemba, M. and Goldberg, D. (2002). Biological Roles of Proteases in Parasitic Protozoa. *Annual Review of Biochemistry, 71*(1), 275 - 305.

Kumar, S. Stecher, G. and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution, *33* (7), 54

Kyrieleis, O. J., Goettig, P., Kiefersauer, R., Huber, R. and Brandstetter, H. (2005). Crystal Structures of the Tricorn Interacting Factor F3 from Thermoplasma Acidophilum, a Zinc Aminopeptidase in Three Different Conformations. *Journal of Molecular Biology, 349*(4), 787 - 800.

Laemmli, U. K. (1970). Laemmli-SDS-PAGE. *Bio-protocol* Bio101: e80. DOI: 10.21769/BioProtoc.80.

Laskowski, R. A., Hutchinson, E.G., Michie, A.D., Wallace, A.C., Jones, M. and Thorntona, J. M. (1997). PDBsum : A Web - Based Database of Summaries and Analyses of All PDB Structures. *Trends in Biochemical Science, 22*(12), 19.

Livak, K.J. and Schmittegen, T. D. (2001). Analysis of relative gene expression data using real time quantitative PCR and the 2 {Delta Delta C (T)} Method. *Methods*, *25*(4), 201.

Li, K., Zhou, S., Guo, Q., Chen, X., Lai, D., Lun, Z. and Guo, X. (2017). The eIF3 Complex of Trypanosoma Brucei : Composition Conservation Does Not Imply the Conservation of Structural Assembly and Subunits Function. *RNA, 23*, 333 - 345.

Lomo, P., Coetzer, T. and Lonsdale-Eccles, J. (1997). Characterization of a Multicatalytic Proteinase Complex (20S Proteasome) from *Trypanosoma brucei brucei*. *Immunopharmacology, 36*: 285 - 293.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C. and Bryant, S. H. (2015). CDD: NCBI's Conserved Domain Database. *Nucleic Acids Research, 43*(D1), D222 - 226.

McGowan, S., Porter, C. J., Lowther, J., Stack, C., Golding, S. J., Skinner-Adams, T. S.,

Trenholme, K. R., Teuscher, F., Donnelly, S. M., Grembecka, J., Mucha, A., Kafarski, P., Degori, R., Buckle, A. M., Gardiner, D. L., Whisstock, J. C. and Dalton, J. P. (2009). Structural Basis for the Inhibition of the Essential *Plasmodium Falciparum* M1 Neutral Aminopeptidase. *Proceedings of the National Academy of Sciences of the United States of America, 106*(8), 2537 - 2542.

Mazet, M., Morand, P., Biran, M., Bouyssou, G., Courtois, P., Daulou, S., Millerioux, Y., Franconi, J. M., Vincendeau, P., Moreau, P. and Bringaud, F. (2013). *PLoS Neglected Tropical Diseases, 7*(12), 1 - 14.

Mantilla, B. S., Marchese, L., Casas-Sánchez, A., Dyer, N. A., Ejeh, N., Biran, M., Bringaud, F., Lehane, M. J., Acosta-Serrano, A. and Silber, A.M. (2017). PLoS Pathogens, *13*(1), 1 - 29.

Melville, S. E., Vanessa, L., Miguel, N. and Cross, G. (2000). The Molecular Karyotype of the Megabase Chromosomes of *Trypanosoma brucei. Molecular and Biochemical Parasitology, 111*(2), 261 - 273.

Meyer, A., Holt, H. R., Selby, R. and Javier, G. (2016). Past and Ongoing Tsetse and Animal Trypanosomiasis Control Operations in Five African Countries: A Systematic Review. *PLoS Neglected Tropical Diseases, 10*(12), 1 - 29.

Molina-Portela, M., Elena, B. L., Esperanza, R. and Jayne, R. (2005). Trypanosome Lytic Factor, a Subclass of High-Density Lipoprotein, Forms Cation-Selective Pores in Membranes. *Molecular and Biochemical Parasitology, 144*(2), 218 - 226.

Moody, A. H. and Chiodini, P. (2000). Methods for the Detection of Blood Parasites. *Ckin Lab Haem, 22*, 189 - 202.

Morrison, L. J. (2011). Parasite-Driven Pathogenesis in *Trypanosoma brucei* Infections. *Parasite Immunology*, 33(8): 448–455.

Morty, R. E., Lonsdale-Eccles, J. D., Mentele, R., Auerswald, E. A. and Coetzer, T. H. (2001). Trypanosome-Derived Oligopeptidase B Is Released into the Plasma of Infected Rodents, Where It Persists and Retains Full Catalytic Activity. *Infection and Immunity, 69*(4), 2757 - 2761.

Morty, R. E., Vadász, I., Bulau, P., Dive, V., Oliveira, V., Seeger, W. and Juliano, L. E. (2005). Tropolysin, a New Oligopeptidase from African Trypanosomes. *Biochemistry, 44*(44), 14658 - 14669.

Motyka, S., Zhao, Z., Gull, K. and Englund, P. (2004). Integration of pZJM library plasmids into unexpected locations in the *Trypanosoma brucei* genome. *Molecular & Biochemical Parasitology, 134*, 163 - 169.

Mucha, A., Drag, M., Dalton, J. P. and Kafarski, P. (2010). Metallo-aminopeptidase inhibitors. Biochemie, *92*(11), 1509 - 1529.

Muñoz, C., Juan, F., Bessy, G. and Jorge, G. (2015). Role of the Ubiquitin-Proteasome Systems in the Biology and Virulence of Protozoan Parasites. *BioMed Research International, 15*:1 - 13.

Murilla, G., Ndung'u, K., Thuita, J., Gitonga, P., Kahiga, D., Auma, J., Ouma, J., Rutto, J. and Ndung'u, J. (2014). Kenya Trypanosomiasis Research Institute Cryobank for Human and Animal Trypanosome Isolates to Support Research: Opportunities and Challenges. *PLoS Neglected Tropical Diseases, 8*(5), 22 - 28.

Mutomba, M. C., Wah-Yuen, T., William, C. H. and Wang, C. (1997). Inhibition of Proteasome Activity Blocks Cell Cycle Progression at Specific Phase Boundaries in African Trypanosomes. *Molecular and Biochemical Parasitology, 90*(2), 491 - 504.

Mutomba, M. C. and Wang, C. (1998). The Role of Proteolysis during Differentiation of *Trypanosoma brucei* from the Bloodstream to the Procyclic Form. *Molecular and Biochemical Parasitology, 93*(1), 11 - 22.

Neil, D. (1993). Evolutionary Families of Peptidases. *Complement, 218*: 205 - 218.

Nett, I. R .E, Martin, D., Miranda-Saavedra, D., Lamont, D., Barber, J. D., Mehlert, A., and Ferguson, M. (2009). *Molecular & cellular proteomics, 8*(7), 1527 - 1538.

Nocek, B., Mulligan, R., Bargassa, M., Collart, F. and Joachimiak, A. (2008). Crystal Structure of Aminopeptidase N from Human Pathogen *Neisseria meningitides. Proteins, 70*(1), 273 - 279.

O'Brien, T., Mackey, Z., Fetter, R., Choe, Y., O'Donoghue, A., Zhou, M., Craik, C., Caffrey, C. and McKerrow, J. (2008). A Parasite Cysteine Protease Is Key to Host Protein Degradation and Iron Acquisition. *Journal of Biological Chemistry, 283*(43), 28934 - 28943.

Oates, M., Stahlhacke, J., Vavoulis, D., Smithers, B., Rackham, O., Sardar, A., Zaucha, J., Thurlby, N., Fang, H. and Gough, J. (2015). The SUPERFAMILY 1.75 Database: A Doubling of Data. *Nucleic Acids Research, 43*(D1), D227 - 233.

Pallen, M. J., Lam, A. and Loman, N. (2001). Tricorn-like Proteases in Bacteria. *Trends in Microbiology, 9*(11), 518 - 521.

Paugam, A., Bulteau, A., Dupouy-Camet, J., Creuzet, C. and Friguet, B. (2003). Characterization and Role of Protozoan Parasite Proteasomes. *Trends in Parasitology, 19*(2), 55 - 59.

Peacock, L., Ferris, V., Sharma, R., Sunter, J., Bailey, M., Carrington, M. and Gibson, W. (2011). Identification of the Meiotic Life Cycle Stage of *Trypanosoma brucei* in the Tsetse Fl Y. *PNAS, 108*(9), 3671 - 3676.

Peña-Diaz, P., Vancová, M., Resl, C., Field, M. C. and Lukeš, J. (2017). A Leucine Aminopeptidase Is Involved in Kinetoplast DNA Segregation in *Trypanosoma brucei*. *PLoS Pathogens, 13*(4), 1 - 23.

Pillay, D., Boulangé, Al., Coustou, V., Baltz, T. and Coetzer, T. (2013). Recombinant expression and biochemical characterisation of two alanyl aminopeptidases of *Trypanosoma congolense*. *Experimental Parasitology*, *135*(4), 675 - 684

Ponting, C. P. and Pallen, M. J. (1999). Beta-Propeller Repeats and a PDZ Domain in the Tricorn Protease: Predicted Self-Compartmentalisation and C-Terminal Polypeptide-Binding Strategies of Substrate Selection. *FEMS microbiology letters*, *179*(2), 447 - 451.

Ranganathan, S. and Gagan, G. (2009). Secretome: Clues into Pathogen Infection and Clinical Applications. *Genome medicine*, *1*(11), 113.

Reits, E., Neijssen, J., Herberts, C., Benckhuijsen, W., Janssen, L., Drijfhout, W. and Neefjes, J. (2004). A Major Role for TPPII in Trimming Proteasomal Degradation Products for MHC Class I Antigen Presentation. *Immunity*, *20*(4), 495 - 506.

Rockel, B. and Baumeister, W. (2008). A Tale of Two Giant Proteases. *Ernst Schering Foundation Symposium Proceedings*, *1*(1), 17 - 40.

Rockel, B., Peters, È., Ku, B., Glaeser, R. M. and Baumeister, W. (2002). A Giant Protease with a Twist : The TPP II Complex from Drosophila Studied by Electron Microscopy. *The EMBO Journal*, *21*(22): 51009 - 51084.

Rotureau, B., I. Subota, J. Buisson, and P. Bastin. (2012). A New Asymmetric Division Contributes to the Continuous Production of Infective Trypanosomes in the Tsetse Fly. *Development*, *139*(10), 1842 - 1850.

Rudenko, G. (2012). Maintaining the Protective Variant Surface Glycoprotein Coat of African Trypanosomes; The Structure of the Protective Variant Surface Glycoprotein Coat. *Biochemical Society Transactions*, *33*(5): 981 - 982.

Rutto, J., Osano, O., Thuranira, E., Kurgat, R., Odenyo, V. (2013). Socio-Economic and Cultural Determinants of Human African Trypanosomiasis at the Kenya - Uganda Transboundary. *PLoS Neglected Tropical Diseases*, *7*(4), 5 - 8.

Shaw, L. N. (2014). The Lone S41 Family C-Terminal Processing Protease in *Staphylococcus aureus* is localized to the cell wall and contributes to virulence. *Microbiology*, *160*: 1737 - 1748.

Simarro, P., Diarra, A., Postigo, J., Franco, J., and Jannin, J. (2011). The Human African Trypanosomiasis Control and Surveillance Programme of the World Health Organization 2000-2009: The Way Forward. *PLoS Neglected Tropical Diseases*, *5*(2), 1 - 7.

Simarro, P., Cecchi, G., Franco, R., Paone, M., Diarra, A., Ruiz-postigo, A., Fe, E., Mattioli, R. and Jannin, G. (2012). Estimating and Mapping the Population at Risk of Sleeping Sickness. *PLoS Neglected Tropical Diseases*, *6*(10), 1 – 12.

