Enzymatic characterisation of a CATL-like protease from *Trypanosoma brucei brucei* and small-subunit rRNA sequence based phylogenetic analysis of freshwater fish trypanosomes

By

Bongumusa Comfort Mthethwa

BSc (Hons) Biochemistry

(University of Zululand)

Submitted in the fulfilment of the

academic requirements for the

MSc degree

in

Biochemistry

School of Life Sciences

University of KwaZulu-Natal

Pietermaritzburg

2021

PREFACE

The experimental work described in this dissertation was carried out in the School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, from January 2019 to December 2020, under the supervision of Professor THT Coetzer and co-supervision of Dr Sandi Willows-Munro. The studies represent original work by the author and have not otherwise been submitted in any other form to another University. Where use has been made of the work of others, it has been duly acknowledged in the text.

[redacted]

Bongumusa Mthethwa
April 2021

As the candidate's supervisors we agree to the submission of this dissertation.

[redacted]

Prof. Theresa H. T. Coetzer

[redacted]

Dr Sandi Willows-Munro

DECLARATION - PLAGIARISM

I Bongumusa Mthethwa, declare that

1. The research reported in this dissertation, except where otherwise indicated, is my original research.

2. This dissertation has not been submitted for any degree or examination at any other university.

3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

    a. Their words have been re-written, but the general information attributed to them has been referenced
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.

5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the Reference section.

_____

Bongumusa Mthethwa

April 2021

ACKNOWLEDGEMENTS

I waited patiently for the Lord; he returned to me and heard my cry (Psalm. 40:1).

I would like to express my gratitude and appreciation to the following persons:

Firstly, to my supervisor, Prof. Coetzer, for the opportunity to complete this degree under your guidance, your time, patience and your belief in my ability.

To my co-supervisor, Dr Willows-Munro, for your time, all the assistance and guidance throughout the phylogenetic studies.

To Charmaine Ahrens, Pat Joubert, Sibongile Khuzwayo and Tanya Karalic for all their assistance in financial and administrative matters.

To my fellow postgraduates: Lucky Marufu, Pretty Gumede, Nomusa Zondo, Thando Maseko, Nxalati Mkhombo, Ephraim Chauke, Faiaz Shaik, Arishka Maikoo, Akira Boodhoo, Bhavana Ramjeawon, Chanelle Grantham and Lauren Eyssen for their advice, encouragement and friendship.

To Prof. Annemariè Avenant-Oldewage, Department of Zoology, University of Johannesburg for providing the fish blood samples.

To Prof. Coetzer and National Research Foundation for the financial support.

To Nomacusi Sibeko, Lucky Khubone, Zolile Malembe and Sibusiso Ngeleka for your help, support and encouragement throughout my postgraduate studies.

To Sibonile Sibiya for the love, patience, support and happy memories.

Finally, to my family, my mom and my siblings for all their support, love and constant encouragement during difficult times.

ABSTRACT

Trypanosomes infect different classes of vertebrates, including humans, animals, birds, fish, amphibians and reptiles. In Sub-Saharan Africa, tsetse flies transmit trypanosomes in both humans and other animals, while leeches possibly transmit fish trypanosomes. African trypanosomiasis mainly affects people and their livestock living in impoverished rural areas, hence affordable diagnostic tests and drugs are required. The development of a vaccine is unlikely due to the ability of trypanosomal parasites to undergo antigenic variation, enabling the parasites to avoid the host immune system. During infection of the vertebrate host, trypanosome peptidases, such as cathepsin L-like cysteine proteases, hydrolyse essential host serum proteins and hormones. Therefore, these proteases have been identified as virulence factors that contribute to the manifestation of various trypanosomiasis symptoms such as anaemia, fever, paralysis, and disturbances in sleep-wake cycles. Proteases are consequently drug and diagnostic targets.

Fish trypanosomes were among the first species of trypanosomes described and are phylogenetically related to those infecting mammals. However, the impact of trypanosome infection on natural fish populations is not known because of a lack of studies. Formerly used fish trypanosome taxonomy, based on host specificity and changing kinetoplast and nucleus positions, is unreliable. Biological marker genes, such as 12S rRNA or 18S rRNA, and glycolytic glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) are reliable for phylogenetic analysis of trypanosomes.

The cathepsin L-like cysteine protease from *Trypanosoma brucei brucei* (*Tb*CATL) and trypanosomes infecting the freshwater fish, *Clarias gariepinus*, were the main focus of this study. The full-length *TbCATL* gene was cloned into a pGEMT®-Easy cloning vector, sub-cloned into a pET28a expression vector and recombinantly expressed using the *Escherichia coli* BL21 DE3 expression system at a size of the approximately 45 kDa. The *Tb*CATL was expressed within inclusion bodies as confirmed by a western blot probed with anti-His tag antibodies. The *Tb*CATL expression was optimised under varying temperatures, and isopropyl β-d-1-thiogalactopyranoside (IPTG) concentrations in an attempt to solubilise the inclusion bodies but sarkosyl solubilisation was required preceding immobilised metal affinity chromatography (IMAC) purification. Enzymatic characterisation of purified recombinant *Tb*CATL revealed gelatinolytic as well as peptidolytic activities. The kinetic parameters, $K_{m,}$ and $k_{cat}/K_{m,}$ were determined with a range for peptide substrates to show the substrate specificity of recombinant *Tb*CATL. The optimum pH for peptide hydrolysis was determined using constant ionic strength buffers and found to be optimal from pH 5 to 5.5. An investigation

of the interaction with a range of catalytic class-specific inhibitors showed that only cysteine protease-specific inhibitors inhibited *Tb*CATL completely.

Genomic DNA was isolated from *Clarias gariepinus* blood samples. Primers targeting the trypanosome small-subunit (SSU) rRNA gene were used to amplify and sequence six representative trypanosome sequences collected from *Clarias gariepinus.* These data were analysed together with 32 NCBI GenBank sequences previously isolated from fish. Maximum likelihood (ML) and Bayesian inference (BI) were used in phylogenetic analyses. The resulting phylogenetic trees recovered two main lineages. One lineage containing trypanosomes collected from freshwater fish species (branch support: ML bootstrap = 89%; BI posterior probabilities = 0.94) and second lineage including parasites collected from marine fish (branch support: ML bootstrap = 100%; BI posterior probabilities = 0.94). All six trypanosomes collected from *Clarias gariepinus* belong to the freshwater clade and were not identical and did not form a well-supported monophyletic clade suggesting that this fish species may be infected by multiple lineages of trypanosomes.

This study laid a foundation for conducting further studies on *Tb*CATL as a chemotherapeutic and a diagnostic target for African trypanosomiasis. This preliminary study of the phylogeny of *Clarias gariepinus* trypanosomes provides a foundation for future studies of the evolution and diversification of these parasites in fish species.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

2 x YT	2 x yeast extract, tryptone
AAT	animal African trypanosomiasis
AMC	7-amino-4-methylcoumarin
BCA	bicinchoninic acid
bp	base pair
BSA	bovine serum albumin
CATL	cathepsin-L
CATT	card agglutination test
dH_2O	distilled H_2O
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
g	relative centrifugal force
gGAPDH	glycolytic glyceraldehyde-3-phosphate dehydrogenase
GPI	glycosylphosphatidylinositol
h	hour(s)
HAT	human African trypanosomiasis
HRPO	horseradish peroxidise
IMAC	immobilised metal ion affinity chromatography
IPTG	isopropyl-β -D-thiogalactopyranoside
kDa	kilodalton
MCA	metacaspase

min	minute(s)
M_r	relative molecular mass
MSP	surface metalloprotease
PBS	phosphate buffered saline
PCE	sarkosyl elution buffer
PCL	sarkosyl lysis buffer
PCR	polymerase chain reaction
PCW	sarkosyl wash buffer
PEG	polyethylene glycol
pI	isoelectric point
RT	room temperature
sarkosyl	N-lauroylsarsocine sodium salt
SBTI	soya bean trypsin inhibitor
SDS	sodium dodecyl sulfate
sec	second(s)
SIF	Stumpy induction factor
SSU rRNA	small subunit ribosomal ribonucleic acid
TAE	tris-acetate-EDTA
TBS	tris buffered saline
TNF	tumour necrosis factor
Tris	2-amino-2-(hydroxymethyl)-1,3-propandiol
VSG	variant surface glycoprotein
X-gal	isopropyl-β-D-thiogalactopyranoside

1 CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

Trypanosomes are protozoan parasites that infect many different classes of vertebrates (Fermino *et al.*, 2015), including humans, animals and aquatic vertebrates (fish, amphibians and reptiles) (Corrêa *et al.*, 2016). African trypanosomes are free-living in the host bloodstream and negatively influence livestock production and human health in 37 Sub-Saharan countries (Muhanguzi *et al.*, 2017). Animal trypanosomes affect livestock production in African countries, South America and Asia (Figure 1.1) (Giordani *et al.*, 2016). Different *Trypanosoma* species cause disease in humans: *Trypanosoma cruzi* causes Chagas disease (American trypanosomiasis), and *Trypanosoma brucei gambiense* and *T. b. rhodesiense* causes sleeping sickness (human African trypanosomiasis, HAT) (Nussbaum *et al.*, 2010), *Trypanosoma congolense*, *T. vivax* and *T. b. brucei* infect animals. While the former two species cause nagana (animal African trypanosomiasis, AAT), the latter is non-pathogenic and a good model organism used in research (Stijlemans *et al.*, 2018). The insect vector often transmits trypanosomes to the host bloodstream and tissue fluid (Barrett *et al.*, 2003). The most prevalent vector for African trypanosomes is the tsetse fly (*Glossina* spp.) found in sub-Saharan Africa (Steverding, 2008; Muhanguzi *et al.*, 2017). *Trypanosoma evansi* is spread by other biting insects such as horseflies to camels, horses, buffalo and cattle, causing surra in these animals in northern Africa, Asia and South America. *Trypanosoma equiperdum*, found in the north of Africa, infects horses through venereal transmission and causes dourine (Desquesnes *et al.*, 2013).

The HAT affects primarily poor people living in remote areas of Africa and may result in the death of the infected hosts if untreated, and ATT results in significant economic loss and makes it difficult to farm with livestock in affected regions (Steverding, 2017). The impact of trypanosomiasis on human health and livestock production requires an integrated strategy for disease control (Hamill *et al.*, 2017). Programmes for controlling trypanosomiasis have been going on for years. However, the disease remains a devastating set-back in livestock production (Meyer *et al.*, 2016). Current interventions are resource-intensive, costly and not feasible in remote areas, making it almost impossible for the complete elimination of the disease (Sutherland *et al.*, 2017). In addition to factors that hinder the effective control of trypanosomiasis, vector control only covers 1.3% of the tsetse infested area, parasites are resistant to trypanocides due to new veterinary drugs for AAT last being produced in 1985 (Meyer *et al.*, 2016). African trypanosomes can periodically switch the expression of variant surface glycoproteins (VSGs) to new VSGs, a process known as antigenic variation, enabling parasites to evade the host's immune system (Horn, 2014). This drawback attracts

researchers to develop new drugs and novel methods for controlling African trypanosomiasis (Meyer *et al.*, 2016). Studies showed that characterisation and a better understanding of *Trypanosoma* proteases are promising for developing and improving existing drugs, as they contribute to these parasites' pathogenesis (Steverding, 2008) as will be further elaborated on in later sections.

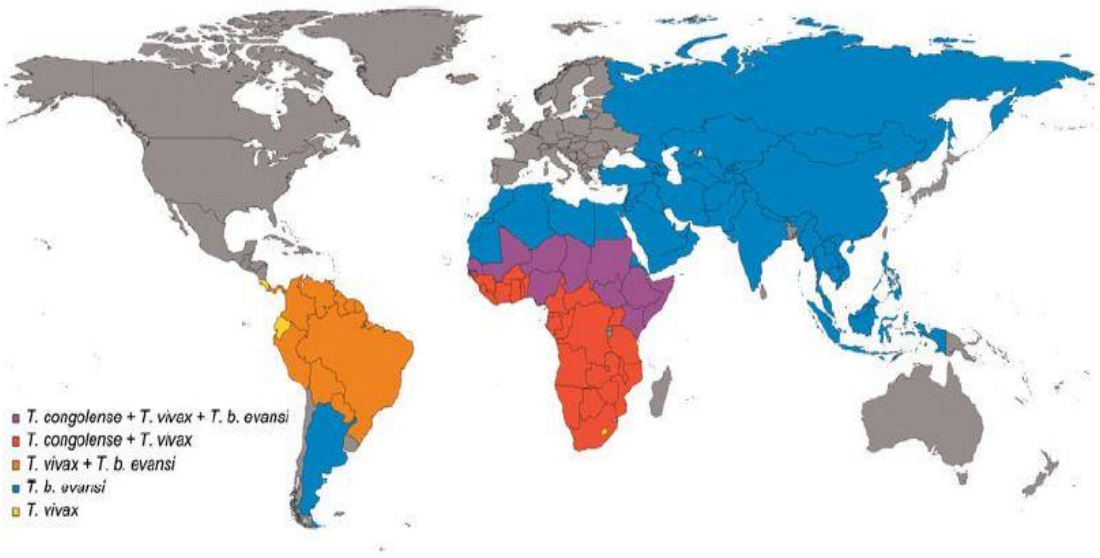

Figure 1.1: Map of animal trypanosomes from different countries. *Trypanosoma* species affecting livestock production from various countries (Giordani et al., 2016).

Mammalian trypanosomes have been the main subject in research, while those that infect fish and birds are being neglected (Maslov *et al.*, 1996; Grybchuk-Ieremenko *et al.*, 2014). Even though fish trypanosomes were among the first described species of the *Trypanosoma* genus, studies showing the impact of fish trypanosomes occurring in the natural environment is still lacking, particularly at the molecular level (Oladiran and Belosevic, 2012; Grybchuk-Ieremenko *et al.*, 2014). Fish trypanosomes infect both freshwater and marine fish. Aquatic blood-sucking leeches species transmit trypanosomes in marine and freshwater fish, causing a disease called piscine trypanosomiasis (Gupta, 2006; Corrêa *et al.*, 2016). Over 200 fish trypanosome species have been identified, and 60 of these *Trypanosoma* species originate in Brazil (Corrêa *et al.*, 2016). In aquaculture production, fish trypanosomiasis results in economic losses due to increasing infections, with the mortality rate reaching up to 35% in farmed fish (Gupta, 2006; Rodrigues *et al.*, 2018). *Trypanosoma carassii* is the most studied fish trypanosome species, and its infection rate reaches up to 100% in farmed fish (Overath *et al.*, 2001). Therefore, understanding these parasites' biology at a molecular level is essential for controlling fish trypanosomiasis (Overath *et al.*, 1998; Overath *et al.*, 2001).