Simo, G., Mbida, J., Eyenga, V., Asonganyi, T., Njiokou, F. and Grébaut, P. (2014).

Challenges towards the Elimination of Human African Trypanosomiasis in the Sleeping Sickness Focus of Campo in Southern Cameroon. *Parasites and Vectors, 7*(1), 374.

Skinner-Adams, T., Stack, C., Trenholme, K., Brown, C., Grembecka, J., Lowther, J., Mucha, A., Drag, M., Kafarski, P., McGowan, S., Whisstock, J., Gardiner, D. and Dalton, P. (2010). *Plasmodium falciparum* Neutral Aminopeptidases: New Targets for Anti-Malarials." *Trends in Biochemical Sciences, 35*(1), 53 - 61.

Smith, T. K. and Bütikofer, P. (2010). Europe PMC Funders Group Lipid Metabolism in *Trypanosoma brucei*. *Molecular Biochemistry and Parasitoly, 172*(2), 66 - 79.

Steverding, D. (2008). The History of African Trypanosomiasis. *Parasites and vectors, 1*(1), 3.

Steverding, D. (2010). The Development of Drugs for Treatment of Sleeping Sickness: A Historical Review. *Parasites and vectors, 3*(1), 15.

Stijlemans, B., Beschin, A., Magez, S., Van Ginderachter, A. and De Baetselier, P. (2015). Iron Homeostasis and Trypanosoma brucei Associated Immunopathogenicity Development: A Battle/Quest for Iron. *BioMed Research International,* doi:*10.1155/2015/819389.*

Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gürtler, R., McKerrow, J., Reed, S. and Tarleton, R. (2008). Kinetoplastids: Related Protozoan Pathogens, Different Diseases. *Journal of Clinical Investigation, 118*(4), 1301 - 1310.

Sutton, R. E. and Boothroyd, J. C. (1986). Evidence for Trans Splicing in Trypanosomes. *Cell, 47*(4), 527 - 535.

Sykes, J. E., Stephen, C. B., Ashley, B. S. and Sykes, J. E. (2014). Trypanosomiasis. *Canine and Feline Infectious Diseases*, 760 - 770.

Szecsi, P. B. (1992). The Aspartic Proteases. *Scandinavian Journal of Clinical and Laboratory Investigation. Supplementum 52*(Suppl. 210): 5 - 22.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L. and Mering, C. (2010). The STRING Database; Functional Interaction Networks of Proteins, Globally Integrated and Scored. *Nucleic Acids Research, 41*(D1), D284 - D289.

Tamura, N., Pfeifer, G., Baumeister, W. and Tamura, T. (2001). Tricorn Protease in Bacteria: Characterization of the Enzyme from *Streptomyces coelicolor*. *Biological Chemistry*, 382(3): 449–458.

Tamura, N., Lottspeich, F., Baumeister, W. and Tamura, T. (1998). The Role of Tricorn Protease and Its Aminopeptidase-Interacting Factors in Cellular Protein Degradation. *Cell, 95*(5), 637 - 648.

Tamura, T., Tamura, N., Cejka, Z., Hegerl, R., Lottspeich, F. and Baumeister, W.

(1996). Tricorn Protease--the Core of a Modular Proteolytic System. *Science, 274*(5291), 1385 - 1389.

Timm, J., Valente, M., García-Caballero, D., Wilson, K. and González-Pacanowska, D. (2017). Structural Characterization of Acidic M17 Leucine Aminopeptidases from the TriTryps and Evaluation of Their Role in Nutrient Starvation in *Trypanosoma brucei*. *American Society for Microbiology, 2*(4), 217-226.

Tomkinson, B. (1999). Tripeptidyl Peptidases: Enzymes That Count. *Trends in Biochemical Sciences, 24*(9), 355 - 359.

Troeberg, L., Pike, R. N., Morty, R. E., Berry, R. K., Coetzer, T. H. and Lonsdale-Eccles, J. D. (1996). Proteases from *Trypanosoma brucei brucei*. *FEBS, 238*(3), 728 - 736.

Troeberg, L., Morty, R. E., Pike, R. N., Lonsdale-Eccles, J. D., Palmer, J. T., McKerrow, J. H. and Coetzer, T. (1999). Cysteine Proteinase Inhibitors Kill Cultured Bloodstream Forms of *Trypanosoma brucei brucei*. *Experimental Parasitology, 91*(4), 349 - 355.

Uilenberg, G. (1998). A Field Guide for the Diagnosis, Treatment and Prevention of African Trypanosomiasis. *FAO*.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J. (2007). Primer3Plus, an Enhanced Web Interface to Primer3. *Nucleic Acids Research, 35*(SUPPL.2), 71–74.

Vaidya, T., Bakhiet, M., Hill, K., Olsson, T., Kristensson, K. and Donelson, J. (1997). The gene for a T lymphocyte triggering factor from African Trypanosomes. *The Journal of experimental medicine, 186*(3), 433 - 438.

Vincendeau, P., Okomo-Assoumou, M. C., Semballa, S., Fouquet, C. and Daulouede, S. (1996). Immunology and Immunopathology of African Trypanosomiasis. *Med Trop, 56*(1), 73 - 78.

Vreysen, M., Baba, S. and Bouyer, J. (2012). Tsetse Flies : Their Biology and Control Using Area-Wide Integrated Pest Management Approaches. *Journal of Invertbrate Pathology, 112*, 15 - 25.

Walz, J., Tamura, T., Tamura, N., Grimm, R., Baumeister, W. and Koster, A. J. (1997). Tricorn Protease Exists as an Icosahedral Supermolecule *in vivo*. *Molecular Cell, 1*(1), 59 - 65.

Wang, C., Bozdech, Z., Liu, C., Shipway, A., Backes, B., Harris, J. and Bogyo, M., (2003). Biochemical Analysis of the 20 S Proteasome of *Trypanosoma brucei*. *The Journal of Biological Chemistry, 278*(18), 15800–15808.

Wang, Z., Morris, J., Drew, M. and Englund, P. (2000). Inhibition of *Trypanosoma brucei* Gene Expression by RNA Interference Using an Integratable Vector with Opposing T7 Promoters. *The Journal of biological chemistry, 275* (51), 40174 - 40179.

Wastling, S. L. and Welburn, S. C. (2011). Diagnosis of Human Sleeping Sickness: Sense and Sensitivity. *Trends in Parasitology, 27*(9), 394 - 402.

Węglarz-Tomczak, E. (2013). An Integrated Approach to the Ligand Binding Specificity of *Neisseria meningitidis* M1 Alanine Aminopeptidase by Fluorogenic Substrate Profiling, Inhibitory Studies and Molecular Modeling. *Biochimie, 95*(2), 419 - 428.

WHO (2016). Centers for Disease Control and Prevention, Human African Trypanosomiasis Annual Report (http://www.who.int/neglected_diseases/news)

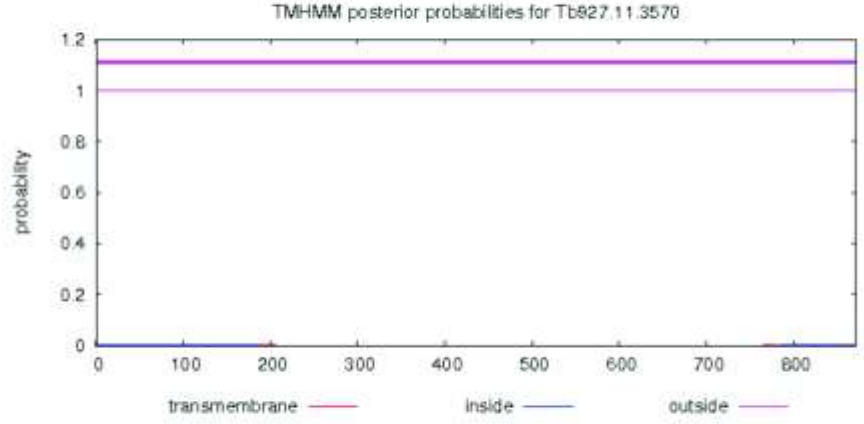
Yao, T. and Cohen, R. (1999). Giant Proteases: Beyond the Proteasome. *Current Biology, 9*(15), 551 - 553.

Yaro, M., Stear, M. and Groth, D. (2016). Combatting African Animal Trypanosomiasis (AAT) in Livestock: The Potential Role of Trypanotolerance. *Veterinary Parasitology, 225*: 43 - 52.

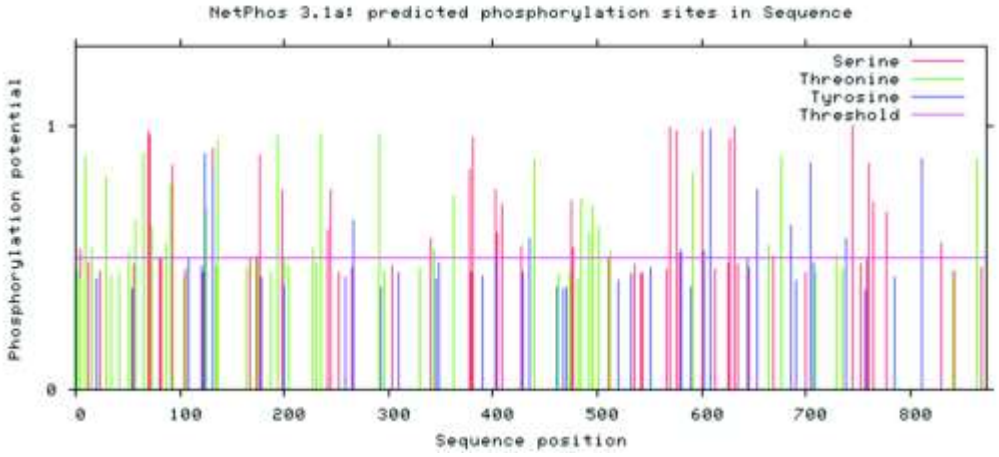
APPENDICES

a

```
# Tb927.11.3570 Length: 871
# Tb927.11.3570 Number of predicted TMHs:  0
# Tb927.11.3570 Exp number of AAs in TMHs: 0.00935
# Tb927.11.3570 Exp number, first 60 AAs:  0
# Tb927.11.3570 Total prob of N-in:        0.00029
Tb927.11.3570   TMHMM2.0        outside      1   871
```

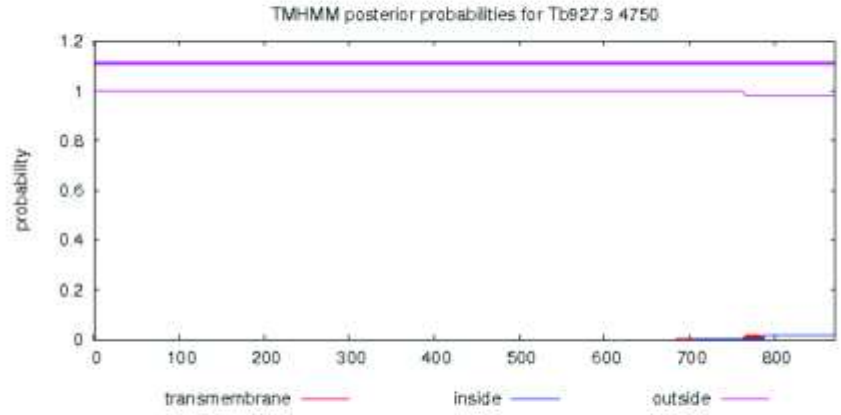


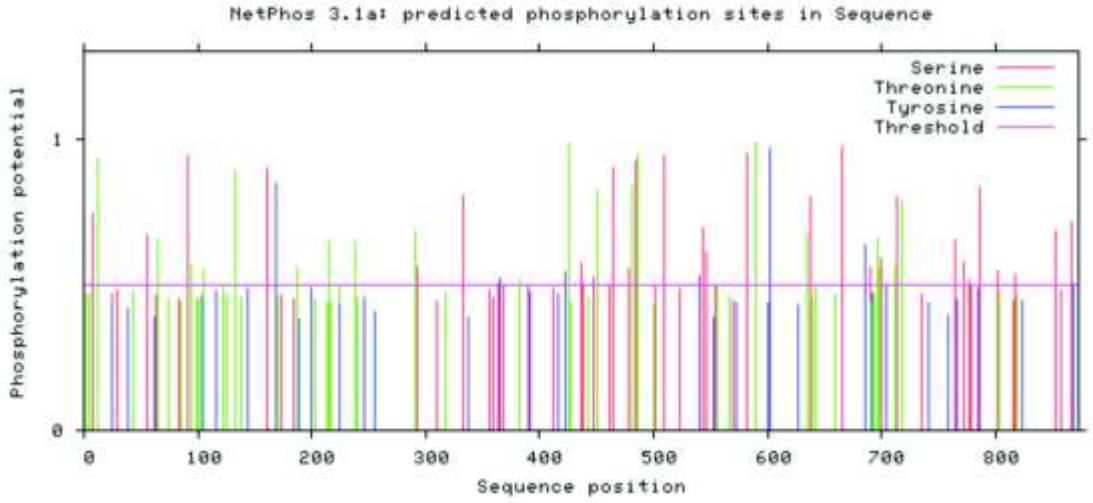
b



Appendix I: Prediction of transmembrane domains and phosphorylation sites for Tb927.11.3570.

```
# Tb927.3.4750 Length: 871
# Tb927.3.4750 Number of predicted TMHs:  0
# Tb927.3.4750 Exp number of AAs in TMHs: 0.39683
# Tb927.3.4750 Exp number, first 60 AAs:  0.00048
# Tb927.3.4750 Total prob of N-in:        0.00006
Tb927.3.4750    TMHMM2.0        outside      1   871
```





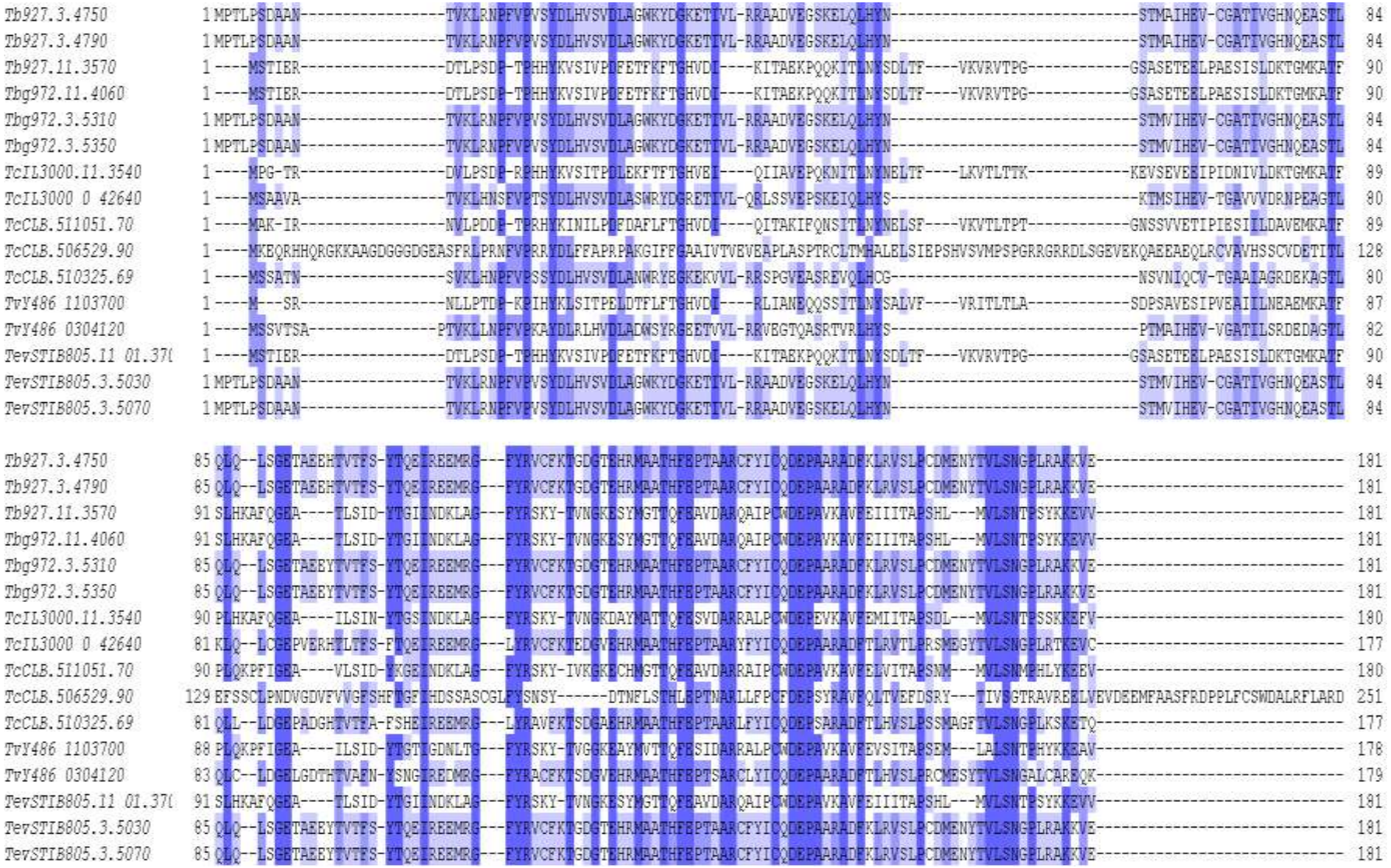
Appendix II: Prediction of transmembrane domains and phosphorylation sites for Tb927.3.4750

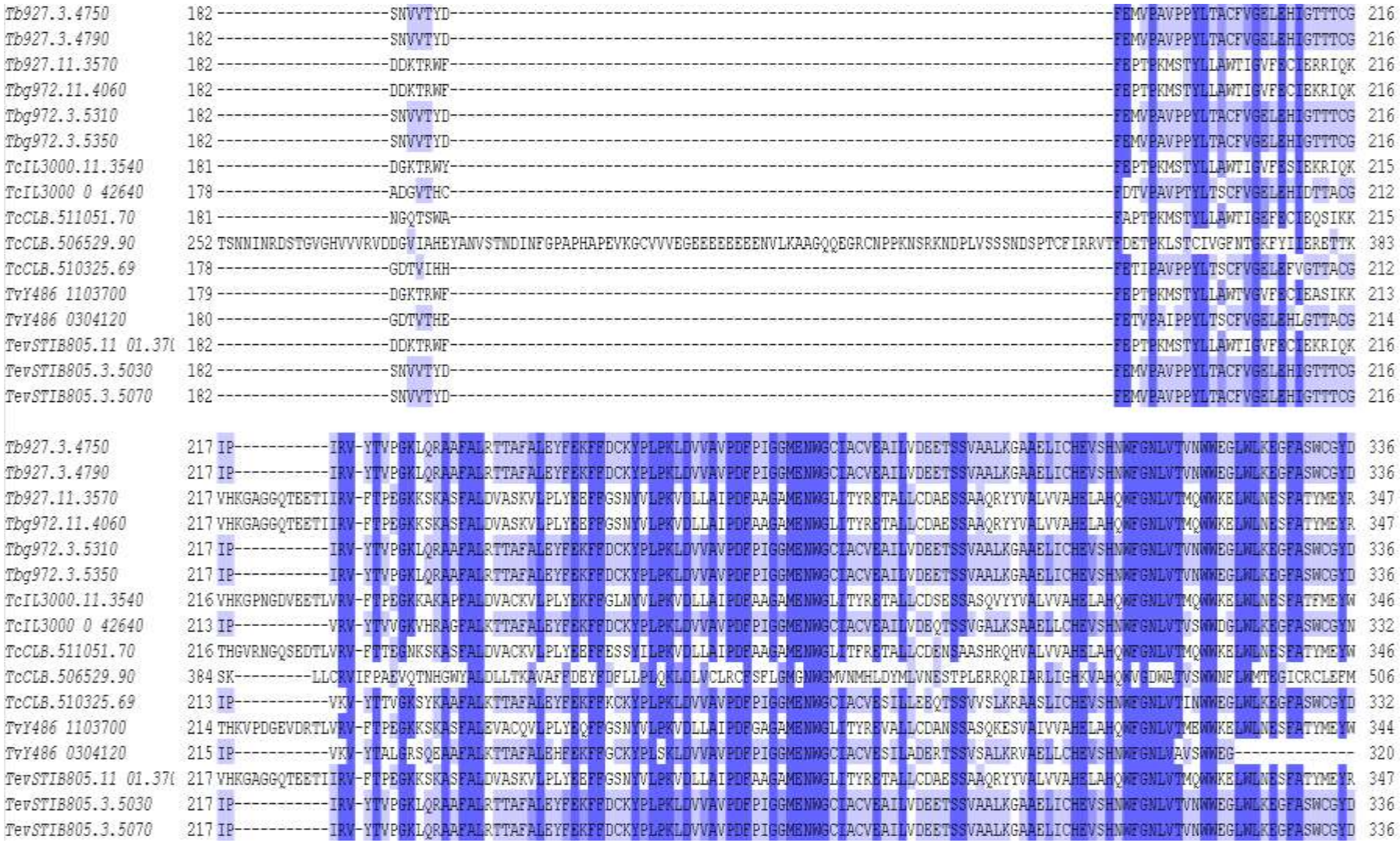
Appendix III: Predicted phosphorylation sites in tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei*