In the past years, the taxonomy of fish trypanosomes was based on morphological features, such as cell shape, the position of the nucleus and kinetoplast, and host specificity (Maslov *et al.*, 1996; Grybchuk-Ieremenko *et al.*, 2014). However, taxonomy using such features have proven unreliable due to several morphological changes that occur during development. Single *Trypanosoma* species can infect different fish hosts (Gupta, 2006; Grybchuk-Ieremenko *et al.*, 2014). *Trypanosoma carassii* infects different freshwater fish (cyprinids) and some non-cyprinid family members (Overath *et al.*, 1998). Various molecular markers such as genes encoding 12S rRNA, 18S rRNA and glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) are more reliable sources of data that can be used in phylogenetic reconstructions to clarify the taxonomy of the group. The 18S rRNA gene in particular has a mutational rate suitable for constructing phylogenetic trees and separating *Trypanosoma* species. Therefore, it is commonly used in fish trypanosomes' phylogeny (Maslov *et al.*, 1996; Grybchuk-Ieremenko *et al.*, 2014).

1.2 Classification and morphology of trypanosomes

The genus *Trypanosoma* comprises eukaryotes belonging to the order Kinetoplastida, in the phylum Euglenozoa, together with two other groups: Euglenida and Diplonemida (Figure 1.2) (Moreira *et al.*, 2004; Lemos *et al.*, 2015). The kinetoplast consists of a network of tightly packed circular DNA (mini- or maxi circles) inside the mitochondrion and is situated near the flagellar pocket (Welburn *et al.*, 2016). Due to trypanosomes' impact on public health and agriculture, the taxonomic classification of trypanosomatids has gained more attention than other kinetoplasts (Kaufer *et al.*, 2020). The phylogenetic analysis of *Trypanosoma* using ribosomal SSU rRNA shows that Euglenozoa is a monophyletic group and share a common ancestor (Maslov *et al.*, 1996; Hamilton *et al.*, 2004). Classification of trypanosomatids based on morphology and lifecycle shows that they represent two suborders: Bodonina and Trypanosomatina. Bodonina includes two families: Bodonidae and Crytobiidae, characterised by large kinetoplast and two free flagella originating from the anterior end. Bodonids are a diverse group of free-living (*Bodo*) and parasitic (*Parabodo*) parasites. *Cryptobia* parasites are transmitted by leech vectors, found in the reproductive tract of snails and fish's gastrointestinal tract (Wright *et al.*, 1999; Welburn *et al.*, 2016). Trypanosomatidae are obligate parasites forming a single family of trypanosomatids, characterised by a single flagellum and a small kinetoplast (Hamilton *et al.*, 2004; Cayla *et al.*, 2019). The genus *Trypanosoma* is further divided into two groups based on their transmission mode: Salivaria and Stercoraria (Wright *et al.*, 1999; Baral, 2010). Salivaria are transmitted through tsetse fly saliva (*T. brucei*, *T. congolense*, and *T. vivax*), and Stercoraria are transmitted through faeces of a vector (*T. cruzi*) (Baral, 2010).

Mixed infections occurring in nature make it difficult to accurately classify trypanosomes (Borges *et al.*, 2016). However, various biomarker genes, such as SSU rRNA, have made it simpler to accurately classify these species (Gibson *et al.*, 2005; Grybchuk-Ieremenko *et al.*, 2014). Using the SSU rRNA genes revealed that fish trypanosomes belong to a monophyletic group of the aquatic clade of *Trypanosoma* (Hamilton *et al.*, 2004). The aquatic clade has two lineages: freshwater fish trypanosomes and marine water fish trypanosomes (Dóró *et al.*, 2019). Fish trypanosomes are a close relative of mammalian trypanosomes belonging to the family Trypanosomatidae (Lemos *et al.*, 2015). *Cryptobia* parasites are sister relatives of fish trypanosomes, which belong in the aquatic clade, and they belong to Cryptobiidae, e.g., *Cryptobia salmostica* and *Trypanoplasma borreli* (Maslov *et al.*, 1996).

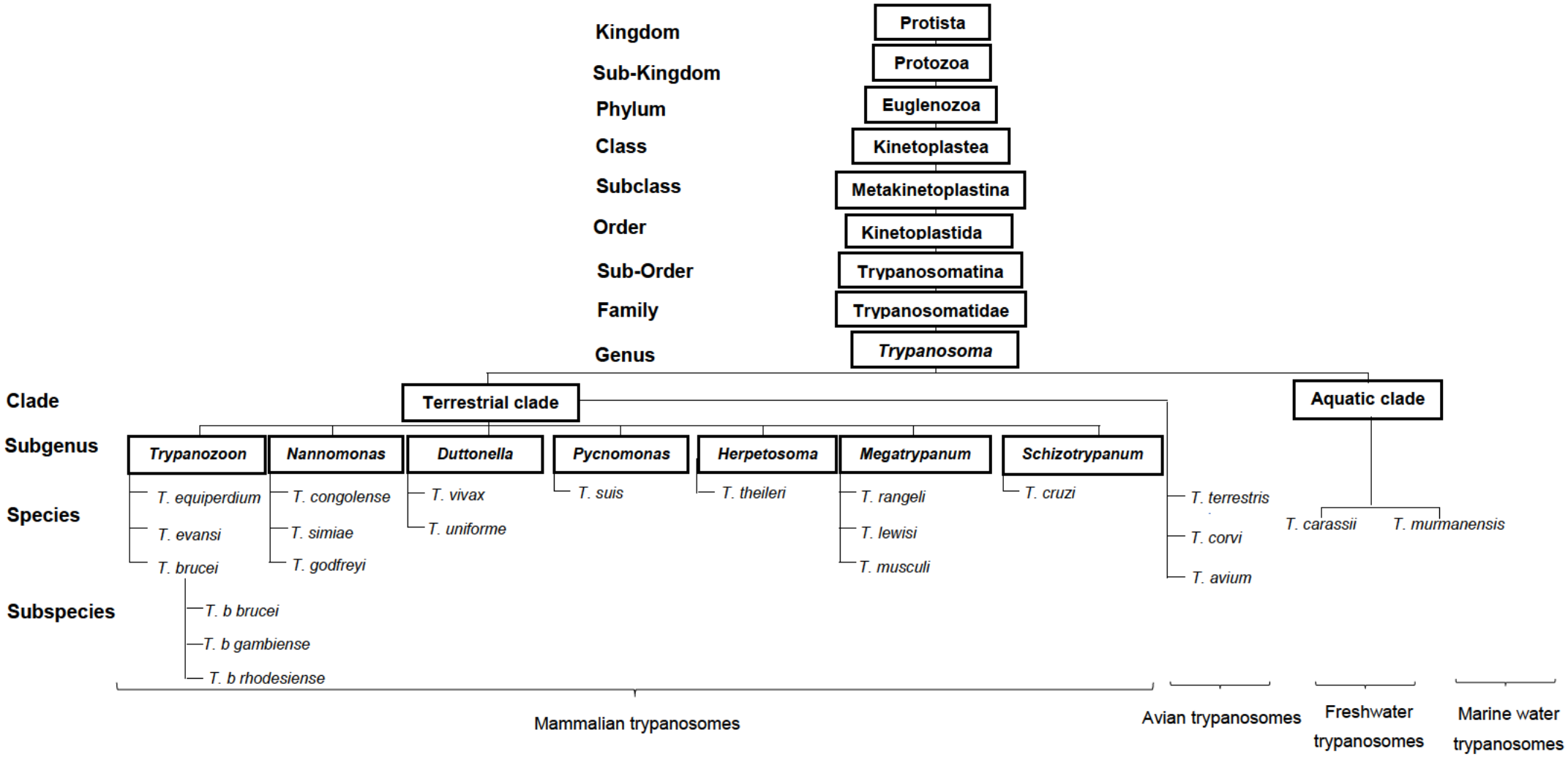

Figure 1.2: Summarised taxonomic classification of trypanosomes. Taxonomic classification showing clades, different sub-genus, species and subspecies of *Trypanosoma*, adapted from (Baral, 2010).

Trypanosomes are unicellular, approximately 15-30 µm long and slender (Figure 1.3). They have a typical eukaryotic cellular organisation, containing one tubular mitochondrion, a kinetoplast and a circular mitochondrial DNA (Brun *et al.*, 2010). A single flagellum of *Trypanosoma* originates from the basal body and is connected to the kinetoplast by filaments at the cell's anterior end (Ooi and Bastin, 2013). It runs through an undulating membrane to the body's posterior end, where it emerges from the flagellar pocket (Lacomble *et al.*, 2010). The flagellum's primary function is to propel trypanosomes through the bloodstream and tissue fluids of the host. All the organelles are located between the posterior end and the centre of the cell. Endo- and exocytosis are the most important processes that take place in the flagellar pocket. These processes are responsible for recycling the variant surface glycoproteins in the bloodstream forms (Matthews, 2005). The nucleus is usually located in the centre of the body, but the position may differ among species (Gupta, 2006).

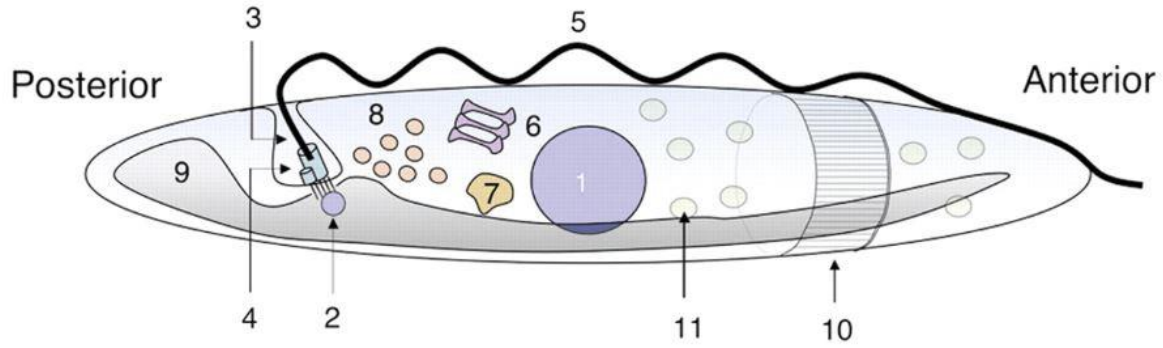

Figure 1.3: Cell structure of a trypanosome. Cell organelles located in various positions in the body of the trypanosome: 1, Nucleus; 2, Kinetoplast; 3, Flagellar pocket; 4, Basal body and probasal body; 5, Axoneme and flagellar rod; 6, Golgi; 7, Lysosome; 8, Endosomes; 9, Mitochondrion; 10, Microtubule cytoskeleton and 11, Glycosomes (Matthews, 2005).

1.3 Transmission of trypanosomes

Tsetse flies, found in sub-Saharan Africa, feed on mammalian blood and act as vectors of African trypanosomes. Tsetse flies belong to two groups: zoophilic (*Glossina palplis*) that prefers to feed on animal blood over human blood, and anthropophilic (*Glossina morsitans*) that has the opposite preference (Smith *et al.*, 1998). African trypanosomes cause mild infections in wild animals but may be acute or deadly in domesticated animals (Steverding, 2008). The transmission of trypanosomes to mammalian hosts depends on various factors, including the density of the tsetse fly population, tsetse flies' longetivity, the vector's

susceptibility to the infection, tsetse fly infestation rates, and the intensity of host-fly contacts (Franco et al., 2014).

Transmission of African trypanosomes is a cycle that revolves around the ingestion of contaminated blood by the tsetse fly from the mammalian host (Silvester et al., 2017). After ingestion, parasites divide in the tsetse fly's midgut, and during the next blood meal, parasites are injected into the next mammalian host, where they spread from blood to the internal organs and brain (Welburn et al., 2016). The most prevalent modes of trypanosome transmission are tsetse fly saliva (*T. brucei*) and insect feaces (*T. cruzi*) (Alizadehrad *et al.*, 2015). Other transmission routes rarely occur, such as transmission through contaminated needles, blood transfusion, sexual contact and placental transfer (Franco *et al.*, 2014).

Aquatic leeches are the primary vector of fish trypanosomes. They transmit trypanosomes to the fish hosts (e.g. *T. carassii*) via saliva (Overath *et al.*, 1998). However, isopod crustaceans and annelids can also transmit fish trypanosomes (Corrêa *et al.*, 2016). When leeches feed on infected fish, they ingest blood contaminated with parasites, and during the next blood meal, they transmit trypanosomes to the next fish host. Aquatic leeches act as both hosts and vectors of fish trypanosomes, and the fashion of transmission for both freshwater and marine trypanosomes is the same (Gupta, 2006; Lemos *et al.*, 2015).

1.4 Lifecycle of trypanosomes

African trypanosomes need both vertebrates (mammals) and invertebrates (tsetse flies) to complete their lifecycle and multiplication (Figure 1.4). Bloodstream form trypomastigotes are taken up in the bloodmeal from the mammalian host by the tsetse fly. The complexity and location of the developmental stages in the tsetse fly differs, depending on the trypanosome species. Whereas the *T. vivax* lifecycle stages are confined to the tsetse fly proboscis and cibarium, those of *T. congolense* occur successively in the midgut, foregut and proboscis, and *T. brucei*'s final transformation is in the salivary glands. In the tsetse fly, midgut *T. brucei* and *T. congolense* trypomastigotes transform into procyclics and multiply by binary fission. In *T. vivax,* this transformation occurs in the proventriculus (Peacock *et al.*, 2012; Rotureau and Van Den Abbeele, 2013). The *T. brucei* procyclics form short and long epimastigotes that move from the midgut to the foregut. Short epimastigotes move to and attach to the epithelium of salivary glands. Epimastigotes proliferate in the epithelium of the salivary glands and subsequently mature into metacyclics. The metacyclics migrate from the epithelium to salivary glands' lumen (Van Den Abbeele *et al.*, 1999; Kolev *et al.*, 2012). During this stage, the variant surface glycoproteins become ready to be activated after infecting the host. Tsetse flies transmit metacyclic forms of *T. brucei* during the blood meal to the mammalian host. In 2-4 weeks, post-infection, parasites change into actively dividing long slender forms and spread

through the mammalian host's blood and lymph. The level of parasitaemia keeps increasing due to a continuous, binary fission division of parasites until they reach a particular peak. Once a specific peak is reached, parasites stop dividing, and the slender forms transform into non-dividing stumpy forms. Tsetse flies become infected during the next blood meal of an infected mammalian host, and the cycle continues (Matthews *et al.*, 2015; Steverding, 2017).