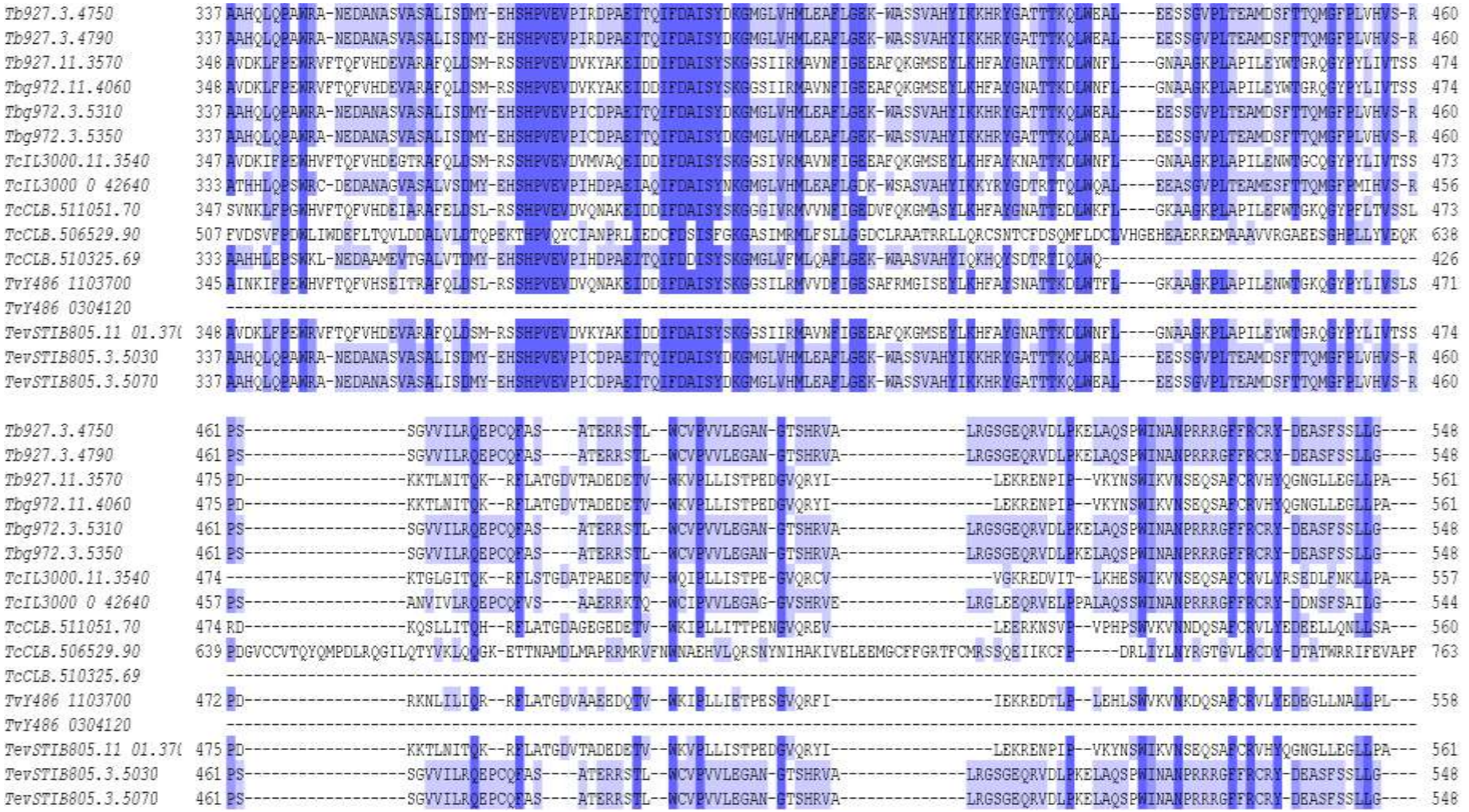
Tb927.3.4750/90			**Tb927.11.3570**		
Kinase type	**Site**	**Score**	**Kinase type**	**Site**	**Score**
PKC	Ser^{6}	0.747	CK II	Ser^{2}	0.558
PKC	Thr^{11}	0.835	Unsp	Thr^{8}	0.880
Unsp	Ser^{53}	0.672	cdk 5	Thr^{14}	0.536
PKC	Thr^{63}	0.651	PKC	Thr^{28}	0.808
CK I	Ser^{89}	0.598	PKG	Thr^{49}	0.516
CK II	Thr^{92}	0.564	PKC	Thr^{56}	0.638
ATM kinase	Thr^{103}	0.554	Unsp	Thr^{63}	0.893
PKC	Thr^{131}	0.621	Unsp	Ser^{67}	0.982
PKA	Ser^{159}	0.750	Unsp	Ser^{69}	0.962
Unsp	Tyr^{167}	0.849	CK II	Thr^{71}	0.614
CK II	Thr^{186}	0.559	PKC	Thr^{84}	0.556
PKC	Thr^{213}	0.650	PKC	Thr^{89}	0.774
PKC	Thr^{237}	0.645	PKC	Ser^{91}	0.851
Unsp	Thr^{289}	0.686	Unsp	Tyr^{122}	0.896
PKC	Ser^{291}	0.546	PKC	Thr^{123}	0.686
Unsp	Ser^{331}	0.805	PKA	Ser^{129}	0.505
CK II	Ser^{361}	0.507	Unsp	Tyr^{130}	0.845
INSR	Tyr^{364}	0.506	CK II	Thr^{134}	0.503
ATM	Thr^{381}	0.517	cdc 2	Ser^{165}	0.501
EGFR	Tyr^{420}	0.539	PKG	Ser^{171}	0.505
PKC	Thr^{423}	0.790	PKC	Ser^{175}	0.846
Unsp	Thr^{424}	0.1008	cdk 5	Thr^{192}	0.515
CK I	Ser^{435}	0.572	PKA	Ser^{196}	0.760
CK II	Ser^{436}	0.504	CK I	Thr^{234}	0.544
CK I	Ser^{446}	0.521	PKG	Ser^{240}	0.602
Unsp	Thr^{449}	0.825	PKA	Ser^{243}	0.759
PKA	Ser^{462}	0.654	Unsp	Tyr^{264}	0.638
Unsp	Ser^{463}	0.849	PKC	Thr^{289}	0.547
PKG	Ser^{477}	0.552	CK I	Ser^{339}	0.576
PKC	Thr^{479}	0.841	CK II	Thr^{342}	0.531
Unsp	Ser^{483}	0.926	Unsp	Thr^{360}	0.733
PKA	Thr^{484}	0.841	PKC	Ser^{375}	0.830
PKA	Ser^{507}	0.579	PKA	Ser^{379}	0.524
EGFR	Tyr^{538}	0.528	cdc 2	Ser^{400}	0.520
Unsp	Ser^{542}	0.694	Unsp	Ser^{406}	0.702
cdc 2	Ser^{544}	0.556	Unsp	Ser^{425}	0.543
CK I	Ser^{545}	0.612	EGFR	Tyr^{433}	0.572
PKC	Thr^{554}	0.500	PKC	Thr^{437}	0.641
PKA	Ser^{580}	0.559	Unsp	Ser^{473}	0.713
Unsp	Thr^{587}	0.983	PKC	Ser^{474}	0.535
Unsp	Tyr^{600}	0.966	PKC	Thr^{483}	0.720
PKC	Thr^{632}	0.675	PKC	Thr^{490}	0.544

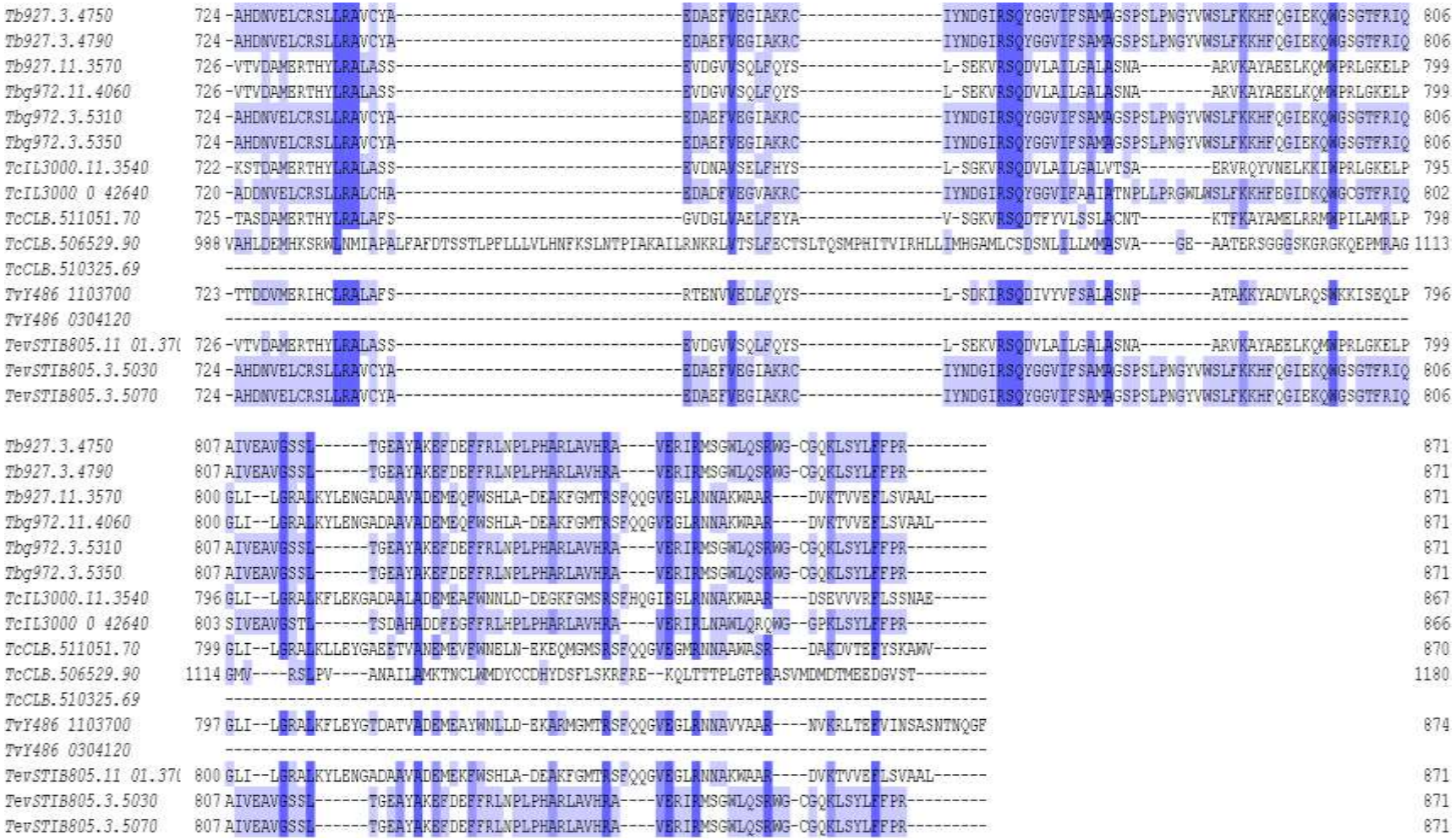
Tb927.3.4750/90			Tb927.11.3750		
Kinase type	**Site**	**Score**	**Kinase Type**	**Site**	**Score**
Unsp	Ser^{636}	0.801	CK II	Thr^{494}	0.699
CK II	Thr^{641}	0.507	PKC	Thr^{500}	0.538
PKG	Ser^{664}	0.510	cdk 5	Thr^{510}	0. 507
Unsp	Tyr^{684}	0.634	Unsp	Ser^{568}	0.993
cdc 2	Ser^{688}	0.564	Unsp	Ser^{574}	0.982
Unsp	Thr^{694}	0.662	CK II	Ser^{577}	0.517
cdc 2	Thr^{696}	0.555	EGFR	Tyr^{579}	0.527
DNAPK	Ser^{698}	0.592	Unsp	Ser^{589}	0.615
Unsp	Thr^{710}	0.559	PKC	Thr^{590}	0.822
Unsp	Ser^{711}	0.801	CK II	Ser^{598}	0.561
Unsp	Thr^{716}	0.788	Unsp	Ser^{599}	0.982
DNAPK	Ser^{762}	0.650	Unsp	Tyr^{600}	0.521
cdc 2	Ser^{770}	0.580	Unsp	Tyr^{606}	0.985
p38MAPK	Ser^{775}	0.515	Unsp	Ser^{625}	0.950
PKC	Ser^{785}	0.829	Unsp	Ser^{629}	0.990
PKA	Ser^{800}	0.547	Unsp	Tyr^{651}	0.759
CK I	Ser^{815}	0.534	Unsp	Thr^{662}	0.551
PKA	Ser^{851}	0.684	cdc 2	Ser^{666}	0.504
PKC	Ser^{865}	0.717	Unsp	Thr^{674}	0.884
Unsp	Tyr^{866}	0.507	CK II	Ser^{675}	0.641
			Unsp	Tyr^{683}	0.624
			Unsp	Tyr^{702}	0.857
			PKC	Thr^{703}	0.536
			CK II	Thr^{727}	0.514
			Unsp	Tyr^{736}	0.575
			Unsp	Ser^{742}	0.601
			PKC	Ser^{758}	0.858
			Unsp	Ser^{763}	0.709
			PKC	Ser^{775}	0.669
			CK II	Ser^{809}	0.553
			PKC	Thr^{861}	0.591

Phosphorylation sites in Tb927.3.4750/90 and Tb927.11.3570 sequences predicted by NetPhos 3.1 server. The numbers represented as superscripts show the position of the amino acid within the sequence. Key: CK: Creatine kinase, PK: Protein kinase, EGFR: Epidermal Growth Factor Receptor, DNAPK: DNA dependent protein kinase, cdc: Cell division cycle, INSR: Insulin receptor precursor, p38MAPK: p38-Mitogen activation protein kinase, GSK: Glycogen synthase kinase, Unsp: Unspecific, ATM kinase: Mutated ataxia-telangiectasia kinase, cdk: Cyclin dependent kinase.

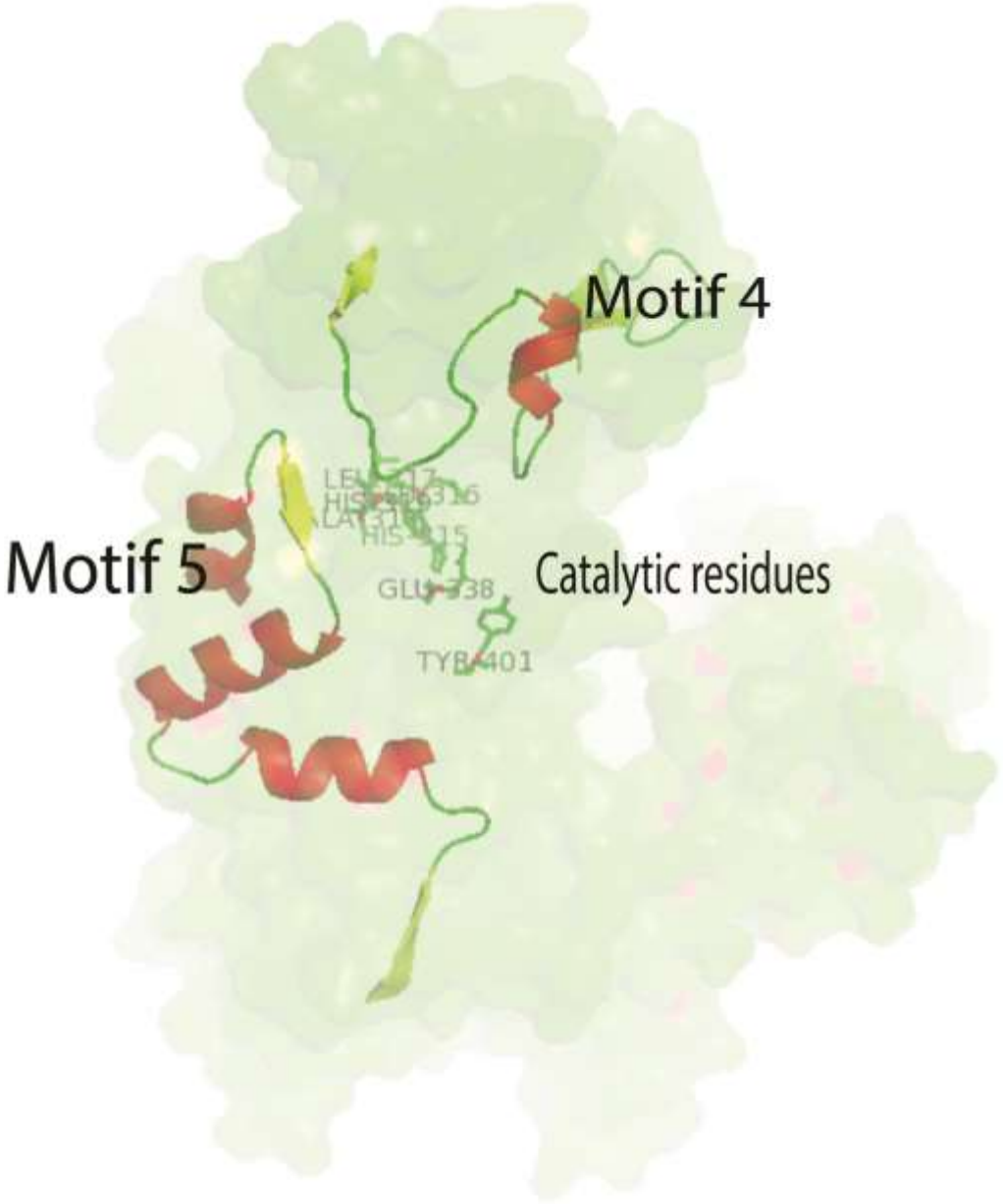





Tb927.3.4750
Tb927.3.4790
Tb927.11.3570
Tbg972.11.4060
Tbg972.3.5310
Tbg972.3.5350
TcIL3000.11.3540
TcIL3000 0 42640
TcCLB.511051.70
TcCLB.506529.90
TcCLB.510325.69
TvY486 1103700
TvY486 0304120
TevSTIB805.11 01.37(
TevSTIB805.3.5030
TevSTIB805.3.5070



Appendix IV: Multiple sequence alignment of tricorn interacting factor 3 orthologs in trypanosomes. The peptides sequences were retrieved from TriTrypsDB. The alignment was generated by MAFFTv7 and was coloured by % identity.



Appendix V: Surface representation of *Trypanosoma brucei brucei* F3 homolog with motifs four (4) and five (5) shown in cartoon representation. The catalytic residues are shown as sticks.

Appendix VI: Pairwise alignment of the N-terminal region tricorn interacting factor 3 orthologs in *Trypanosoma brucei brucei*

```
Query  8    TLPSDP----------TPHHYKVSIVPDFETFKFTGHVDI---KITAEKPQQKITLNYSD  54
            TLPSD            P  Y + +  D   +K+ G   I   +    +  +++ L+Y+
Sbjct  3    TLPSDAANTVKLRNPFVPVSYDLHVSVDLAGWKYDGKETIVLRRAADVEGSKELQLHYNS  62

Query  55   LTFVKVRVTPGGSASETEELPAESISLDKTGMKATFSLHKAFQGEATLSIDYTGIINDKL  114
             T     V         +E  A ++ L  +G  A          E T++  YT  I +++
Sbjct  63   -TMAIHEVCGATIVGHNQE--ASTLQLQLSGETAE---------EHTVTFSYTQEIREEM  110

Query  115  AGFYRSKY-TVNGKESYMGTTQFEAVDARQAIPCWDEPAVKAVFEIIITAPSHL---MVL  170
             GFYR  + T +G E  M  T FE   AR    C DEPA +A F++ ++ P  +    VL
Sbjct  111  RGFYRVCFKTGDGTEHRMAATHFEPTAARCFYICQDEPAARADFKLRVSLPCDMENYTVL  170

Query  171  SNTPSYKKEVVDDKTRWFFEPTPKMSTYLLAWTIGVFECIERRIQKVHKGAGGQTEETII  230
            SN P   K+V  +   + FE  P +  YL A  +G  E I           G  T    I
Sbjct  171  SNGPLRAKKVESNVVTYDFEMVPAVPPYLTACFVGELEHI-----------GTTTCGIPI  219

Query  231  RVFTPEGKKSKASFALDVASKVLPLYEEFFGSNYVLPKVDLLAIPDFAAGAMENWGLITY  290
            RV+T  GK  +A+FAL   +  L  +E+FF   Y LPK+D++A+PDF  G MENWG I
Sbjct  220  RVYTVPGKLQRAAFALRTTAFALEYFEKFFDCKYPLPKLDVVAVPDFPIGGMENWGCIAC  279

Query  291  RETALLCDAESSAAQRYYVALVVAHELAHQWFGNLVTMQWWKELWLNESFATYMEYRAVD  350
             E  L+ +  SS A     A ++ HE++H WFGNLVT+ WW+ LWL E FA++  Y A
Sbjct  280  VEAILVDEETSSVAALKGAAELICHEVSHNWFGNLVTVNWWEGLWLKEGFASWCGYDAAH  339

Query  351  KLFPEWRVFTQFVHDEVARAFQLDSMRSSHPVEVDVKYAKEIDDIFDAISYSKGGSIIRM  410
            +L P WR   +  +  VA A   D    SHPVEV ++   EI  IFDAISY KG  ++ M
Sbjct  340  QLQPAWRA-NEDANASVASALISDMYEHSHPVEVPIRDPAEITQIFDAISYDKGMGLVHM  398

Query  411  AVNFIGEEAFQKGMSEYLKHFAYGNATTKDLWNFLGNAAGKPLAPILEYWTGRQGYPYLI  470
               F+GE+ +   ++ Y+K   YG  TTK LW  L  ++G PL   ++ +T + G+P L+
Sbjct  399  LEAFLGEK-WASSVAHYIKKHRYGATTTKQLWEALEESSGVPLTEAMDSFTTQMGFP-LV  456

Query  471  VTSSPDKKTLNITQK--RFLATGDVTADEDETVWKVPLLISTPEDGVQRYIL----EKRE  524
              S P    + + Q+  +F +      +   T+W VP+++        R  L    E+R
Sbjct  457  HVSRPSSGVVILRQEPCQFAS----ATERRSTLWCVPVVLEGANGTSHRVALRGSGEQRV  512

Query  525  N-PIPVKYNSWIKVNSEQSAFCRVHYQGNGLLEGLLPAIASKNLSDIDRFSIISDYHAFA  583
            + P  +  + WI  N  +  F R  Y        LL A   K+L+  DR  +I+D  A
Sbjct  513  DLPKELAQSPWINANPRRRGFFRCRYD-EASFSSLLGAY--KSLTIPDRCGLIADTLASV  569

Query  584  RAGYCSTVDVLKILSSYVDEDDYT--VW  609
              G  + V+ L IL   + E +    VW
Sbjct  570  YMGN-NDVERLSILRCVLTERELNAGVW  596
```

Key: Query; The gene product of Tb927.11.3570, Sbjct; gene product of Tb927.3.4750/90. +; Non-identical but similar residues. The sequence identity: 32% while sequence similarity: 69%.

Appendix VII: Occurrence of tricorn interacting factor 3 motifs in *Trypanosoma brucei brucei* genome

Sequence Name	Strand	Start	End	p-value	q-value
Motif 1					
AAZ13253	+	34	62	1.39E-07	0.266
AAZ11472	+	638	666	3.65E-07	0.301
AAZ12778	+	362	390	1.26E-06	0.711
EAN77617	+	100	128	1.27E-06	0.711
EAN80265	+	196	224	1.41E-06	0.711
EAN78869	+	362	390	1.70E-06	0.711
EAN79295	+	2323	2351	1.73E-06	0.711
EAN76691	+	290	318	2.23E-06	0.711
EAN78338	+	10	38	2.28E-06	0.711
EAN80233	+	1846	1874	2.68E-06	0.711
AAZ10735	+	149	177	2.80E-06	0.711
AAZ11676	+	366	394	2.86E-06	0.711
AAZ12077	+	36	64	3.42E-06	0.711
EAN79283	+	112	140	3.46E-06	0.711
EAN78188	+	106	134	3.66E-06	0.711
AAZ10863	+	514	542	3.73E-06	0.711
Motif 2					
AAZ10952	+	253	293	6.11E-08	0.377
AAZ11656	+	38	78	3.88E-07	0.393
AAZ12397	+	4118	4158	4.07E-07	0.393
EAN80097	+	170	210	5.18E-07	0.394
EAN79599	+	318	358	5.84E-07	0.394
AAZ11403	+	410	450	8.09E-07	0.397
EAN77526	+	167	207	1.06E-06	0.427
AAZ11856	+	178	218	1.32E-06	0.483
EAN80546	+	136	176	1.78E-06	0.508
EAN78731	+	31	71	1.90E-06	0.508
EAN76317	+	174	214	1.96E-06	0.508
AAZ11451	+	282	322	2.01E-06	0.508
AAZ13033	+	1366	1406	2.04E-06	0.508
EAN78819	+	446	486	2.18E-06	0.508
EAN78143	+	302	342	2.36E-06	0.508
AAZ11352	+	152	192	2.39E-06	0.508
EAN78082	+	89	129	2.87E-06	0.533
EAN77653	+	403	443	2.87E-06	0.533