The developmental stages of *T. congolense* only differ from that of *T. brucei* in the final stages, and its surface coat consists of carbohydrates instead of glycoproteins. The *T. congolense* procyclic trypomastigotes move from the midgut to the foregut, where they become elongated. These trypomastigotes migrate from the foregut to the proboscis, where they develop into epimastigote forms. Epimastigotes divide though the division does not occur asymmetrically as seen in *T. brucei* (Peacock *et al.*, 2012). The epimastigotes mature into the infective metacyclic form in the labrum and hypopharynx that are transmitted to the mammalian host during the tsetse fly's next blood meal (Eyford *et al.*, 2013; Rotureau and Van Den Abbeele, 2013). The development of *T. vivax* is quite different from that of *T. brucei* and *T. congolense* in that it begins and end in the proboscis instead of the midgut of the insect (Giordani *et al.*, 2016). That is the reason why *T. vivax* can be transmitted by different kinds of insects and is also found in northern Africa and South America (Jackson *et al.*, 2015).

Fish trypanosomes undergo a complex life cycle that involves distinct morphological types, and the life cycle takes place in two hosts, vertebrates (fish) and blood-sucking invertebrates (leeches) (Rodrigues *et al.*, 2018). *Hemisclepsis* and *Piscicola* transmit freshwater fish trypanosomes, while *Pontobdella* and *Trachobdella* transmit marine water fish trypanosomes (Kelly *et al.*, 2018). Fish trypanosomes (trypomastigotes) undergo a sequential developmental phase in the leech's digestive tract. The development begins with the formation of the flagellum, kinetoplast, extension of the body's posterior end, and synthesis of the undulating membrane. The nucleus begins to split into two, and the formation of two trypanosomes resulting in infectious forms. Division patterns of fish trypanosomes can occur in two distinct forms, e.g. *T. carassii* divide by transverse binary fission, whereas *T. trichogasteri* divide by longitudinal binary fission (Gupta, 2006). Inside the leech's intestine, the slender trypomastigote transforms through amastigote, promastigote and sphaeromastigote lifecycle stages into metacyclic trypanosomes that migrate into the proboscis and spread to the new host during the next blood meal (Hayes *et al.*, 2014).

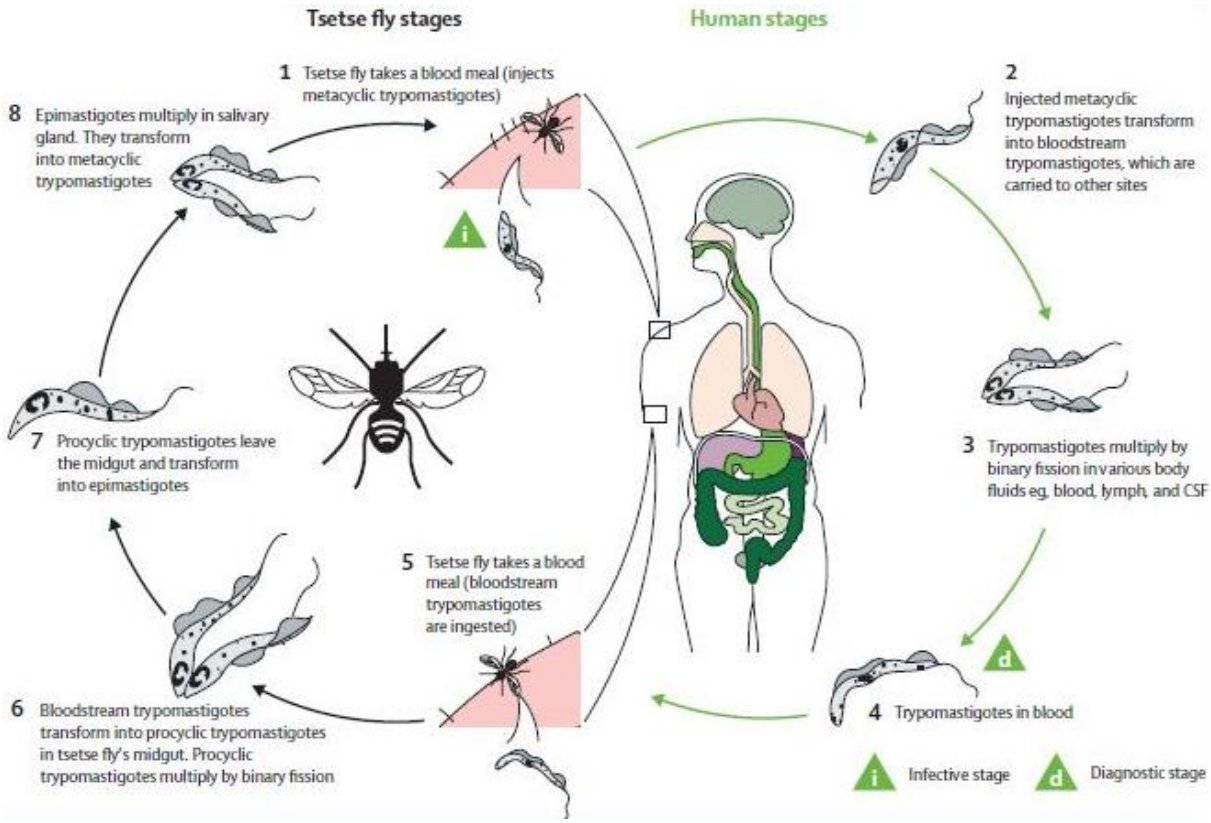

Figure 1.4: The lifecycle of African trypanosomes. A complete lifecycle of African trypanosomes shows different morphological stages between the tsetse fly and mammalian host (Kennedy, 2013).

1.5 Clinical symptoms

One of the common symptoms in African trypanosomiasis is anaemia. The disease shows different symptoms (acute and chronic forms) among different infected host species (Steverding, 2017). The acute forms occur immediately after the infection, and it is characterised by high parasitaemia followed by a rapid fall of packed cell volume (PCV). In the chronic stage, parasites persist for a long time, and the infected animals succumb to anaemia (Latif *et al.*, 2019). African animal trypanosomiasis is caused by *T. congolense, T. vivax* and *T. b brucei* (Steverding, 2008). In wild animals, these parasites cause relatively mild infections, whereas, in all domestic animals, the disease can be severe and result in decreased milk and livestock production, and as the disease progresses, the animals get progressively weaker. As a result, the animal loses physical function (Steverding, 2008; Steverding, 2017). In addition to nagana's clinical symptoms, including fever, listlessness, emaciation, hair loss, eye discharge, oedema, paralysis and miscarriage, the disease results in high mortality rates in domestic animals (Osório *et al.*, 2008; Steverding, 2017).

Human African trypanosomiasis caused by *T. b rhodesiense* is an acute disease that only lasts a few weeks if untreated, whereas the chronic disease caused by *T. b gambiense* lasts

for several months or years (Kennedy, 2013). Both forms of the disease are fatal if untreated. During the early haemolymphatic stage, the parasites multiply in the infected patient's bloodstream and lymphatic system. Symptoms of early-stages of HAT include fever, headache, weakness, anaemia, enlargement of the spleen, liver, endocrine dysfunction including menstrual abnormalities, fertility problems such as sterility, abortion and heart disease (MacLean *et al.*, 2012; Franco *et al.*, 2014; Steverding, 2017). During the late- or encephalitic stage, parasites cross the blood-brain barrier and invade the central nervous system leading to sleep disorders, involuntary movements, sensory disorders, seizures, and coma. Sleep disorder is the main symptom of late-stage of HAT. It includes night-time insomnia and day-time drowsiness, and frequent uncontrollable short sleep episodes (Steverding, 2017). In addition to the late-stage symptoms, the disease is associated with severe headache, and it changes the infected host's personality after several months or a year of post-infection (MacLean *et al.*, 2012).

Fish trypanosomiasis occurs in an acute and a chronic phase. The acute phase is characterised by an initial rise in parasitaemia in the blood and then a sharp decline. A year post infection, low levels of parasites can be found in the host blood, causing chronic disease (Overath *et al.*, 1999). Fish trypanosomes cause various blood-related disorders in the infected hosts, such as decreased serum protein concentration, leukocytosis, hypoglycaemia and hypocholesterolemia (Gupta, 2006; Gupta and Gupta, 2012). In addition to fish trypanosomiasis symptoms, parasites secrete lytic factors that lyse the fish red blood cells and result in anaemia in the absence of specific antibodies. Anaemia causes spleen and heart disorders in infected fish (Corrêa *et al.*, 2016). The severity of anaemia depends on the level of parasitaemia (Gupta, 2006). The infection alters erythrocytes' shape and decreases the host cells' lifespan, causing erythropenia (Rodrigues *et al.*, 2018). Infected salmon and goldfish lose appetite, and the disease results in abnormally low body weight (Islam and Woo, 1991; Corrêa *et al.*, 2016). Parasites utilise a high level of sugar from an infected host's blood and this result in a decline of the hosts' metabolic activity due to a complete depletion of the carbohydrates reserve. A rapid decrease in carbohydrate supply results in a severe liver burden (Gupta, 2006). This causes a high mortality rate in fish and negatively impact aquaculture production (Gupta and Gupta, 2012).

1.6 Host-parasite interaction

African trypanosomes exist extracellularly throughout their lifecycle in the host bloodstream and tissue and provoke sequential activation of the innate and adaptive immune response in mammalian hosts (Matthews *et al.*, 2015). Therefore, they must survive the innate and adaptive immune responses of their mammalian hosts to cause an infection (Matthews *et al.*,

2015; Cayla *et al.*, 2019). Parasites achieve this by expressing distinct VSGs connected to the parasite membrane by a glycosylphosphatidylinositol (GPI) anchor (Matthews, 2005). These VSGs of the slender bloodstream trypomastigotes are highly immunogenic. As soon as the host immune system detects parasites expressing VSGs, it starts producing antibodies, specifically against the VSG coat, facilitating parasite lysis (Matthews *et al.*, 2015; Ponte-Sucre, 2016). However, trypanosomes avoid detection by switching to a new VSG coat, and the next wave of infection is initiated (Figure 1.5). The previously raised antibodies are unable to detect new VSGs (Ponte-Sucre, 2016). The GPI anchor consists of glycosylated proteins forming a dense membrane, and VSGs serve as a shield mechanism that protects the parasites from the host immune response ensuring the infection (Matthews, 2005; Horn, 2014). The essential feature of *VSG* genes is that expression of only one gene occurs at any time, allowing the parasites to evade the host immune response continually (Kennedy, 2013; Matthews *et al.*, 2015). After uptake of stumpy trypomastigotes by the insect vector, the VSG coat is replaced by a procyclin coat and the VSG coat is acquired by the metacyclics in the salivary glands in preparation for transmission to their mammalian host (Matthews, 2005).

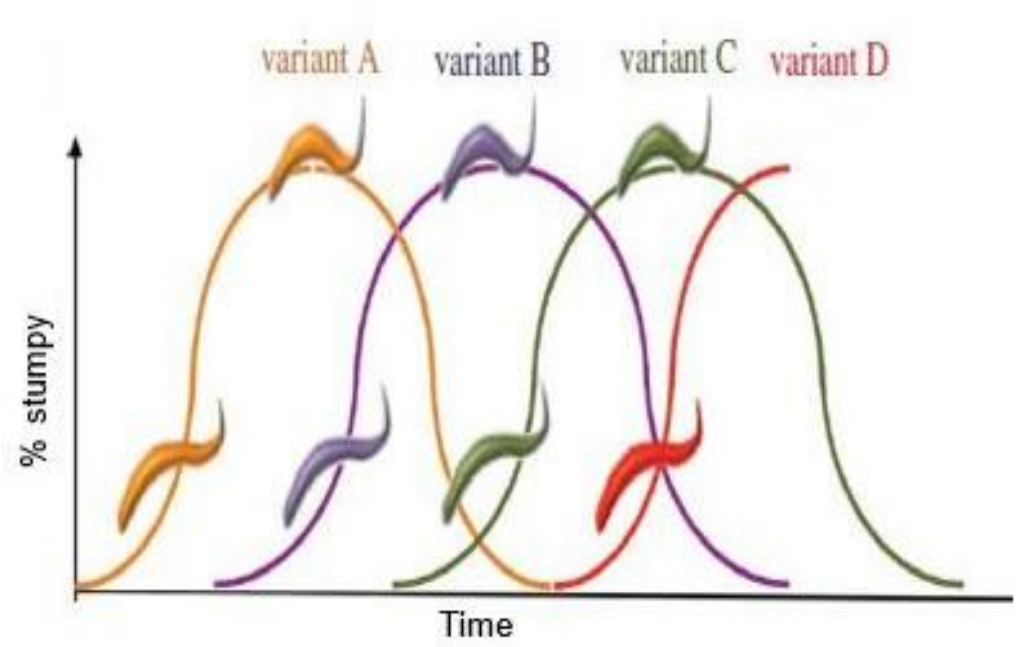

Figure 1.5: Host-parasite interaction of African trypanosomes. During the infection, *Trypanosoma brucei* antigenic variation frequently changes VSGs to evade the host immune system. As the number of proliferative slender forms increases, the host immune detects parasites. Parasites evade the host immune response by replacing the proliferative forms with the non-proliferative stumpy forms. Therefore, the wave of proliferative slender forms decreases (Matthews *et al.*, 2015).

Trypanosomes maintain their virulence and spread through quorum sensing (QS) in their host bloodstream. The proliferative slender forms multiply until they receive a signal to switch into

non-proliferative stumpy forms that are taken up by a vector (Rojas *et al.*, 2019). This QS is stimulated by the accumulation of soluble, low molecular weight, heat-stable factors known as stumpy induction factor (SIF) (Matthews *et al.*, 2015; Rojas *et al.*, 2019). Even though trypanosomes do not have G protein-coupled receptors (GPCRs), the signal is received by the GPR89 protein found on the surface of slender forms in the host bloodstream. It drives the formation of stumpy forms in response to the SIF signal (Rojas *et al.*, 2019). Stumpy forms can survive extreme pH and proteolytic attack, which helps them survive in the tsetse midgut and complete the parasites' cell cycle (Antoine-Moussiaux *et al.*, 2009; Matthews *et al.*, 2015).

The VSGs form a further important molecular interface between trypanosomes and the host immune system by inducing over-production of tumour necrosis factor-alpha (TNF-α) by macrophages (Antoine-Moussiaux *et al.*, 2009; Nok, 2009). The continuous release of the unregulated TNF-α from the activated macrophages results in the neurological manifestation of late-stage HAT that includes symptoms such as fever, asthenia, cachexia, and hypertriglyceridemia. Therefore, TNF-α plays a considerable role in the central nervous system (CNS) disorders of trypanosomiasis (Ponte-Sucre, 2016).