Sequence Name	Strand	Start	End	p-value	q-value
Motif 2					
AAZ10354	+	298	338	2.91E-06	0.533
EAN79519	+	35	75	3.20E-06	0.553
EAN79761	+	81	121	3.29E-06	0.553
EAN78651	+	15	55	3.49E-06	0.563
AAZ11317	+	150	190	3.69E-06	0.572
EAN79151	+	492	532	4.05E-06	0.604
AAZ13023	+	105	145	4.40E-06	0.634
AAZ12226	+	343	383	4.69E-06	0.652
AAZ12420	+	599	639	5.17E-06	0.695
EAN79423	+	34	74	5.54E-06	0.707
AAZ11904	+	112	152	5.61E-06	0.707
EAN77575	+	35	75	6.15E-06	0.751
AAZ13522	+	324	364	6.65E-06	0.788
AAZ12184	+	300	340	6.97E-06	0.803
EAN76902	+	313	353	7.30E-06	0.818
EAN80499	+	267	307	7.89E-06	0.82
EAN79283	+	412	452	7.91E-06	0.82
AAZ12184	+	843	883	8.20E-06	0.82
EAN79200	+	665	705	8.63E-06	0.82
AAZ13515	+	422	462	9.15E-06	0.82
EAN80524	+	95	135	9.57E-06	0.82
AAZ11111	+	215	255	9.99E-06	0.82
CAJ16215	+	498	538	1.01E-05	0.82
AAZ13586	+	82	122	1.01E-05	0.82
AAZ12668	+	98	138	1.03E-05	0.82
Motif 3					
EAN80621	+	247	296	2.66E-06	0.609
EAN80438	+	3037	3086	6.35E-09	0.611
EAN78595	+	516	565	1.29E-08	0.613
EAN76289	+	25	74	2.46E-07	0.615
CAJ16537	+	1798	1847	4.67E-07	0.617
AAZ11041	+	1011	1060	1.47E-06	0.619
EAN77991	+	136	185	1.70E-06	0.619
EAN76500	+	574	623	1.74E-06	0.619
EAN76624	+	602	651	1.75E-06	0.619
AAZ12854	+	311	360	1.90E-06	0.619
AAZ13431	+	331	380	2.04E-06	0.619
AAZ10788	+	108	157	2.27E-06	0.637
CAJ16008	+	3791	3840	2.53E-06	0.654

Sequence Name	Strand	Start	End	p-value	q-value
Motif 3					
EAN80268	+	164	213	3.06E-06	0.709
AAZ10671	+	2166	2215	3.28E-06	0.716
EAN80091	+	472	521	3.60E-06	0.716
AAZ10349	+	309	358	3.85E-06	0.716
AAZ11858	+	731	780	3.95E-06	0.716
EAN78480	+	283	332	4.25E-06	0.716
AAZ13047	+	700	749	4.36E-06	0.716
EAN77298	+	312	361	4.37E-06	0.716
AAZ10935	+	78	127	4.92E-06	0.775
AAZ12320	+	451	500	5.15E-06	0.778
EAN79908	+	494	543	5.34E-06	0.778
EAN78497	+	104	153	5.61E-06	0.788
EAN79306	+	48	97	6.30E-06	0.843
EAN77762	+	41	90	6.63E-06	0.843
EAN79265	+	711	760	6.89E-06	0.843
EAN79092	+	2016	2065	6.99E-06	0.843
AAZ13574	+	62	111	7.24E-06	0.843
AAZ11574	+	131	180	7.44E-06	0.843
EAN78289	+	460	509	7.50E-06	0.843
AAZ10886	+	799	848	7.72E-06	0.843
EAN79235	+	97	146	7.94E-06	0.844
AAZ10268	+	474	523	8.44E-06	0.854
Motif 4					
AAZ13515	+	236	267	5.73E-10	0.000579
CAJ16666	+	352	383	3.67E-07	0.165
CAJ16669	+	352	383	3.67E-07	0.165
CAJ16660	+	352	383	3.67E-07	0.165
CAJ16664	+	352	383	3.67E-07	0.165
CAJ16662	+	352	383	3.67E-07	0.165
AAZ13116	+	102	133	5.67E-07	0.229
AAZ10583	+	814	845	8.55E-07	0.314
AAZ10300	+	577	608	1.83E-06	0.617
EAN76467	+	333	364	2.47E-06	0.671
AAZ11407	+	221	252	2.55E-06	0.671
EAN77905	+	10	41	2.59E-06	0.671
EAN77772	+	87	118	2.77E-06	0.671
EAN77891	+	295	326	2.82E-06	0.671
EAN79772	+	85	116	3.09E-06	0.693
EAN79557	+	947	978	3.72E-06	0.792

Sequence Name	Strand	Start	End	p-value	q-value
Motif 4					
AAZ11230	+	273	304	4.15E-06	0.839
AAZ13123	+	73	104	4.57E-06	0.88
AAZ13423	+	153	184	4.79E-06	0.88
AAZ13156	+	102	133	5.02E-06	0.882
EAN78120	+	444	475	6.03E-06	0.913
EAN79473	+	358	389	6.38E-06	0.913
EAN79320	+	813	844	6.42E-06	0.913
EAN77998	+	762	793	6.48E-06	0.913
EAN79868	+	1403	1434	6.48E-06	0.913
EAN80227	+	971	1002	6.55E-06	0.913
AAQ15906	+	132	163	6.94E-06	0.935
AAZ13558	+	3856	3887	7.21E-06	0.94
AAZ12391	+	772	803	7.51E-06	0.949
AAZ10548	+	176	207	8.25E-06	0.956
AAZ12906	+	3864	3895	8.65E-06	0.956
EAN78524	+	602	633	8.85E-06	0.956
EAN79979	+	204	235	8.94E-06	0.956
EAN80438	+	435	466	9.33E-06	0.956
EAN79889	+	297	328	9.90E-06	0.956
CAJ16398	+	95	126	1.01E-05	0.956
Motif 5					
EAN80603	+	650	721	1.89E-07	0.493
EAN76611	+	662	711	1.94E-07	0.491
EAN79519	+	262	311	3.86E-07	0.497
EAN80011	+	393	442	7.34E-07	0.498
AAQ16070	+	487	536	8.52E-07	0.498
AAZ10240	+	888	937	1.07E-06	0.553
AAZ12540	+	192	241	1.29E-06	0.566
EAN78590	+	366	415	1.52E-06	0.572
AAZ12854	+	744	793	1.59E-06	0.572
AAZ13111	+	68	117	2.00E-06	0.612
AAZ11834	+	48	97	2.02E-06	0.612
EAN78001	+	1700	1749	2.51E-06	0.638
AAZ12991	+	268	317	2.55E-06	0.638
CAJ16294	+	128	177	2.74E-06	0.638
EAN78847	+	16	65	2.86E-06	0.638
EAN76294	+	526	575	2.98E-06	0.638
AAZ11971	+	244	293	3.12E-06	0.638
AAZ13015	+	207	256	3.25E-06	0.638

Sequence Name	Strand	Start	End	p-value	q-value
Motif 5					
AAZ10388	+	185	234	3.40E-06	0.638
EAN77397	+	2	51	3.74E-06	0.648
AAQ15909	+	208	257	3.87E-06	0.648
EAN78473	+	13	62	3.94E-06	0.648
EAN79940	+	229	278	4.47E-06	0.686
EAN80236	+	67	116	5.27E-06	0.686
AAZ11040	+	994	1043	5.40E-06	0.686
AAZ13286	+	11	60	5.51E-06	0.686
AAZ12297	+	332	381	5.54E-06	0.686
EAN79482	+	1246	1295	5.58E-06	0.686
AAZ11259	+	59	108	5.68E-06	0.686
AAZ12825	+	1648	1697	5.75E-06	0.686
EAN77796	+	203	252	5.84E-06	0.686
AAZ10329	+	194	243	5.91E-06	0.686
EAN80193	+	180	229	6.43E-06	0.725
EAN78541	+	760	809	6.93E-06	0.734
AAQ15930	+	257	306	7.00E-06	0.734

The p-value of a motif occurrence is the probability of a random sequence of the same length as the motif matching that position of the sequence with as good or a better score while the q-value of a motif occurrence is the false discovery rate if the occurrence is accepted as significant.

Appendix VIII: Nucleic acid purity and quantity determination by a Nanodrop

User name	Date and Time	Nucleic Acid Conc.	Unit	A260	A280	260/280	260/230	Sample Type	Factor
APU-Pathology Internship	10/29/2015 4:30:39 PM	2553.5	ng/µl	51.071	29.642	1.72	1.49	DNA	50
APU-Pathology Internship	10/29/2015 4:31:23 PM	2804.4	ng/µl	56.088	32.214	1.74	1.54	DNA	50
APU-Pathology Internship	10/29/2015 4:31:51 PM	2582.9	ng/µl	51.658	30.175	1.71	1.47	DNA	50
APU-Pathology Internship	10/29/2015 4:32:20 PM	1029.1	ng/µl	20.582	12.064	1.71	1.28	DNA	50
APU-Pathology Internship	10/29/2015 4:32:47 PM	913.4	ng/µl	18.267	11.016	1.66	1.26	DNA	50
APU-Pathology Internship	10/29/2015 4:33:13 PM	775.6	ng/µl	15.511	9.795	1.58	1.54	DNA	50
APU-Pathology Internship	10/29/2015 4:33:44 PM	538.5	ng/µl	10.77	6.795	1.58	1.14	DNA	50
APU-Pathology Internship	10/29/2015 4:34:12 PM	950.6	ng/µl	19.013	11.669	1.63	1.64	DNA	50
APU-Pathology Internship	10/29/2015 4:34:37 PM	1161.5	ng/µl	23.229	13.606	1.71	1.66	DNA	50
APU-Pathology Internship	10/29/2015 4:35:04 PM	1232.8	ng/µl	24.657	13.857	1.78	1.66	DNA	50
APU-Pathology Internship	10/29/2015 4:35:54 PM	1466.1	ng/µl	29.322	17.087	1.72	1.65	DNA	50
APU-Pathology Internship	10/29/2015 4:36:21 PM	1123.7	ng/µl	22.475	13.748	1.63	1.69	DNA	50
APU-Pathology Internship	10/29/2015 4:36:45 PM	934.9	ng/µl	18.699	11.469	1.63	1.47	DNA	50
APU-Pathology Internship	10/29/2015 4:37:10 PM	491.2	ng/µl	9.824	6.535	1.5	1.25	DNA	50
APU-Pathology Internship	10/29/2015 4:37:35 PM	820.7	ng/µl	16.413	9.711	1.69	1.58	DNA	50

User name	Date and Time	Nucleic Acid Conc.	Unit	A260	A280	260/280	260/230	Sample Type	Factor
APU-Pathology Internship	6/3/2016 9:52:34 AM	1412.2	ng/µl	35.306	17.35	2.03	2.33	RNA	40
APU-Pathology Internship	6/3/2016 9:55:29 AM	1201.5	ng/µl	30.037	14.69	2.04	2.2	RNA	40
APU-Pathology Internship	6/3/2016 9:56:24 AM	1335.7	ng/µl	33.394	16.246	2.06	2.34	RNA	40

The highlighted samples represent the samples that were used in subsequent experiments.

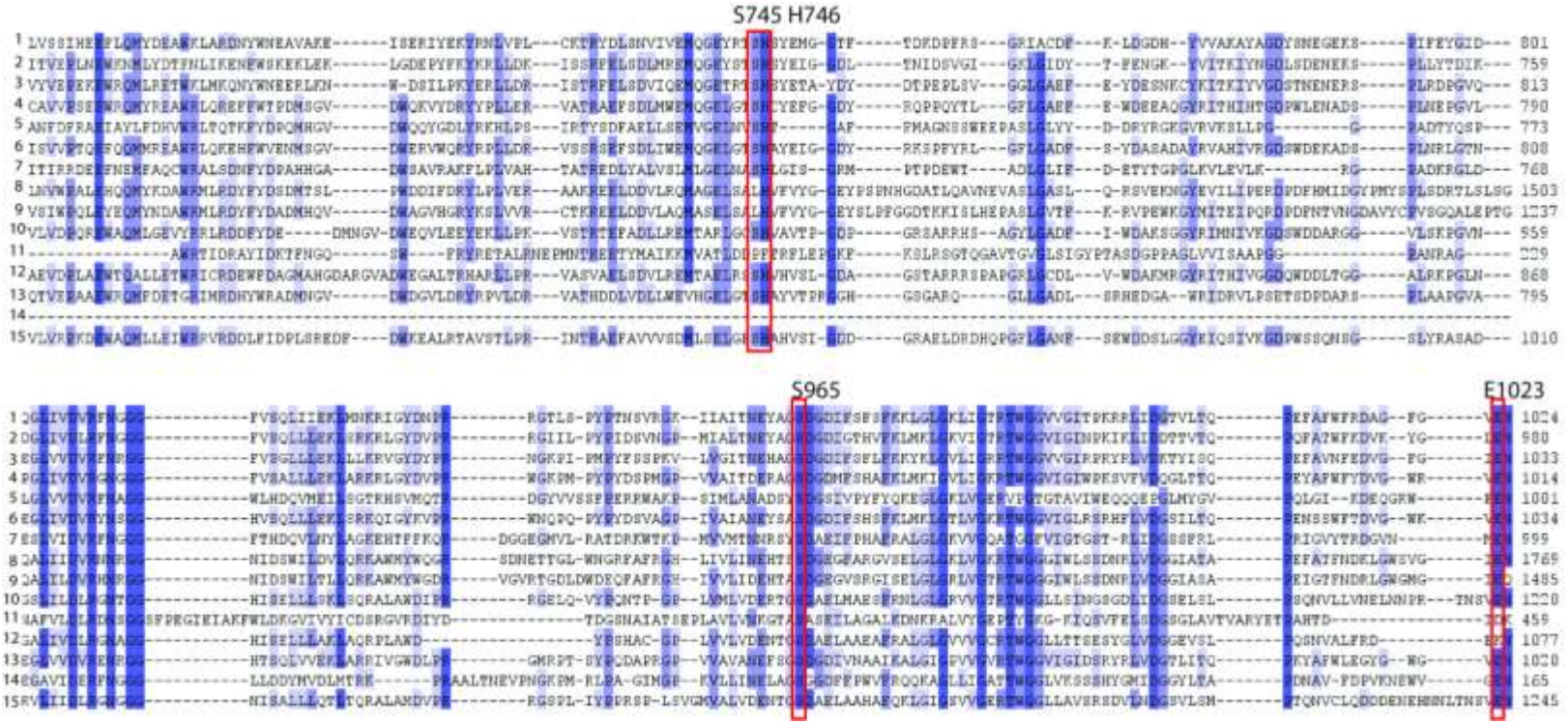
Appendix IX: BLAST homologies between tricorn protease, *Thermoplasma acidophilum* and its homologs

Description	Max score	E value	Identity	Accession
Tricorn protease, *Thermoplasma volcanium*	1138	0	54%	WP_010917153.1
Hypothetical protein, *Thermoplasmatales archaeon*	776	0	42%	EQB68734.1
Tricorn core peptidase, *Cuniculiplasma divulgatum*	774	0	42%	SIM28710.1
Tricorn protease, *Ferroplasma acidarmanus*	756	0	39%	WP_019841748.1
Peptidase S41, *Picrophilus torridus*	753	0	41%	WP_011177615.1
Peptidase S41, *Thermoproteus uzoniensis*	656	0	37%	WP_013679574.1
Tricorn protease, *Pyrobaculum aerophilum*	643	0	36%	AAL63388.1
Peptidase S41, *Sulfolobus solfataricus*	604	0	37%	WP_0099810076.1
Hypothetical protein, *Scytonema hofmannii*	602	0	34%	WP_017743915.1
Photosystem II D1 protease, *Micromonas pusilla*	600	2.00 E-179	34%	micr|ACO63313
Tricorn protease, *Nonomuraea solani*	559	7.00 E-179	33%	SEH01969.1
Tricorn protease homolog, *Candidatus acidaminovorans*	557	5.00 E-178	33%	WP_015423959.1
Peptidase S41, *Herbidospora cretacea*	556	1.00 E-177	33%	WP_061299287.1
Hypothetical protein, *Longispora albida*	554	1.00 E-176	34%	WP_018350392.1
Periplasmic protease, *Chlorobi bacterium*	550	9.00 E-175	34%	KXK57595.1
Peptidase S41, *Petrotoga mobilis*	541	8.00 E-172	34%	KUK16346.1
Hypothetical protein, *Kutzneria sp.*	443	3.00 E-135	31%	WP_052396111.1

Description	Max score	E value	Identity	Accession
Tricorn protease, *Lentzea waywayandensis*	429	1.00 E-129	30%	SFR27171.1
Tricorn protease, *Actinokineospora terrae*	429	2.00 E-129	30%	SES10996.1
Tricorn protease, *Alloactinosynnema album*	429	3.00 E-129	29%	SDI48145.1
Tricorn protease, *Streptomyces davawensis*	428	5.00 E-129	29%	WP_015660463.1
Carboxyl-terminal-processing, *Ricinus communis*	422	2.00 E-124	28%	rcom|29904.m003051
C-terminal processing protease, *Arabidopsis thaliana*	420	3.00 E-119	27%	atha|NP_193509
Predicted protein, *Micromonas pusilla*	257	1.00 E-67	25%	XP_003064769.1
Tricorn protease homolog, *Ostreococcus tauri*	200	3.00 E-49	29%	XP_003080833.1
Carboxyl-terminal protease, *Brucella suis*	189	6.00 E-44	28%	bsui|NP_698817
Carboxyl-terminal protease, *Deinococcus radiodurans*	180	1.00 E-42	29%	drad|NP_295274
Carboxyl-terminal protease, *Bacillus anthracis*	179	6.00 E-39	28%	bant|YP_022076
Protease, *Bathycoccus prasinos*	164	6.00 E-38	29%	XP_007515703.1
Carboxy-terminal protease, *Agrobacterium tumefaciens*	160	5.00 E-36	29%	atum|NP_355704
Predicted protein, *Phaeodactylum tricornutum*	157	1.00 E-35	27%	XP_002181061.1
Tricorn protease-like 1, *Symbiodinium microadriaticum*	153	3.00 E-34	34%	OLP82703.1
Predicted protein, *Thalassiosira pseudonana*	128	7.00 E-27	36%	XP_002295534.1
Hypothetical protein, *Nematostella vectensis*	117	1.00 E-26	33%	XP_001617415.1
Protease-related protein, *Vibrio cholera*	110	7.00 E-24	30%	vcho|NP_232446
Tail-specific protease, *Chlamydophila pneumoniae*	110	2.00 E-23	29%	cpne|NP_224751

Description	Max score	E value	Identity	Accession
Carboxy-terminal protease, *Escherichia coli*	110	7.00 E-18	29%	ecol|AP_002450
Carboxy-terminal protease, *Salmonella enterica*	110	2.00 E-17	28%	sent|NP_456350
Periplasmic tail-specific proteinase, *Rhodopirellula baltica*	110	9.00 E-17	28%	rbal|NP_870341
Carboxy-terminal protease, *Shigella flexneri*	110	5.00 E-16	27%	sfle|NP_707289
Carboxyl-terminal protease, *Clostridium perfringens*	110	9.00 E-07	31%	cper|YP_694758
Carboxyl-terminal protease, *Thermotoga maritima*	110	1.00 E-04	30%	tmar|NP_228556
Carboxyl-terminal protease, *Aquifex aeolicus*	110	1.00 E-04	28%	aaeo|NP_213546
Carboxyl-terminal protease, *Dehalococcoides ethenogenes*	108	1.00 E-03	27%	deth|YP_181110
Carboxyl-terminal protease, *Chlorobium tepidum*	107	2.00 E-03	26%	ctep|NP_661945
Carboxyl-terminal protease, *Staphylococcus aureus*	107	1.10 E-02	29%	saur|NP_371944
Tail specific periplasmic protease, *Rickettsia typhi*	105	1.40 E-01	27%	rtyp|YP_067184

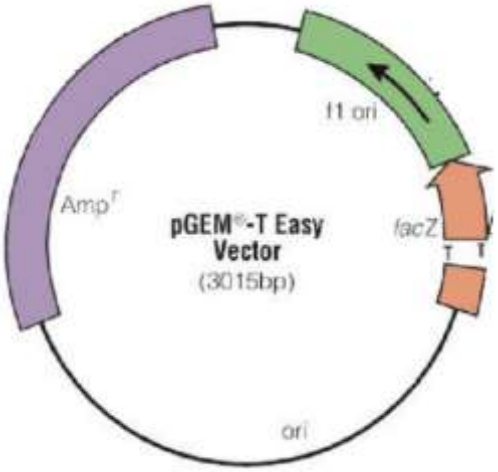
The table was sorted by increasing E value.



Appendix X: Multiple sequence alignment of catalytic domain 1 (C 1) and catalytic domain 2 (C 2) of tricorn protease and its homologs. The active site tetrad S 745, H 746 are located in C 1 while S 965 and E 1023 are located in C 2 domain (as shown in the red boxes) with the positions corresponding to tricorn protease in *Thermoplasma acidophilum*. The sequences were retrieved from GenBank: 1; Tricorn protease, *Thermoplasma acidophilum*, 2; tricorn protease, *Ferroplasma acidarmanus*, 3; Tricorn protease, *Sulfolobus salfataricus*, 4; Tricorn protease, *Pyrobaculum aerophilum*, 5; hypothetical protein, *Aminicenantes bacterium* 6; tricorn protease, *Vibrio cholerae*, 7; tricorn protease, *Streptomyces coelicor*, 8; Hypothetical protein, *Scytonema hofmanni*, 9; MdsD protein, *Gemmata obscuriglobus*, 10; Predicted protein, *Thalasiossira pseudonana*, 11; Tricorn protease, *Ostreococcus tauri*, 12; D 1 processing protease, *Arabidopsis thaliana*, 13; Predicted protein, *Micromonas pusilla*, 14; Hypothetical protein, *Nematostella vectensis*, 15; Protease, *Bathycoccus prasinos*.