Fish trypanosomes are close relatives of African trypanosomes. There is little information about the fish host-trypanosome relationship, but some leeches are obligatory intermediate hosts of fish trypanosomes (Wiegertjes *et al.*, 2005). *Trypanosoma carassii* has been used as a model for studying the host-parasite relationship. Like salivarian trypanosomes, *T. carassii* lives extracellularly in the host blood and tissue fluids (Ruszczyk *et al.*, 2008; Oladiran and Belosevic, 2012). However, *T. carassii* does not express antigenic variation (Lischke *et al.*, 2000). Therefore, during the infections of *T. carassii*, parasites do not switch from slender to stumpy forms, as observed in other salivarian trypanosomes. Therefore, *T. carassii* apply a different approach to survive the host immune system. The surface coat of *T. carassii* is carbohydrate-rich, and it is connected to the parasite membrane by GPI residues (Lischke *et al.*, 2000; Wiegertjes *et al.*, 2005). The carbohydrate coat consists of highly glycosylated mucin molecules that are essential for parasite protection against the host immune response. Therefore, the surface coat of fish trypanosomes is similar to that of mammalian-infective *T. cruzi,* which have a GPI anchored mucin-like coat. This could mean that *T. carassii* share characteristics with both salivarian and non-salivarian trypanosomes (Lischke *et al.*, 2000; Overath *et al.*, 2001). Fish trypanosomes secrete haemolytic agents within the host blood and lyse the red blood cells; the rate of lysis depends on the level of parasitaemia in the host (Corrêa *et al.*, 2016). The increase in the number of parasites in fish blood leads to haemodilution (increased blood volume). Both haemolysis and haemodilution can cause death in the infected hosts (Islam and Woo, 1991).

1.7 Proteolytic enzymes in trypanosomes

Proteases are enzymes that hydrolyse peptide bonds and are described as exopeptidases or endopeptidases. Exopeptidases are responsible for the cleavage of peptide bonds at the amino- and carboxy-termini, and endopeptidases hydrolyse peptide bonds within the polypeptide. Most proteases are endopeptidases and are grouped into classes according to their catalytic mechanism: serine (EC 3.4.21), cysteine (EC 3.4.22), aspartic (EC 3.4.23), metallo (EC 3.4.24) and threonine (EC 3.4.35) peptidases (Rawlings, 2020). Proteases are present in all living organisms, and they are involved in different biological activities such as activation of other enzymes, modulation of the host organisms' immune response processes, cell differentiation, and autophagy. In parasites, proteases are involved in nutrient acquisition, ensure that parasites survive the host immune response and sustain infection (Caffrey and Steverding, 2009; Siqueira-Neto *et al.*, 2018). In trypanosomes, proteases play vital roles in the parasite lifecycle and in pathogenesis of trypanosomiasis. These include nutrition, immune evasion and intracellular survival (McKerrow, 2006; Steverding, 2013). Trypanosome proteases are therefore targeted for the development of new drugs against trypanosomiasis (Steverding, 2013).

The Schechter and Berger (1967) nomenclature is used to describe the interaction between a peptide substrate and the substrate binding pockets of a protease (Figure 1.6). The substrate amino acid residues N-terminal to the scissile bond (cleavage site) are numbered P1, P2, P3 etc. and those C-terminal to the scissile bond are numbered P’1, P’2, P’3 etc. The corresponding protease substrate binding pockets are numbered S1, S2, S3, etc. and S’1, S’2, S’3 etc.

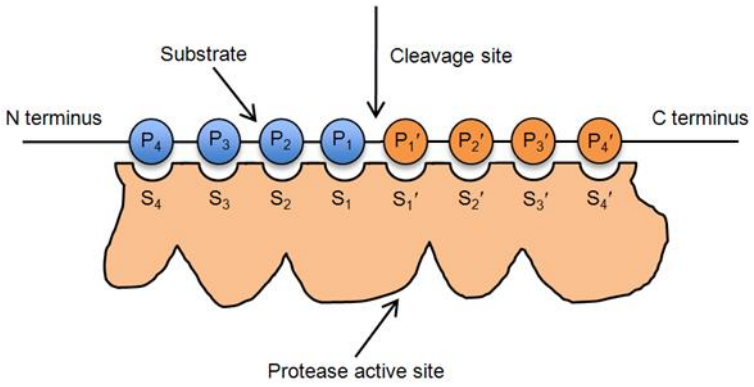

Figure 1.6: Schematic representation of the interaction between enzyme and substrate illustrating the Schechter and Berger (1967) nomenclature. Accessed from https://prosper.erc.monash.edu.au/methodology.html.

1.7.1 Cysteine proteases

Cysteine proteases consist of 72 different families, but not all are present in protozoan parasites, and cysteine proteases are divided into ten clans (CA, CD, CE, CF, CH, CL, CM, CN, CO and C- not allocated to any particular family. Clans are evolutionary related proteins (Caffrey and Steverding, 2009; Siqueira-Neto *et al.*, 2018). Clan CA represents all proteases with a papain-like fold (Yamamoto *et al.*, 2002). In kinetoplastids parasites, clan CA of the C1 family is the best-characterised cysteine proteases, and it consists of homologous proteases of mammalian cathepsins B, C, K, L and S (Caffrey and Steverding, 2009; Siqueira-Neto *et al.*, 2018). All cysteine proteases' catalytic activity depends on the cysteine thiol group that acts as the nucleophile and together with a histidine residue, form a catalytic triad with asparagine (Duschak and Couto, 2009; Erez *et al.*, 2009). Glutamine is the fourth residue that usually participates during catalysis (Yamamoto *et al.*, 2002). The histidine residue donates protons to stabilise the nucleophilicity of the cysteine residue (Figure 1.7) (Erez *et al.*, 2009). The cysteine residue attacks the carbon of the reactive peptide bond, thus, yields the first tetrahedral thioester intermediate and releasing an amine or amino terminus fragment of the substrate. The glutamine residue stabilises the hydrogen bond between the intermediate and the substrate oxyanion. After that, the thioester bond is broken down by hydrolysis, resulting in carboxylic acid moiety production from the remaining substrate fragment (Verma *et al.*, 2016; Siqueira-Neto *et al.*, 2018).

To activate cysteine proteases requires removing the propeptide, either by autoproteolytic activation or by another protease (Yamamoto *et al.*, 2002). The prodomain plays a significant role in folding the new proteins, maintaining the peptidase inactive, and stabilising the enzyme against denaturing environments (Yamamoto *et al.*, 2002; Kerr *et al.*, 2010).

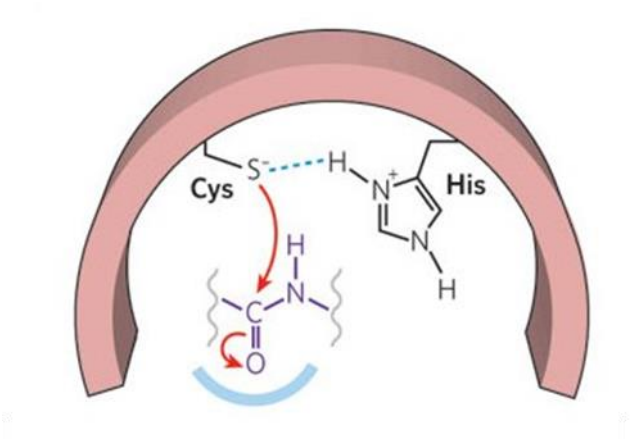

Figure 1.7: The reaction mechanism of cysteine proteases. The catalytic dyad consists of Cys and His catalytic residues whereby the Cys nucleophile attacks the substrate carbonyl carbon (Erez *et al.*, 2009).

Cysteine proteases have been identified and characterised in the African trypanosomes, *T. b. brucei*, *T. b. rhodesiense, T. congolense* and *T. vivax* (Mendoza-Palomares *et al.*, 2008; Eyssen *et al.*, 2018). The cathepsin B and L of *T. brucei* are called *Tb*CATB and *Tb*CATL, respectively (Steverding, 2013). The *Tb*CATB and *Tb*CATL function in the pathological processes of sleeping sickness. The *Tb*CATL is predominantly in the bloodstream forms and is involved in cell differentiation, and it participates in cell regulatory processes that allow parasites to penetrate the brain barrier by activating host G-protein coupled receptors protease-activated receptor 2 (PAR2), followed by triggering the host calcium-signalling pathways, resulting in brain disorders (Steverding, 2013; Siqueira-Neto *et al.*, 2018). The *Tb*CATL protects *T. brucei* from being lysed by host serum by hydrolysing pore-forming APOL1 in human serum and plays a significant role in the hydrolysis of endocytosed proteins in the host bloodstream (Steverding, 2017; Siqueira-Neto *et al.*, 2018). *Tb*CATB function in the hydrolysis of transferrin and iron acquisition pathways of the parasites. Transferrin is a source of iron for *T. brucei* from the host's lysosomal or endosomal compartments (O'Brien et al., 2008).

In *T. congolense* cathepsin B- and L-like proteases, known as *Tco*CATB and *Tco*CATL, participate in the pathogenesis of nagana (Mendoza-Palomares *et al.*, 2008; Boulangé *et al.*, 2011). The *Tco*CATL plays a significant role in anaemia and immunosuppression during the infection. (Lalmanach *et al.*, 2002). Both *Tco*CATB and *Tco*CATL are present in developing bloodstream forms and are essential in protecting the parasite against the host immune system (Mbawa *et al.*, 1992).

The cathepsin L-like cysteine protease of *T. cruzi* is called cruzipain or *Tcr*CATL (Klemba and Goldberg, 2002). The protease is produced in an inactive form and is autocatalytically

activated by hydrolytic cleavage of the N-terminal prodomain, like all cathepsin L-like proteases and it is involved in the invasion of host cells, nutrient metabolism and evasion of the host immune response by hydrolysing host IgG (Dos Reis *et al.*, 2006; McKerrow, 2006). Cathepsin B-like proteinase of *T. cruzi* is responsible for signal transduction and metacyclogenesis (differentiation of non-pathogenic epimastigotes into pathogenic metacyclic trypomastigotes) in *T. cruzi* (McKerrow, 2006; Siqueira-Neto *et al.*, 2018).

The *T. carassii* cathepsin L is one of the essential molecules contributing to fish trypanosomiasis pathogenicity by host tissue invasion and by degrading host proteins and red blood cells, resulting in blood-related disorders in the fish hosts (Lemos *et al.*, 2015). Apart from involvement in pathogenicity, the cysteine protease also plays a crucial role in metabolic processes and cell division of parasites (Overath *et al.*, 2001; Ruszczyk *et al.*, 2008). Therefore, cysteine proteases can be used as targets for developing vaccines and drugs for piscine trypanosomes (Ruszczyk *et al.*, 2008).

African animal trypanosomes are becoming increasingly drug resistant and cysteine protease have been identified as potential drug targets (Mendoza-Palomares *et al.*, 2008). An improved understanding of the biochemical characteristics of cysteine proteases found in protozoan parasites is thus essential as it gives an idea about how enzymes facilitate a substrate's hydrolysis, therefore can be used in developing drugs or improving the already existing drugs against these pathogens (Mendoza-Palomares *et al.*, 2008; Roberta *et al.*, 2016; Siqueira-Neto *et al.*, 2018). The vinyl sulfone inhibitor, K11777 of cysteine proteases is a promising drug against *T. cruzi* and *T. brucei* since it targets the cathepsin L-like enzyme of *T. brucei* when combined with eflornithine (Siqueira-Neto *et al.*, 2018).

1.7.2 Other proteases

In addition to the Clan CA cysteine proteases (Section 1.7.1) trypanosomes also express metacaspases (MCA) that belong to family C14, Clan CD of cysteine proteases. These proteases have a caspase-like structure and a similar active site catalytic dyad comprising cysteine and histidine, but unlike caspases that hydrolyse peptides C-terminal to Asp residues, metacaspases require Arg or Lys in the P1 position (Szallies *et al.*, 2002; Tsiatsiani *et al.*, 2011; Minina *et al.*, 2017). Metacaspases (MCAs) are attractive drug targets for the treatment of African trypanosomiasis as they are not found in the metazoan kingdom and their action has been implicated in cell cycle and cell death pathways in kinetoplastid parasites (Mottram *et al.*, 2003). *Trypanosoma brucei* encodes five metacaspases (*Tb*MCA1-*Tb*MCA5) of which *Tb*MCA2, *Tb*MCA3 and *Tb*MCA5 are active due to the presence of a Cys-His catalytic dyad (McLuskey *et al.*, 2012). The MCA5 from *T. congolense* is able to autoprocess into an active form that hydrolyse host proteins such as gelatin (Eyssen and Coetzer, 2019).

The main serine proteases in trypanosomes belong to the S9 family of prolyl oligopeptidases and include oligopeptidase B (OPB, E.C. 3.4.21.83) that hydrolyses peptides smaller than 30 residues C-terminal to Arg residues and prolyl oligopeptidase (POP, EC 3.4.21.26) that require a Pro residue in P1. Both OPB and POP have been identified in African trypanosomes (*T. brucei*, *T. congolense*, *T. vivax* and *T. evansi*) and in South American *T. cruzi* and shown to hydrolyse host peptide hormones in the bloodstream, resulting in disturbing host endocrine rhythms (Morty *et al.*, 1999; Coetzer *et al.*, 2008; Bastos *et al.*, 2013). In *Tb*OPB double knockout strains, increased levels of *Tb*POP were observed showing that the trypanosomes have back-up systems to ensure host peptide hormone hydrolysis (Kangethe *et al.*, 2012). In *T. cruzi* trypomastigotes and metacyclics, *Tcr*OPB mediates host cell invasion by triggering an increase in the free intracellular calcium concentration in non-phagocytic cells (Burleigh and Andrews, 1995).

The metalloproteases of trypanosomes are Zn^{2+}-dependent endopeptidases, known as glycoprotein 63 (Gp63) or major surface proteases (MSPs). *Trypanosoma brucei* expresses three MSP families, *Tb*MSP-A, -B, and -C, with *Tb*MSP-B only present in the procyclic and bloodstream forms, while *Tb*MSP-A and *Tb*MSP-C are only present in the bloodstream forms (Lacount *et al.*, 2003; Moreno *et al.*, 2019).

1.8 Diagnosis of African trypanosomiasis

The ability of African trypanosomes to avoid the host immune system by periodically switching VSGs (antigenic variation) and the toxicity of the already existing drugs (Melarsoprol) are challenges for controlling the disease (Chappuis *et al.*, 2005). Disease control depends mainly on detecting a new case, determining the stage, treatment, and vector control. The diagnosis process of *T. b. gambiense* involves three steps screening, diagnosis, confirmation and staging (Chappuis *et al.*, 2005; Mitashi *et al.*, 2012). Screening can either be active that refers to professionals going directly to the communities to examine the whole population in areas affected by HAT or passively, meaning individuals from the HAT affected areas seek help from health care infrastructures (Radwanska, 2010).