Appendix XI: List of vectors maps and sequences

1. pGEM_T Easy vector map

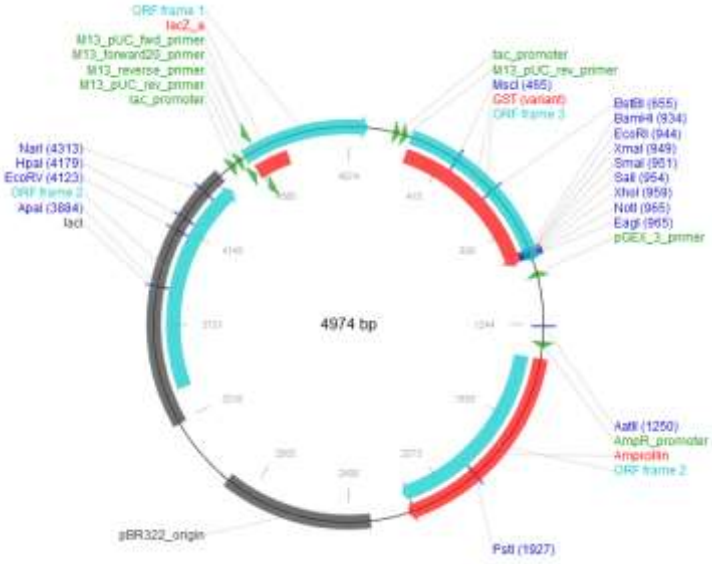


>pGEM-T Easy Vector

gggcgaattgggcccgacgtcgcatgctcccggccgccatggcggccgcgggaattcgatnatcactagtgaattcgcggccgcctgcaggtcgacc
atatgggagagctcccaacgcgttggatgcatagcttgagtattctatagtgtcacctaaatagcttggcgtaatcatggtcatagctgtttcctgtgtgaaatt
gttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgc
tcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccg
cttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggga
taacgcaggaaagaacatgtgagcaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctg
acgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgc
tctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggt
gtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaa
gacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaact
acggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccacc
gctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagt
ggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagta
tatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgt
gtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaa
ccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcg
ccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaag
gcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggtt
atggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcg
accgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaa
ctctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagca
aaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcag
ggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgatgcggtgt
gaaataccgcacagatgcgtaaggagaaaataccgcatcaggaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttt
taaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaa
gaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtg
ccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaag

cgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcg
tccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggc
gattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgaattgtaatacgactcactata

2. pGEX-5x-3 map

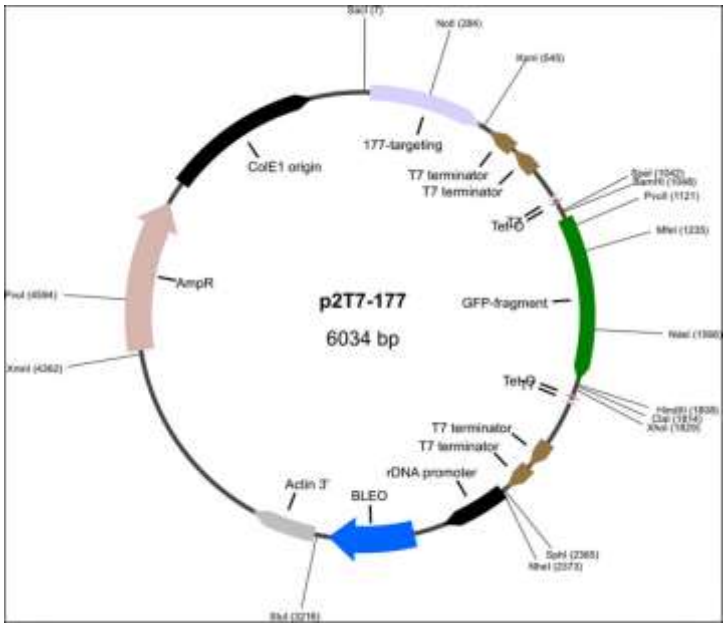


>pGEX-5x-3 sequence

agcttatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtc
gctcaaggcgcactcccgttctggataatgtttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttgacaattaatcatcggctcgt
ataatgtgtggaattgtgagcggataacaatttcacacaggaaacagtattcatgtcccctatactaggttattggaaaattaagggccttgtgcaacccact
cgacttcttttggaatatcttgaagaaaaatatgaagagcatttgtatgagcgcgatgaaggtgataaatggcgaaacaaaaagtttgaattgggtttggagt
ttcccaatcttccttattatattgatggtgatgttaaattaacacagtctatggccatcatacgttatatagctgacaagcacaacatgttgggtggttgtccaaa
agagcgtgcagagatttcaatgcttgaaggagcggttttggatattagatacggtgtttcgagaattgcatatagtaaagactttgaaactctcaaagttgattt
tcttagcaagctacctgaaatgctgaaaatgttcgaagatcgtttatgtcataaaacatatttaaatggtgatcatgtaacccatcctgacttcatgttgtatgac
gctcttgatgttgttttatacatggacccaatgtgcctggatgcgttcccaaaattagtttgttttaaaaaacgtattgaagctatcccacaaattgataagtactt
gaaatccagcaagtatatagcatggcctttgcagggctggcaagccacgtttggtggtggcgaccatcctccaaaatcggatctgatcgaaggtcgtggg
atccccaggaattcccgggtcgactcgagcggccgcatcgtgactgactgacgatctgcctcgcgcgtttcggtgatgacggtgaaaacctctgacacat
gcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcgggg
cgcagccatgacccagtcacgtagcgatagcggagtgtataattcttgaagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataata
atggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaat
aaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgttttg
ctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttga
gagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtgttgacgccgggcaagagcaactcggt
cgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgc
cataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgta
actcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaa
actattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttcc
ggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtag
ttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagacca

agtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagt
tttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccg
ctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgt
agccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgt
gtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaac
gacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcag
ggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgattt
ttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcc
tgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagc
gaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcataaattccgacaccatcgaatggtgcaaaacctttt
cgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctctt
atcagaccgtttcccgcgtggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattccc
aaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgcaaattgtcgcggc
gattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttct
cgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgat
gtctctgaccagacacccatcaacagtattattttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgc
gctgttagcgggcccattaagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaacggg
aaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgatcagatggc
gctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttatat
cccgccgttaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggca
atcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagct
ggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgctt
ccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacggattcactggccgtcgttttacaacgtcgt
gactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgccct
tcccaacagttgcgcagcctgaatggcgaatggcgctttgcctggtttccggcaccagaagcggtgccggaaagctggctggagtgcgatcttcctgag
gccgatactgtcgtcgtcccctcaaactggcagatgcacggttacgatgcgcccatctacaccaacgtaacctatcccattacggtcaatccgccgtttgtt
cccacggagaatccgacgggttgttactcgctcacatttaatgttgatgaaagctggctacaggaaggccagacgcgaattatttttgatggcgttggaatt

3. p2T7-177



>p2T7-177 sequence

gaattcgatgcgaggcgaatcgctcagtagtgttagtgcaggcattgtttctcttttcctttcggctcgctctctgtgccagaaaggcttttggagtcttggtta
ctcgtgggactagggcaagcgtcgcttccctaataattgttgctaacgtttcaccgtgcgtgtgtggatatcttccaagttgtcattttctctacccctctctttttc
cggcgcggtctaacttcctccaactagcgcatttccatgataacggggtccacatagtccgttgtagtgaaattggcatgccatcagggccaaggcataca
tcgcggccgcgataaaaatgatacgatgatctccgtgtggagcagaccgtggcgaccagtttgtaccatggaagttcttcgatagagttatatcgagtgcg
gaattgatgctacggtcacagtgtcgtatcgtgtggtctcccatgtgtttttgacggtgtaccaattatctttgcactaatgggctatggaatctttccttttaatca
gggggacatgtttatcccacgccagtagtctggaattagaatacatacccaactatatggcgcgccgagctcgaattctgcattaatgaatcggccaacgcg
cggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactc
aaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaagg
ccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaa
gataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtgg
cgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgc
gccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatg
taggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaa
aaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaa
gaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagat
ccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatc
tgtctatttcgttcatccatagttgcctgactccccgtggtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcga
gacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatcca
gtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgt
ttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgt
tgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtga
gtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaa
aagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgat
cttcagcatcttttacttttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaat
actcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttcc
gcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtctcgc
gcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtc
agggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccattcgacgctctcccttat
gcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcaaggagatggcgcccaacagtcc
cccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgata
taggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatctggctaggaatactgcatagataacaaacgc
atcaaacacaactggtggcactatttcaatcatgtcgacggcgcgccctcggacgagtgctggggcgtcggtttccactatcggcgagtacttctacaca
gccatcggtccagacggccgcgcttctgcgggcgatttgtgtacgcccgacagtcccggctccggatcggacgattgcgtcgcatcgaccctgcgccc
aagctgcatcatcgaaattgccgtcaaccaagctctgatagagttggtcaagaccaatgcggagcatatacgcccggagccgcggcgatcctgcaagct
ccggatgcctccgctcgaagtagcgcgtctgctgctccatacaagccaaccacggcctccagaagaagatgttggcgacctcgtattgggaatccccga
acatcgcctcgctccagtcaatgaccgctgttatgcggccattgtccgtcaggacattgttggagccgaaatccgcgtgcacgaggtgccggacttcggg
gcagtcctcggcccaaagcatcagctcatcgagagcctgcgcgacggacgcactgacggtgtcgtccatcacagtttgccagtgatacacatggggat
cagcaatcgcgcatatgaaatcacgccatgtagtgtattgaccgattccttgcggtccgaatgggccgaacccgctcgtctggctaagatcggccgcagc
gatcgcatccatggcctccgcgaccggctgcagaacagcgggcagttcggtttcaggcaggtcttgcaacgtgacaccctgtgcacggcgggagatg
caataggtcaggctctcgctgaattccccaatgtcaagcacttccggaatcgggagcgcggccgatgcaaagtgccgataaacataacgatctttgtaga
aaccatcggcgcagctatttacccgcaggacatatccacgccctcctacatcgaagctgaaagcacgagattcttcgccctccgagagctgcatcaggtc
ggagacgctgtcgaacttttcgatcagaaacttctcgacagacgtcgcggtgagttcaggctttttcatcactagttctagagcttattttatggcagcaacga
gaccttacgtaaagcaaaactaaataagagcggagactgcaatgcagagtaaaaaaaaaaaaaaacatgaatttcagaagaccttgctgtgcccgggctc
tgagagcggtcagttgctctcctgctccgtacgccgtaagcgctacttttactgcttaatctttcctatataagatatatttctattactatatatacttttttaacttta
tcattattatcctattgtgtgttgtattgactttacaataaaacatattatatctaacatgtcaaaagcatcggctcttcaatacagttacccgtcccttatcaacac
agtacaatatactctgtcatgacaataagtatatataagagccagaatgcacccgcgctgggtggaaagctagcttgcatgcctgcaggggcccaatccg
gatatagttctcctttcagcaaaaaacccctcaagacccgtttagaggccccaaggggttatgctagttattgctcagcggtggcagcagccaactcagctt
cctttcgggctttgttagcagccctatccggatatagttctcctttcagcaaaaaacccctcaagacccgtttagaggccccaaggggttatgctagttattgc
tcagcggtggcagcagccaactcagcttcctttcgggctttgttagcagcctgcagaatactgcatagataacaaacgcatcaaacacaactggtggcact
atttcaatcatgtcgaggaggaccgaatactaataacgacacttgcggtcaaaaagtagaagaacaaatgctcaacgatgagtgaatcaggttagggtag

tcggaaaattatacagaatgtctttggcaacacaccggctgattaatacgactcactatagggagatctccctatcagtgatagagatctccccgggccccc
cctcgagttaacgaccttgcgtccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcg
aaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagtttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaa
gggagcccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctg
gcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtccattcgccattcaggctgcgcaactgtt
gggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcc
cagtcacgacgttgtaaaacgacggccagtgaattgtaatacgactcactatagggcgaattgggcccgacgtcgcatgctcccggccgccatggcgg
ccgcgggaattcgatatcactagtgaattcgcggccgcctgcaggtcgaccatatgggagagctcccaacgcgttggatgcatagcttgagtattctatag
tgtcacctaaatagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagt
gtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgcgacgcaaggtcgttaacggatccactagttctagagcggccggccgc
caccgcggctagatctctatcactgatagggagatctccctatagtgagtcgtattaatcagccggtgtgttgccaaagacattctgtataattttccaactac
cctaacctgattcactcatcgttgagcatttgttcttctacttttgaccgcaagtgtcgttattagtattcggtcctgggctgcaggctgctaacaaagcccgaa
aggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgagggtttttgctgaaaggagaactat
atccggatagggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgg
gtcttgaggggtttttgctgaaaggagaactatatccggattgggcccctgcag