Serological tests follow the active and passive screening processes for HAT. Possible tests are Card Agglutination Tests for Trypanosomiasis (CATT/*T. b gambiense*), Latex/*T. b. gambiense*, immune trypanolysis tests, enzyme-linked immunosorbent assay (ELISA) and immunofluorescence antibody test (IFAT) (Chappuis *et al.*, 2005; Mitashi *et al.*, 2012). The CATT is inexpensive, quick and easy to use, the results can be visualised after 5 min and can be used for screening populations at high risk (Büscher *et al.*, 2017). The agglutination assay for detecting *T. b gambiense*-specific antibodies present in the blood, plasma or serum of *T. b gambiense* infected hosts is the most preferred serological test, especially for large

populations human trypanosome infected areas (Chappuis *et al.*, 2005; Radwanska, 2010). The antigen used in the CATT is complete bloodstream form *T. b gambiense* variable antigen type LiTAt 1.3 (Radwanska, 2010; Büscher *et al.*, 2017). The CATT makes use of Coomassie blue-stained, formalin-fixed freeze-dried trypanosomes (Mitashi *et al.*, 2012). Patients infected with a strain of trypanosomes that does not express the LiTat 1.3 gene may give false negative CATT results. The CLN palpation and puncture subsequently follows the positive results from the CATT tests. The CLN is punctured, and fresh aspirate is spread on a slide, and the coverslip is applied to spread the sample and immediately examined by a microscope (magnification x400) to detect motile trypanosomes (Chappuis *et al.*, 2005; Kennedy, 2013).

The LATEX/*T. b. gambiense* is another serological test that may be used instead of CATT, and it uses a mixture of three purified *T. b. gambiense* variable surface antigens, LiTat 1.3, -1.5 and -1.6, coupled to suspended latex particles (Penchenier *et al.*, 2003; Chappuis *et al.*, 2005). The specific antibodies in the patient blood agglutinate the antigen-coupled latex particles. The results of LATEX/*T. b gambiense* is easier to interpret than those of CATT (Penchenier *et al.*, 2003). However, pipettes, microtiter plates, cold chain and trained technicians are equired for the LATEX test. Due to these requirements, it is not preferred to be used at the health centre level (Mitashi *et al.*, 2012). Therefore, new diagnostic techniques for African trypanosomes are required for the complete elimination of trypanosomiasis (Chappuis *et al.*, 2005).

The pathogenicity or animal trypanosomes (*T. congolense*, *T. evansi*, *T. vivax* and *T. brucei*) vary among species and symptoms alone do not present definitive diagnoses. Attempts to control these infections proves to be ineffective, due to vector control being expensive and parasites develop resistance to the existing animal trypanocidal drugs. In addition to factors that contribute to failed attempts of controlling animal African trypanosomiasis and a lack of veterinary professionals, farmers rely on themselves to make diagnoses, which may lead to misdiagnosis and incorrect drug administration. The development of new and user friendly diagnostic tests is important for successful treatment and correct use of the existing trypanocidal drugs (Boulangé *et al.*, 2017).

Diagnosis of AAT can be carried out by detection of the parasite or its constituents (antigen or DNA) and detection of the antibodies it induces. Parasite detection methods use microhaematocrit centrifugation technique, which is rapid and easy to perform. However, the set back about this technique is that it has low sensitivity compared to molecular detection such as polymerase chain reaction (PCR) or Loop-mediated isothermal amplification (LAMP). The disadvantages of PCR are that it requires expensive equipment and trained technicians (Boulangé *et al.*, 2017). Serological tests for infection and diagnosis, based on antibody-

enzyme-linked immunosorbent assay (Ab-ELISA) have been developed for animal trypanosomes (Desquesnes *et al.*, 2001). The advantages of these tests, are that they can be used for large-scale samples, are highly sensitive, reproducible and use less expensive equipment compared to PCR (Desquesnes *et al.*, 2001; Boulangé *et al.*, 2017). The ELISAs are carried out in the laboratory on serum, plasma, or dried blood spot samples. They are used in large-scale sampling, because of their high specificity and sensitivity. They can also be used in monitoring post elimination diagnosis, and animal reservoir studies (Büscher *et al.*, 2017). The disadvantages of ELISAs, are that detection of antibody uses trypanosome lysate and it is difficult to obtain trypanosome lysate due to different preparation protocols, strains, preservation conditions, and even propagation methods (culture or animal propagation) and none of these methods are easy to use in *T. vivax* (Boulangé *et al.*, 2017). As a result, specific protein antigens that can be recombinantly expressed have been evaluated for their diagnostic potential (Pillay *et al.*, 2013; Boulangé *et al.*, 2017; Eyssen *et al.*, 2018).

1.9 Rationale

Trypanosomes infect all vertebrates, including humans, animals, birds and fish. Tsetse flies transmit African trypanosomes, and leeches transmit fish trypanosomes (Grybchuk-Ieremenko *et al.*, 2014). In sub-Saharan Africa, *Trypanosoma* species cause a devastating disease in humans (HAT) and animals (AAT). *Trypanosoma b. gambiense* causes a chronic form of HAT in western and central Africa, while *T. b. rhodesiense* is responsible for an acute form in the eastern parts of the continent. Animal AT or nagana is caused by *T. congolense* and *T. vivax*, while *T. b. brucei* infection of animals is asymptomatic, making the parasite a good model organism for studies of HAT and AAT. Although the number of cases of HAT has declined significantly as a result of effective diagnosis and chemotherapy, AAT continues to cause severe economic losses in livestock production (Büscher *et al.*, 2017). African trypanosomes live extracellularly in their host bloodstream and tissue fluids. Inside the mammalian hosts, trypanosomes frequently switch to different VSGs, evading the host immune system (Matthews *et al.*, 2015). The periodic expression of different VSGs prolongs the infection, sustains parasites within the host for a long time, and contributes to drug resistance. Previously the attempts to eliminate trypanosomes in Africa was based on vector control. However, this method proves to be inefficient because it is expensive and challenging to reach remote tsetse fly-infested areas. Trypanosomes affect people and their livestock living in poor rural areas in Africa. The available drugs for controlling animal African trypanosomiasis are either expensive, toxic or ineffective and drug resistance is on the increase. In addition to factors that contribute to the virulence of trypanosomes are proteases. Proteases play different roles in parasites' biological process, including cell differentiation and hydrolysis of host

proteins. Proteases have been identified as promising diagnostic and drug targets for animal and human African trypanosomiasis.

Mammalian trypanosomes have been the main subject in research, while fish trypanosomes are being neglected. Even though they affect aquaculture production, the information about fish trypanosomes' molecular biology is very scarce. None of the fish trypanosomes genomes has been sequenced to date. In the past, the taxonomy of trypanosomes was based on the nucleus and kinetoplast position and the host specificity (Grybchuk-Ieremenko *et al.*, 2014). Results obtained from these features were unreliable because the nucleus and kinetoplast position may change as parasites grow. A single species of the *Trypanosoma* parasite may infect more than one host. Biomarker genes have been identified as a promising solution to this issue. Various molecular markers such as genes encoding 12S rRNA, 18S rRNA, and glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) prove to give more reliable phylogenies of trypanosomes. The 18S rRNA gene is far more suitable for constructing a tree with better resolution. Therefore, it is widely used in phylogeny and shows that fish trypanosomes are monophyletic.

As previously stated, trypanosome proteases are pathogenic factors, and a better understanding of their enzymatic characteristics will inform drug development. A cathepsin L-like cysteine protease of *T. b. brucei* is a potential drug target, and enzymatic characterisation will inform drug development. Regarding trypanosomes that potentially infect a local freshwater fish, *Clarias gariepinus,* isolation of genomic DNA from blood samples for amplification of a trypanosome biomarker gene, followed by sequencing of the amplicons, may be used in phylogenetic studies to determine the diversity of trypanosomes infecting these species.

1.10 Aims and objectives of the present study

The first aim of the present study was to determine the enzymatic characteristics of a cathepsin L-like cysteine protease from *T. b. brucei*, *Tb*CATL, which has been identified as a potential drug target in trypanosome infections. The second aim was to amplify trypanosomes in blood samples of freshwater fish (*Clarias gariepinus*) from the Vaal River and to investigate the genetic diversity of the trypanosomes using phylogenetic analyses.

Aim 1: To recombinantly express, purify and enzymatically characterise the *T. b. brucei* cathepsin L-like protease, *Tb*CATL

The first aim was to clone the gene coding for *Tb*CATL from genomic DNA for recombinant expression and purification, so that enzymatic characterisation studies could be undertaken. Aim 1 can be divided into the following objectives:

Objective 1: Amplify the gene coding for the T. b brucei cathepsin L-like cysteine protease (TbCATL) from genomic DNA, clone the gene into a pGEM®-T Easy vector and identify recombinants.

The gene coding for *Tb*CATL will be identified in GenBank, and primers will be designed to amplify the gene using genomic DNA, which will be isolated from *T. b. brucei* strain 927 as the template. The amplicon will be cloned into a T-vector for transformation into *E. coli* JM109 cells to perform blue-white colony screening to identify recombinant plasmids. Plasmid DNA will be isolated from white colonies, and colony PCR performed to confirm the presence of the DNA insert.

Objective 2: Subclone the gene into pET28a and pET32a expression vectors.

White colonies containing the DNA insert will be grown up, plasmid DNA isolated, and restriction digests performed to obtain the DNA inserts following agarose gel electrophoresis and excision of the inserts for subcloning into the two expression vectors. Nucleic acid sequencing will be used to ensure that the inserts are mutation-free before ligation into expression vectors and transformation into *E. coli* BL21 (DE3) cells.

Objective 3: Recombinantly express TbCATL using the E. coli BL21 (DE3) expression system.

Induction of expression of *Tb*CATL from the expression vector and optimise the temperature for expression. Should *Tb*CATL be expressed in insoluble inclusion bodies, sarkosyl solubilisation will be performed.

Objective 4: Purify the His-tagged TbCATL using Co^{2+} immobilised metal affinity chromatography (IMAC).

Should sarkosyl solubilisation *Tb*CATL be required, the His-tagged protein will be refolded and purified using Co^{2+}-IMAC. The concentration of the purified protein will be determined using dye-binding assays and purity assessed using sodium dodecyl sulfate gel electrophoresis (SDS-PAGE).

Objective 5: Conduct enzymatic assays on TbCATL to determine substrate specificity, optimal pH conditions and inhibitor susceptibility.

The activity of *Tb*CATL against the protein substrate, gelatin, will be evaluated using gelatin-substrate SDS-PAGE. This assay will also be used to determine the pH optimum for protein digestion and the effect of protease class-specific inhibitors on gelatin hydrolysis. Fluorescently labelled peptide substrates, which differ in the residues in the P1-P3 positions, will be used to determine catalytic specificity by measuring k_{cat} K_m, V_{max} and k_{cat} values. The

best peptide substrate will be used to determine the effect of protease class-specific inhibitors on peptidolytic activity.

Aim 2: To identify trypanosomes in genomic DNA isolated from *Clarias gariepinus* blood samples using SSU rRNA gene amplification and perform phylogenetic analysis

The second aim was to screen for the gene coding for *Trypanosoma* SSU rRNA from *Clarias gariepinus* blood samples and sequence the genes so that phylogenetic studies could be undertaken.

Aim 2 can be divided into the following objectives:

Objective 1: Amplify the SSU rRNA genes from Clarias gariepinus trypanosomes to generate sequences.

The gene coding for SSU rRNA will be amplified, using primers available in the literature (Sehgal *et al.*, 2001) to amplify the gene from *Clarias gariepinus* blood samples. The amplicons will be sent for Sanger sequencing.

Objective 2: Align the SSU rRNA sequences from Clarias gariepinus trypanosomes with the SSU rRNA from GenBank fish trypanosomes.

The generated sequences will be confirmed as fish trypanosomes through BLAST against GenBank, and these will be combined with all SSU rRNA sequences matching fish trypanosomes from GenBank.

Objective 3: Estimate best-fit substitution models using Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc) in JmodelTest for maximum likelihood (ML) and Bayesian inference (BI).

The alignment will be used to estimate the best-fit substitution models for maximum likelihood (ML) and Bayesian inference (BI) using Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc) in JmodelTest. The most likely tree, will be estimated using Garli 2.0 (Zwickl, 2006). Bayesian analysis will be performed using 20 million generations using MrBayes version 3 (Ronquist and Huelsenbeck, 2003) on CIPRES.

2 CHAPTER 2: MATERIALS AND METHODS

2.1 Identification of *T. b brucei* CATL-like protease (*Tb*CATL) ORF and designing primers to amplify the *TbCATL* gene

The polymerase chain reaction (PCR) is used to obtain multiple copies of a DNA fragment by replication of the double-stranded DNA template using short single-stranded DNA sequences (primers) complementary to the flanking regions of the DNA to be amplified. The amplification process comprise three steps: denaturation, hybridisation and elongation (Mullis and Faloona, 1987). During the denaturation step the two strands of the double-stranded DNA are separated by increasing the temperature to around 94 °C. This is followed by the hybridisation step where the primers anneal with the complementary single stranded DNA when the temperature is decreased to between 40 and 70 °C, allowing the inter-base hydrogen bonds to form. The complementary DNA strand is synthesised by *Taq* polymerase during the elongation step, using deoxyribonucleic acid bases (dATP, dCTP, dGTP, dTTP) at around 72 °C. *Thermus aqauticus* (*Taq*) polymerase is heat stable (Brock and Freeze, 1969) and allows the repeated cycles of denaturation, hybridisation and elongation for the amplification of the target gene.

The DNA sequence coding for the CATL-like protein from *T. b. brucei* strain 927 (GenBank accession: X16465.1) was selected to design primers using primer 3 (Rozen and Skaletsky, 2000) that amplified the 921 bp gene. BamHI and XhoI restriction sites [for nomenclature, see Roberts *et al.* (2003)] were added to the forward and reverse primers, respectively, to allow subsequent subcloning into expression vectors. Primers used for PCR amplification of the *TbCATL* gene and colony PCR of the recombinant cloning and expression plasmids were synthesised at Inqaba Biotec, are shown in Table 2.1. The genomic DNA isolated as described in Section 2.2 was used as a PCR template. The final concentrations used in the PCR master mix were: 0.6 µM of each gene primer, 1 x *Ex Taq* buffer (Takara Bio Inc, Shiga, Japan), 1 U *Takara Ex Taq* HS polymerase (Takara Bio Inc, Shiga, Japan) and 2.5 mM dNTPs (Sigma Aldrich, St. Louis, Missouri, USA) in a total reaction volume of 50 µL. The PCR amplification of the *TbCATL* gene was carried out as follows: incubation at 95 °C for 3 min as the initial DNA denaturation step, followed by 30 cycles of 95 °C for 30 s, 59 °C for 30 s and 72 °C for 1 min. A final elongation step was carried out at 72 °C for 7 min. The amplified PCR product (3 µL) was added to 1 µL loading dye and electrophoresed on a 1% agarose gel, containing 10 µg/mL SYBR Safe in 1 x TAE buffer, as described in Section 2.3. The remaining PCR product was purified using the ZymoResearch Clean and Concentrator™ kit (Orange, CA, USA) as per the manufacturer's protocol.

Table 2.1: Primer sequence used for amplification and cloning process of *TbCATL.*

Primer name	Primer sequence 5' to 3'	Basic Tm^{o}
TbCATL forward	TT GGA TCC C**TC C**AC GTG GAG GAG TCA TTG	56 ^{o}C
TbCATL reverse	TT CTC GAG T**CA T**GC GGA GGA TAC GGC CTG	55 ^{o}C

The underlined sequence shows restriction sites.
The sequence in bold shows start and stop codons, respectively.

2.2 Isolation of genomic DNA from *Trypanosoma brucei brucei*

The genomic DNA of *T b. brucei* strain 927 was isolated using a Qiagen DNeasy Blood & Tissue Kit as per the manufacturer's protocol (Qiagen, Hilden, Germany). Briefly, a trypanosome culture pellet, previously cultured in the lab, was thawed in a microcentrifuge tube, centrifuged (11000 x *g*, 10 min, RT), and the supernatant was discarded. The cell pellet was resuspended with 20 µL proteinase K (Qiagen, Hilden, Germany) and the volume adjusted to 220 µL with PBS [100 mM Na_2HPO_4, 2mM KH_2PO_4, 2.7 mM KCl and 137 mM NaCl, pH 7.4 (1 mL)]. Lysis buffer AL (200 µL) was added into the cell suspension and mixed thoroughly by vortexing and incubated at 56 ^{o}C for 10 min. Ethanol (100% (v/v), 200 µL) was added to the sample and mixed thoroughly by vortexing. The resulting mixture was transferred into the DNeasy Mini spin column, placed in a 2 mL collection tube, and centrifuged (6000 x *g* for 1 min, RT). The flow-through and the collection tubes were discarded. The DNeasy Mini spin column was placed into a new 2 mL collection tube, followed by the addition of 500 µL wash buffer AW1 and centrifuged (6000 x *g*, 1 min, RT). The flow-through and the collection tubes were again discarded. The DNeasy Mini spin column was placed into a new 1.5 mL microcentrifuge tube, and 100 µL elution buffer AE was added directly into the DNeasy membrane, incubated at RT for 1 min centrifuged (6000 x *g,* 1 min, RT) to elute the DNA. The isolated DNA sample (4 µL) was added to (1 µL) of a gel loading dye (Solis BioDyne, Tartu, Estonia) and electrophoresed on a 1% (w/v) agarose gel containing 10 µg/mL SYBR Safe (Thermo Fisher Scientific, Waltham, MA USA), in 1xTris-Acetate-EDTA (TAE) buffer, as described in Section 2.3.

2.3 Agarose gel electrophoresis analysis of DNA

Agarose gel electrophoresis, which involves the migration of ions in an electric field, is used for analysing the purity and size of DNA or RNA fragments. When placed in an electric field, charged nucleic acid molecules are separated based on their sizes, from a few hundred base pairs to several thousand base pairs in length. Nucleic acid molecules carry a negative charge and migrate towards the positively charged anode. Agarose forms a complex reticular network

when dissolved and cooled, thus forming the sieving gel matrix. The smaller molecular weight molecules migrate faster than the larger molecules. DNA fragments are visualised with the aid of a dye such as ethidium bromide or SYBR Safe that intercalates between DNA bases and hence does not affect the size of DNA fragments (Davis *et al.*, 1986). Since ethidium bromide is a potential carcinogen, safer nucleic acid stains such as SYBR Safe was developed that has the added advantage that a blue-light transmitter can be used for visualisation, rather than UV light, when the DNA amplicons need to be excised from the agarose gel for cloning.

A 1% (w/v) agarose gel was used to analyse DNA samples obtained from DNA isolation or PCR products. Briefly, 0.4 g Seakem®LE agarose (Lonza, Rockland, ME, USA) was dissolved in 40 mL 1 x Tris-Acetate-EDTA (TAE) buffer (40 mM Tris-HCl buffer, 100 mM acetic acid, 1 mM Na_2EDTA, pH 8.0), and the agarose was completely dissolved by heating. Once the solution was cooled to approximately 50 ^oC, CYBR Safe (Thermo Fisher Scientific, Waltham, MA USA), was added to give a final concentration of 10 µg/mL, and the solution was poured into the casting tray and allowed to set. Once the gel was set, it was submerged in an electrophoresis tank containing 1 x TAE buffer. DNA samples were prepared by the addition of 1 µL loading dye. The DNA was electrophoresed at 90 V for approximately 1 h to separate DNA bands into various sizes. After electrophoresis, the DNA was visualised by fluorescence of CYBR Safe dye.

2.4 Ligation of the insert into T-vector

The *TbCATL* amplicon was ligated into the pGEM®-T Easy cloning vector (Figure 2.1) (Promega, Madison, WI, USA) using a 3:1 ratio of vector to purified PCR product. A reaction volume of 15 µL was used containing 3 µL insert, 1 µL pGEM®-T Easy, 1.5 µL 10 x ligation buffer (Promega, Madison, WI, USA), 8.5 µL nuclease-free water, and 1 µL (1 U) T4 DNA ligase (Promega, Madison, WI, USA), followed by incubation of the ligation mixture at 4 ^oC for 16 h.

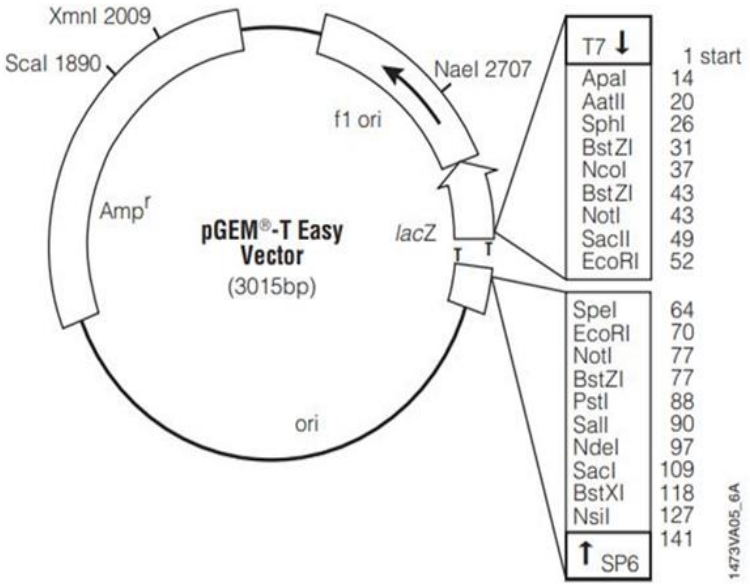

Figure 2.1: Map for the pGEM®-T Easy vector. The vector has multiple cloning sites, the origin of replication (ori), ampicillin resistance (Ampr) used for selection of recombinant colonies and the lacZ gene for blue/white colony screening (Promega technical manual).

2.5 Transformation of *E. coli* JM 109 cells with the ligation mixture

The pGEM®-T Easy vector - *TbCATL* insert ligation mixture was transformed into competent *E. coli* JM 109 cells (New England Biolabs, Ipswitch, MA, USA), using the TransformAid™ Bacterial transformation Kit (Fermentas, Vilnius, Lithuania) following the manufacturer's protocol. The resulting *E. coli* JM 109 cells (50 µl) from the transformation mixture were plated on pre-warmed 2 x YT plates [1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl, 1% (w/v) bacteriological agar (Merck Biolab, Darmstadt, Germany)] containing 100 µg/mL ampicillin (USB Corporation, Cleveland, OH, USA), 20 µg/mL isopropyl-β-D-thiogalactopyranoside (IPTG), 10 µg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (Fermentas, Vilnius, Lithuania) and incubated at 37 ºC for 16 h, according to the manufacturer's protocol. The addition of X-gal onto the 2 x YT plates allows for blue/white colony screening due to alpha complementation of the β-galactosidase gene of the pGEM®-T Easy cloning vector. The ligation of the *TbCATL* insert into pGEM®-T Easy cloning vector (Figure 2.1) results in the disruption of the *lacZ* gene. Non-recombinant colonies were blue since the *lacZ* gene was not disrupted. Therefore, they were able to produce β-galactosidase that utilise X-gal synthetic substrate. Recombinant colonies were white since they were unable to utilise the synthetic substrate.

2.6 Colony PCR

After transforming *E. coli* JM 109 cells with pGEM®-T Easy-containing *TbCATL*, 2 x YT agar

plates containing X-gal, IPTG and ampicillin were examined for the presence of blue and white colonies. Plasmid DNA was isolated from the white recombinant colonies using the NucleoSpin® Plasmid kit (Macherey Nagel, Duren, Germany) according to the manufacturer's protocol. The isolated DNA was used as the template for colony PCR. The reaction mixture contained gene primers synthesised at Inqaba Biotec, are shown in Table 2.1, 1 U FIREpol® Taq polymerase (Solis BioDyne, Tartu, Estonia), 10 x PCR reaction buffer B (Solis BioDyne, Tartu, Estonia), 25 mM $MgCl_2$ and 10 µm dNTPs (Solis BioDyne, Tartu, Estonia). The PCR reaction was performed in a total volume of 25 µL with reaction conditions, as described in Section 2.1. A sample of 5 µL of the colony PCR products was added into 1 µL loading dye and electrophoresed on 1% (w/v) agarose gel containing 10 µg/mL CYBR Safe in 1 x TAE buffer, as described in section 2.3.

2.7 Restriction digestion of insert and plasmid

White colonies containing the recombinant pGEM®-T Easy-*TbCATL* plasmid were selected and grown in 5 mL 2 x YT liquid medium containing 100 µg/mL ampicillin at 37 °C for 16 h, with agitation. The plasmid DNA was isolated from the overnight cultures using the NucleoSpin® Plasmid DNA isolation kit (Macherey Nagel, Germany). The presence of the *TbCATL* insert in the pGEM®-T Easy vector was assessed by restriction digestion of the plasmid with BamHI and XhoI restriction enzymes (ANZA products, Thermo Fisher Scientific, Waltham, MA USA), according to the manufacturer's protocol. Briefly, a total volume of 50 µL restriction digestion mixture containing 6.1 µL pGEM®-T Easy *TbCATL*, 5 µL 10 x buffer (Thermo Fisher Scientific, Waltham, MA USA), 1 µL ANZA BamHI, 1 µL ANZA XhoI and 36.9 µL nuclease-free water was incubated at 37 °C for 2 h. The reaction mixture (50 µL) was added to (5 µL) loading dye and electrophoresed on a 1% (w/v) agarose gel in 1 x TAE buffer, containing 10 µg/mL SYBR Safe dye, as described in Section 2.3. The SYBR Safe dye was used to visualise the DNA bands (instead of ethidium bromide) before excision to avoid DNA mutation by UV light which is usually used for visualisation of ethidium bromide-stained bands (Huang and Fu, 2005). The gel was viewed on a light box, and the 921 bp *TbCATL* product was excised, followed by purification using the Zymoclean™ Gel DNA Recovery Kit (ZymoResearch, Orange, CA, USA), according to the manufacturer's protocol. After the colony PCR (Section 2.6) and restriction digestion, recombinant colonies that tested positive for the presence of an insert at the expected size of 921 bp *TbCATL* were sequenced at the sequencing unit of the Central Analytical Facility (CAF), Stellenbosch University.

2.8 Subcloning of the *TbCATL* gene construct into the bacterial pET28a and pET32a expression vectors

The *TbCATL* inserts found to be free from mutations by sequencing were used for sub-cloning into the pET28a and pET32a expression vectors (Novagen, Darmstadt, Germany) (Figure 2.2). The expression vectors were subjected to restriction digestion with BamHI and XhoI restriction enzymes as described in Section 2.7. The reaction mixture (50 µL) was added to (5 µL) loading dye and electrophoresed in a 1% (w/v) agarose gel in 1 x TAE buffer, containing 10 µg/mL SYBR Safe as described in Section 2.3. The restriction digested pET28a and pET32a expression vectors were excised and purified using the Zymoclean™ Gel DNA Recovery Kit as detailed in Section 2.7. Relevant DNA quantities necessary for ligation were determined using the NanoDrop™ 2000/2000c Spectrophotometer (Thermo Fisher Scientific, Waltham, MA USA).

The digested *TbCATL* inserts and expression vectors were ligated with vectors in a 3:1 ratio of vector to gene insert, as previously described in section 2.4. Transformation of ligation mixture into *E. coli* JM 109 cells were performed using the TransformAid® Bacterial Transformation Kit, as described in Section 2.5. The cells were plated onto 2 x YT plates containing antibiotics, kanamycin for pET28a and ampicillin for pET32a to select recombinant colonies. Recombinant colonies were screened by colony PCR, as described in Section 2.6, to ensure that the correct insert was present before the expression of recombinant proteins was initiated. The recombinant colonies were transformed into the expression host *E. coli* BL21 (DE3) cells (New England Biolabs, Ipswitch, MA, USA) prepared by $CaCl_2$ transformation (Singh *et al.*, 2010). Briefly, 10 mL 2 x YT medium was inoculated with a single colony of *E. coli* BL21 (DE3) and grown at 37 ^{o}C for 16 h, with agitation. The overnight culture was diluted 1:100 with 90 mL 2 x YT media and grown at 37 ^{o}C, with agitation until an OD_{600} of 0.3 to 0.4 was reached. The cultures were transferred into ice-cold, sterile centrifuge tubes and incubated on ice for 10 min. The cells were pelleted (4500 x *g*, 10 min, 4 ^{o}C) and resuspended in ice-cold, sterile $CaCl_2$ solution [60 mM $CaCl_2$, pH 7.0 (40 mL)]. The cell solution was pelleted (4500 x *g*, 10 min, 4 ^{o}C), and the pellet was resuspended in ice-cold sterile $CaCl_2$ solution (2 mL). The competent *E. coli* BL21 (DE3) cells (30 µL) were incubated with recombinant colonies on ice for 30 min. After that, cells were heat-shocked at 42 ^{o}C for 90 s, followed by an immediate incubation on ice for 2 min. The cells were added to pre-warmed super cataboliser with catabolite expression (SOC) medium [2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 20 mM glucose 80 µl] and incubated at 37 ^{o}C for 1 h, with gentle agitation. The cells (100 µL) were plated onto pre-warmed 2 x YT plates containing 34 µg/mL kanamycin for pET28a- and 100 µg/mL ampicillin for pET32a recombinant clones and incubated at 37 ^{o}C for 16 h. Plasmid DNA was isolated

from the recombinant colonies using the NucleoSpin® Plasmid kit and used as the template for colony PCR, as described in Section 2.6 to confirm the presence of the *TbCATL* inserts.

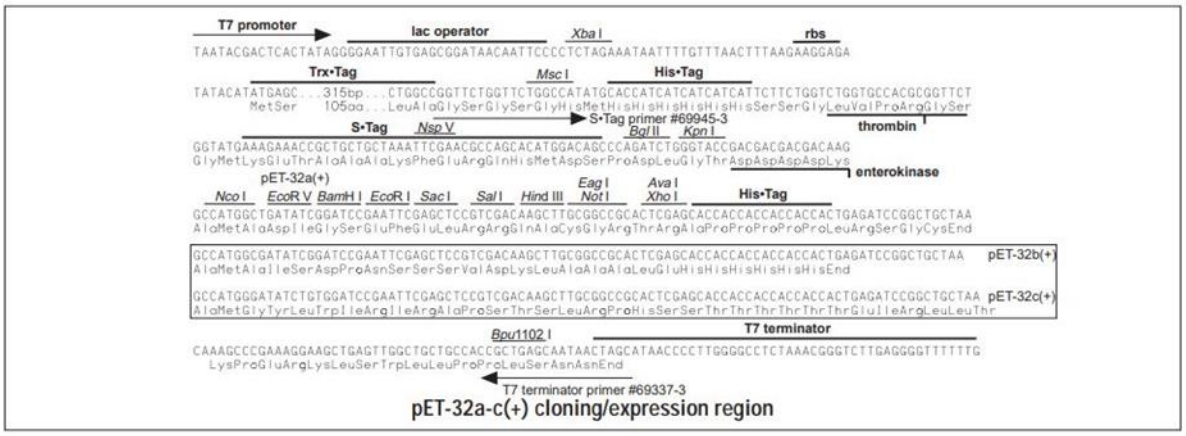

Figure 2.2: Map of pET28a and pET32a expression vector multiple cloning sites. (Novagen Technical Manual).

2.9 Recombinant expression of *TbCATL*

Recombinant BL21 (DE3) pET28a/pET32a *TbCATL* colonies were inoculated into 2 x YT medium (10 mL) containing 34 µg/mL kanamycin for pET28a and 100 µg/mL ampicillin for pET32a recombinant clones. The cultures were incubated overnight at 37 ^{o}C, with agitation before diluted 1:100, using 90 mL of fresh 2 x YT liquid media containing kanamycin or ampicillin, as described before. The cultures were grown at 37 ^{o}C in baffled flasks, with agitation until an OD_{600} of 0.68 was reached. Cultures were incubated at 4 ^{o}C for 15 min, and recombinant expression was induced with 0.3 mM IPTG followed by incubation at 16 ^{o}C, for 24 h, with agitation. The *E. coli* BL21 (DE3) allows the expression of different proteins. This is achieved by placing the desired gene into the multi-copy plasmid vector under transcriptional control of a regulatable strong promoter. The expression of a foreign gene in a regulatable promoter system depends on various factors, and these factors include the concentration of the inducer (lactose or IPTG). The *lac* operon in *E. coli* codes for enzymes required for the transport and metabolism of the sugar lactose. The *lac* repressor regulates transcription from the *lac* promoter by inhibiting the *lac* operon in the absence of the inducer (allolactose). This inhibits the proper binding of the T7 RNA polymerase to the promoter site. In the presence of lactose, cells metabolise it, forming an allolactose isomer that removes a *lac* repressor from the *lac* operon to induce transcription and protein expression. The lactose usually gets depleted in the medium. The IPTG mimics the structure of allolactose. Therefore, it is added in the medium to remove the *lac* repressor from the *lac* operon. As a result, T7 RNA

polymerase binds to the T7 promoter and initiates transcription and induces protein expression (Studier and Moffatt, 1986; Donovan *et al.*, 1996). The cells were pelleted by centrifugation (6000 x *g*, 10 min, 4 ^{o}C) and the pellets were resuspended in 1% (v/v) Triton X-100-PBS (2 mL). A final concentration of 1 mg/mL lysozyme (Sigma, St. Louis, MO, USA) was added. In the T7-based promoters, leaky expression is controlled by the co-expression of T7 lysozyme from the pLysS or pLysE plasmids. Following the induction of expression, lysozyme is added to hydrolyse the glycosidic bonds between N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) in the peptidoglycan cell wall of bacteria, causing lysis of the bacterial cell. After the addition of lysozyme, it was incubated at 37 ^{o}C, for 30 min, with agitation followed by freezing at -80 ^{o}C for 16 h. The cells were thawed at RT, sonicated four times for 30 s each, and cellular debris was removed by centrifugation (6000 x *g*, 10 min, and 4 ^{o}C). The insoluble pellet was resuspended in 1% (v/v) Triton X-100-PBS (2 mL) and stored at -20 ^{o}C.

2.10 Solubilising, refolding and purification of recombinant *TbCATL*

A high level of expression of the recombinant proteins from the *E. coli* usually results in the protein produced within inclusion bodies. Inclusion bodies protect host proteins from degradation by foreign (heterogeneously expressed) proteases or the toxic effects of foreign proteins (Kaur *et al.*, 2018). Inclusion bodies were solubilised and refolded to obtain the native conformation of *Tb*CATL. Solubilisation of the inclusion bodies is usually performed with a high concentration of denaturants and chaotropic agents, such as guanidine hydrochloride and urea or non-chaotropic agents such as N-lauroylsarcosine sodium salt, also known as sarkosyl (Singh *et al.*, 2015). After solubilisation, proteins are refolded in an optimised refolding buffer and denaturants removed during affinity chromatography using an immobilised metal (cobalt or nickel) affinity resin that results in high purity of the recombinant proteins (Phan *et al.*, 2003). The solubilisation and refolding were performed using sarkosyl (non-chaotropic), which was adapted and modified from Schlager and co-workers (Schlager *et al.*, 2012).

Recombinant *Tb*CATL was expressed at high levels within the inclusion bodies. The insoluble expression pellet was resuspended in lysis buffer (PCL) [8 mM Na_2HPO_4, 286 mM NaCl, 1.4 mM KH_2PO_4, 2.6 mM KCl, 1% (w/v) SDS, pH 7.4, 1 mM 2-mecarptoethanol (BioRad, CA, USA), (5 mL for a 50 mL culture)], sonicated twice for 4 min on ice and incubated on ice for 1 h. Cell debris was removed from the soluble protein by centrifugation (13000 x *g*, 20 min, 4^{o}C). The *Tb*CATL solubilisation was followed by refolding and purification using a TALON® (Cu^{2+}) immobilised metal affinity chromatography (IMAC) resin. Briefly, 1 mL of TALON® IMAC resin (Takara Bio Inc, Shiga, Japan) was placed in a 10 mL chromatography column (Bio-Rad, Hercules, California, USA). The column was washed with distilled water (dH_2O;2 mL) and

equilibrated with 5 mL sarkosyl wash buffer (PCW) [8 mM Na_2HPO_4, 286 mM NaCl, 1.4 mM KH_2PO_4, 2.6 mM KCl, 0.1% sarkosyl, 5 mM imidazole pH 7.4]. The solubilised protein lysate (5 mL) was incubated with the resin at 4 ^oC for 3 h with agitation using an end-over-end rotator. The unbound proteins were collected, and the resin was washed with PCW until an absorbance value at 280 nm of 0.02 was reached. The bound proteins were eluted in 1 mL fractions with sarkosyl elution buffer (PCE) [8 mM Na_2HPO_4, 286 mM NaCl, 1.4 mM KH_2PO_4, 2.6 mM KCl, 0.1% sarkosyl, 50 mM imidazole pH 7.4, 10 mL]. The column was regenerated with dH_2O (2 mL), washed with 6 M guanidine hydrochloride (5 mL), dH_2O (3 mL), PCW (3 mL) and stored in 30% (v/v) ethanol at 4^oC. As described in Section 2. 10, samples of unbound and eluted fractions were analysed on a 12.5% SDS-PAGE gel (Laemmli, 1970) and a western blot as detailed in Sections 2.12 and 2.13. Suitably purified fractions were pooled and subjected to dialysis overnight against PBS at 4 ^oC with magnetic stirring using 12.4 kDa cut-off Snakeskin® pleated dialysis tubing (Thermo Fisher Scientific, Waltham, MA USA). Three changes of PBS dialysis buffer achieved the optimal removal of all sarkosyl and imidazole. The dialysed sample was subsequently concentrated using dialysis against solid polyethylene glycol (PEG) M_r 20000 (Merck, Darmstadt, Germany). The protein concentration of the PEG concentrated samples was determined using BCA and Bradford assays (Section 2.11) and purity was analysed using 10% reducing Tris-tricine SDS-PAGE, (Section 2.12).

2.11 Protein quantification

The protein concentration of purified *Tb*CATL was determined using two methods: the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA USA) and the Bradford protein assay (Bradford, 1976). The BCA assay has a working reagent that includes bicinchoninic acid (BCA) (Smith *et al.*, 1985). The reduction of Cu^{2+} to Cu^{+} by protein in an alkaline medium result in a feint blue colour. The BCA reagent then chelates the cuprous ions (Cu^{+}) resulting in a dark purple colour, enabling the quantification of protein. The BCA/copper complex has a strong linear absorbance at 562 nm with increased protein concentration. The Pierce™ BCA Protein Kit makes use of bovine serum albumin (BSA) (Thermo Fisher Scientific, Waltham, MA USA) standards of known protein concentration and reagents A and B. The BSA standards were used for the construction of a calibration curve (Figure 2.3). The BSA standards at concentrations ranging from 25 to 2000 µg/ML were prepared from a stock solution of 2 mg/mL. The BCA reagents A and B were mixed in a ratio of 50:1 (v/v) to prepare the BCA working reagent. A volume of 25 µl of each BSA standard and the protein of unknown concentrations were pipetted (in duplicate) into separate wells of a 96 well plate Nunc® plate (Thermo Fisher Scientific, Waltham, MA USA), followed by the addition of 200 µl BCA working reagent. The plate was covered and incubated at 37 ^oC for 30 min. Colour development was

measured at 562 nm using the Versamax spectrophotometer plate reader (BMG Labtech, Offenburg, Germany).

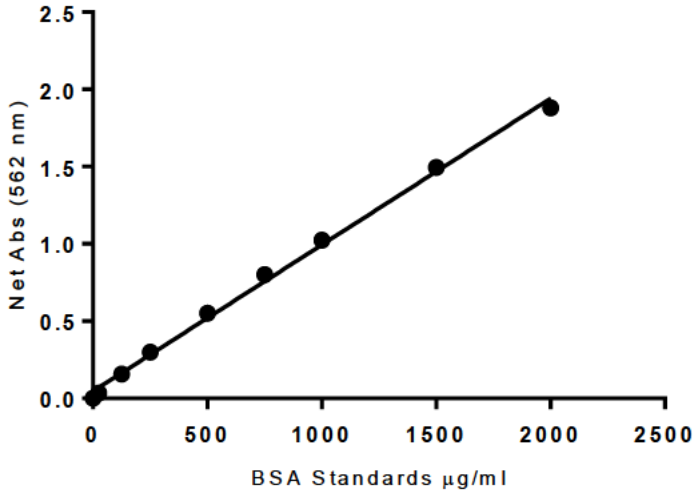

Figure 2.3: The BCA standard curve for quantification of recombinant *Tb*CATL. BSA standards (25 -2000 µg/mL) were mixed with BCA working reagent. After incubating BCA standards containing BCA working reagent, absorbance values were measured at 562 nm. The standard curve was constructed by plotting the absorbance at 562 nm against BSA concentration (µg/mL). The equation of the trend line is Absorbance = 0.0009478[Protein] + 0.04575 with a correlation coefficient of 0.9966.

The Bradford dye-binding assay is also used to determine protein concentration (Bradford, 1976). It involves binding proteins to Serva blue G-250 dye under acidic conditions, resulting in a shift in the dye's absorbance maximum from 465 nm (cationic red form) to 595 nm (anion blue form) (Bradford, 1976; Read and Northcote, 1981). The presence of basic amino acid residues, arginine, lysine, and histidine, results in the protein-dye complex formation (Compton and Jones, 1985). The extinction coefficient of the blue dye protein complex is far greater than that of the free dye at 595 nm. This increases the sensitivity in protein quantification (Bradford, 1976). A standard curve was constructed using BSA (Sigma Aldrich, St. Louis, Missouri, USA) in a range of 100 - 10 µg/mL (Figure 2.4). Briefly, triplicate samples of 0 - 50 µL of the 100 µg/mL BSA standards were diluted to 50 µL with distilled water and 950 µl of Bradford dye (Merck Biolab, Darmstadt, Germany) [50 mg Serva blue G in 50 mL of 88% (v/v) phosphoric acid and 23.5 mL 99.5% (v/v) ethanol] added. The colour development after 2 min was recorded at a wavelength of 595 nm. The calibration curve was constructed (Figure 2.4) and used for the determination of the unknown concentration of purified protein.

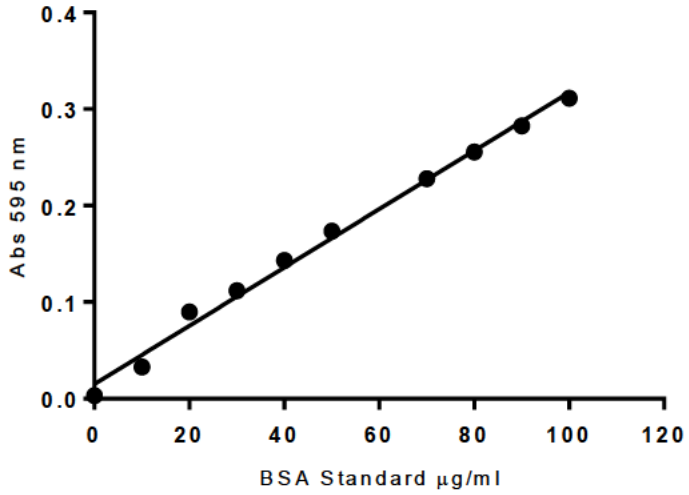

Figure 2.4: The Bradford standard curve for quantification of the recombinant *Tb*CATL. BSA standards (100-10 µg/mL) or recombinant *Tb*CATL samples were mixed with Bradford reagent. After incubation for 2 min, absorbance values were measured at 595 nm. The standard curve was constructed by plotting the absorbance at 595 nm against BSA concentration. The equation of the trend line is given by Absorbance = 0.003035[BSA] + 0.01413 with a correlation coefficient of 0.9923.

2.12 Laemmli and Tris-tricine SDS-PAGE analysis of recombinant *Tb*CATL expression

Electrophoresis is a technique used to separate charged molecules contained in a gel matrix in an electrical field based on their molecular weight. If proteins are subjected to polyacrylamide gel electrophoresis (PAGE), separation will be based on the charge:mass ratio of the proteins. Proteins are denatured when they react with the detergent sodium dodecyl sulfate (SDS), forming a negatively charged complex in a ratio of about 1.4 g of detergent to 1 g of protein (Reynolds and Tanford, 1970). These negatively charged proteins migrate through the gel matrix towards the anode, and the migration rate only depends on the relative size of the protein. Smaller molecular weight proteins migrate faster than larger molecular weight proteins. In PAGE, the effects of size, shape, and conformation on the migration of complex cannot be determined, and an estimate of molecular mass is thus inaccurate. In reducing SDS-PAGE gel, boiling of protein samples in the presence of a reducing agent (2-mercaptoethanol) breaks the disulfide bonds, preventing proteins from adopting their secondary and tertiary structures (Laemmli, 1970). Migration of reduced standard proteins of known molecular mass alongside the reduced proteins of unknown molecular mass can be used to estimate the size of proteins (Neville Jr, 1971).

To this end, protein samples were treated with reducing treatment buffer [125 mM Tris-HCl buffer, 4% (w/v) SDS, 20% (v/v) glycerol and 10% (v/v) 2-mecarptoethanol] at 1:1 ratio

(Laemmli, 1970) before loading onto the gel. The gel was run at 20 mA per gel until the bromophenol blue tracker dye reached a distance of about 0.5 cm from the bottom of the running gel in a tank buffer [250 mM Tris-HCl, 192 mM glycine and 0.1% (w/v) SDS]. The proteins were visualised by staining the gel with Coomassie blue R-250 [0.125% Coomassie brilliant blue R-250, 50% (v/v) methanol and 10% acetic acid], followed by destaining the background using a destaining solution [50% (v/v) methanol, 10% (v/v) acetic acid]. To determine the relative molecular mass (M_r) of protein analysed by reducing SDS-PAGE, a standard curve (Figure 2.5) was constructed using proteins of known M_r [phosphorylase B (97.4 kDa), bovine serum albumin (BSA, 68 kDa), ovalbumin (45 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (SBTI, 21.5 kDa) and lysozyme (14 kDa) (Sigma, St. Louis, MO, USA)] and plotting the relative migration distance (R_f) of individual proteins against the log M_r of each protein.

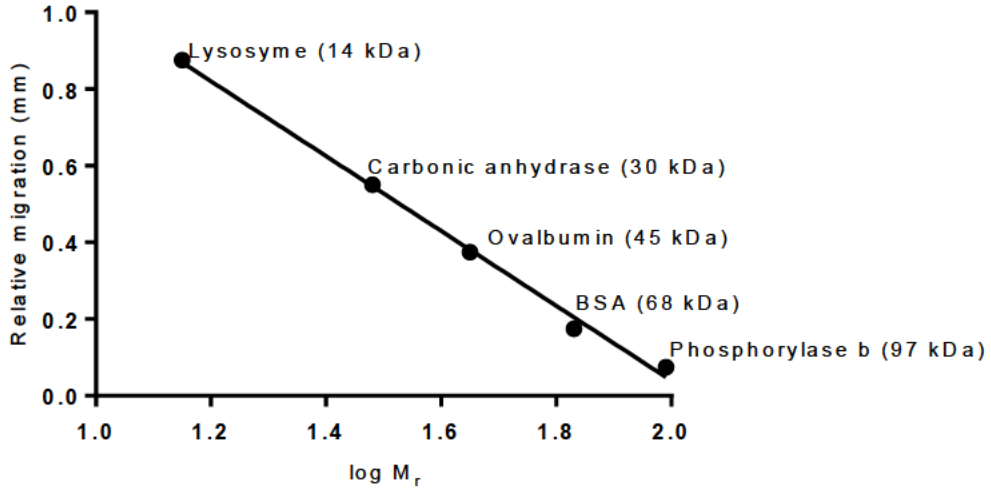

Figure 2.5: Calibration curve for standard proteins analysed by reducing SDS-PAGE. A standard curve was constructed with proteins of known sizes after separation on a 12.5% reducing SDS-PAGE. The relative migration distance of each protein was measured and plotted against log Mr of each standard protein. The equation of a standard curve was Relative migration = -0.9766(log Mr) + 1.992, with a correlation coefficient of 0.9959.

The Tris-Glycine buffer system used in the Laemmli SDS-PAGE system does not provide good resolution of protein bands with sizes smaller than 30 kDa and was replaced by a Tris-Tricine buffer system to improve the resolution of smaller proteins (Schägger and von Jagow, 1987).

To analyse the solubility of the recombinantly expressed *Tb*CATL, the expression supernatant and pellet samples were electrophoresed on 12.5% reducing SDS-PAGE gels as per Laemmli (1970). A discontinuous buffer system was used containing a running gel buffer (Ornstein,

1964) (1.5 M Tris-HCl, pH 8.8) and a stacking gel buffer (500 mM Tris-HCl, pH 6.8). The protein samples were combined with the reducing treatment buffer and SDS-PAGE was conducted at 40 mA per 2 gels in the tank buffer described before using the BioRad Mini protein gel electrophoresis apparatus (BioRad, CA, USA). For Tris-Tricine SDS-PAGE (Schägger and von Jagow, 1987), gel buffer [3.0 M Tris-HCl, 0.3% (m/v) SDS, pH 8.45] was used in a system containing cathode buffer [0.1 M Tris-HCl, 0.1 M Tricine, 0.1% (m/v) SDS, pH 8.25] and was conducted at 40 mA per 2 gels in an anode buffer (0.2 M Tris-HCl, pH 8.9). After electrophoresis gels were stained with Coomassie Blue and destained as described above.

2.13 Western blot to detect the presence of a His-tag on recombinant *Tb*CATL

In western blotting, proteins electrophoretically transferred from an SDS-PAGE gel to a nitrocellulose membrane are identified using antibodies (Towbin *et al.*, 1979). Methanol is included in the blotting buffer to enhance the hydrophobic binding of the protein-SDS complex to the nitrocellulose membrane (Gooderham, 1984). Proteins were subjected to SDS-PAGE as described in Section 2.12 and transferred onto a nitrocellulose membrane for selective immunodetection of the immobilised protein (antigens). The specificity of antibody-antigen interaction plays an important role in identifying specific proteins.

Western blotting was performed according to Towbin *et al.* (1979), and proteins separated by SDS-PAGE were transferred onto nitrocellulose (PALL Corp, Ann Arbor, USA) using a Sigma (model B 2529) Semi-phor® semi-dry blotting apparatus (Sigma Aldrich, St. Louis, Missouri, USA) at a constant voltage (20V), maximum current for 55 minutes. To confirm the transfer of proteins, the nitrocellulose membrane was transiently stained with Ponceau S [0.1% (w/v) Ponceau S, 15% (v/v) acetic acid] and the positions of the molecular weight markers marked using a pencil. The unoccupied sites on the nitrocellulose membrane were blocked with 5% (w/v) non-fat milk powder (Amresco, Solon, OH, USA) in Tris-buffered saline (TBS) (20 mM Tris-HCl buffer, 200mM NaCl, pH 7.4)] at RT for 1 h. The nitrocellulose was washed three times for 5 min each with TBS. The primary antibody, anti-His [1:1500 in 0.5% (w/v) BSA-TBS] was added and incubated at 4^{o}C for 16 h. The nitrocellulose membrane was washed two times for 5 min each with TBS and incubated in goat anti-mouse-horseradish peroxidase (HRPO) conjugate (Sigma Aldrich, St. Louis, Missouri, USA) [1:5000 in 0.5% (w/v) BSA-PBS] at RT for 1 h. The nitrocellulose membrane was washed three times for 5 min each with TBS followed by the addition of 4-chloro-1-naphthol (Thermo Fisher Scientific, Waltham, MA USA)-H_2O_2 chromogen-substrate [0.06% (w/v) 4-chloro-1-naphthol, 0.1% (v/v) methanol and 0.0015%

(v/v) H_2O_2 in BPS] and colour allowed to develop in the dark. The image was captured using the Sys Gene® imaging system (Syngene®, Frederick, Maryland, USA).

2.14 Enzymatic characterisation of recombinant *Tb*CATL

2.14.1 Gelatin containing SDS-PAGE to visualise the proteolytic activity of recombinant *Tb*CATL

Zymography is a technique used to detect and identify protein activity on a gelatin substrate-containing non-reducing SDS-PAGE gel (Heussen and Dowdle, 1980). The proteolytic activity of recombinant *Tb*CATL was analysed on a gelatin-containing Laemmli SDS-PAGE gel. Gelatin 1% (w/v) was dissolved in 1.5 mL of SDS-PAGE running gel buffer at 37 ^{o}C, and added to 2.25 mL running buffer for casting a 12.5% SDS-PAGE as described in Section 2. 12. Samples were treated with equal volumes of non-reducing treatment buffer [125 mM Tris-HCl buffer pH, 6.8, 4% (w/v) SDS and 20 % (v/v) glycerol]. After electrophoresis, gels were washed twice over 1 h with 2.5% (w/v) Triton X-100 (50 mL) at RT to replace the denaturing detergent SDS with Triton X-100 to restore protease activity. The gel was subsequently incubated in *Tb*CATL assay buffer [50 mM sodium acetate buffer, pH 5, 150 mM NaCl, 6 mM EDTA, 0.02 (w/v) NaN_3, 6 mM dithiothreitol (DTT) (Fermentas, Vilnius, Lithuania)] overnight at 37 ^{o}C. The gel was stained with amido black solution [0.1% (w/v) amido black, 30% (v/v) methanol, 10% (v/v) glacial acetic acid] for 1 h and subsequently destained in several changes of destaining solution [30% (w/v) methanol, 10% (v/v) glacial acetic acid]. The effect of inhibitors on the proteolytic activity of *Tb*CATL on gelatin containing SDS-PAGE gel was analysed by adding different protease class-specific inhibitors in the *Tb*CATL assay buffer. The gels were incubated at 37 ^{o}C overnight in *Tb*CATL assay buffer containing the indicated final concentrations of the respective inhibitors [E-64 (10 µM), leupeptin (10 µM), antipain (10 µM), iodoacetate (100 µM), chymostatin (10 µM), bestatin (10 µM), SBTI (10 µM), and EDTA (10 mM) (Sigma, St. Louis, MO, USA)].

2.14.2 Active site titration of recombinant *Tb*CATL

L-trans-Epoxysuccinyl-leucylamido(4-guanidino) butane (E-64) is an irreversible inhibitor for different cysteine protease, including cathepsins B and L (Barrett and Kirschke, 1981). Varying E-64 concentrations may be used to determine the residual activity of cysteine proteases (Barrett *et al.*, 1982). The concentration of active *Tb*CATL was determined using the irreversible cysteine protease inhibitor E-64 as described by Barrett *et al.* (1982) with a few changes. Briefly, 50 µL of *Tb*CATL (1 µM diluted in 50 mM Na-acetate pH 5, 150 mM NaCl, 4 mM EDTA, 0.02% NaN_3, 6 mM DTT) was incubated with 50 µL of *Tb*CATL assay buffer (no inhibitor control) or E-64 ranging from 0.1 to 1 µM, diluted in *Tb*CATL assay buffer for 15 min

at 37ºC. Following the incubation, each reaction volume was increased to 450 µL with *Tb*CATL assay buffer. A volume of 80 µL of each reaction was added in triplicate to the wells of a black FluoroNunc™ 96-well plate (Thermo Fisher Scientific, Waltham, MA USA), and assayed for activity by the addition of benzyloxycarbonyl (Z)-Phe-Arg-7-amido-4-methylcoumarin (AMC) to a final concentration of 20 µM in a 100 µl reaction volume. The fluorescence was measured in a continuous assay at 37 ºC, every 30 s for 15 min ($EX_{360\,nm}$; $EM_{460\,nm}$) using the FLUORStar Optima spectrophotometer (BMG Labtech, Offenburg, Germany). The active concentration of *Tb*CATL was determined by plotting the percentage residual activity against E-64 concentrations (Figure 2.6). The percentage activity was determined using the equation % Residual activity = V_i/V_0 x 100, where V_1 is the initial reaction velocity in the presence of inhibitor and V_0 is the initial reaction velocity in the absence of the inhibitor.

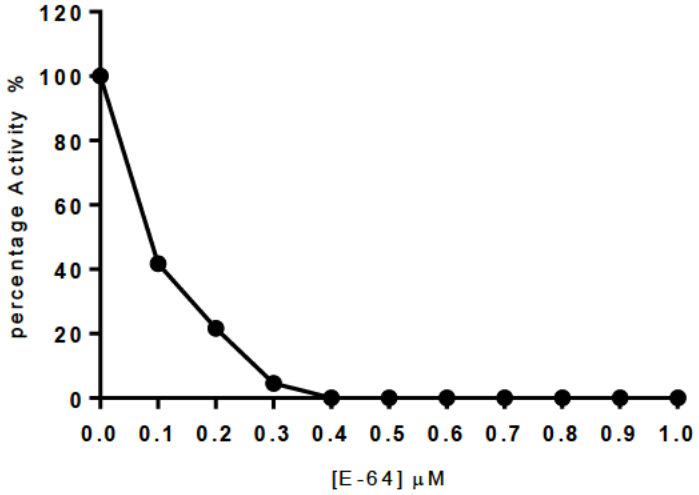

Figure 2.6: Active site titration of *Tb*CATL. Recombinant *Tb*CATL (1 µM) was titrated against varying concentrations of E-64 (0-1 µM) and incubated at 37 ˚C for 15 min and residual activity against Z-Phe-Arg-AMC was measured for each sample. Each data point represents the mean activity of two replicates.

2.14.3 The effect of pH on recombinant *Tb*CATL activity

The acetate-MES-Tris buffers (AMT) have a constant ionic strength across a wide pH range (Ellis and Morrison, 1982). Therefore, AMT buffers can be used to study the catalytic behaviour of enzymes over a range of pH values without masking the effect of pH by varying ionic strength. The optimum pH for the hydrolysis of a peptide substrate by recombinant *Tb*CATL was determined using the AMT constant ionic strength buffer (100 mM sodium acetate, 10 mM MES, 200 mM Tris-HCl, 5 mM EDTA, 6 mM DTT) titrated with HCl to pH-values ranging from 4 to 9 (Ellis and Morrison, 1982). Briefly, 1 µM of *Tb*CATL was incubated with AMT buffers at varying pH values for 10 min at 37 ºC. After incubation, 80 µL of each

reaction was added to the wells of the black FluoroNunc™ 96-well plate, in triplicate and assayed by the addition of Z-Phe-Arg-AMC (Bachem, Bubendorf, Switzerland) to a final concentration of 20 µM in a 100 µL reaction volume. The fluorescence was measured in a continuous assay at 37 ºC every 30 s for 15 min ($Ex_{360\ nm}$; $Em_{460\ nm}$) using the FLUORStar Optima spectrophotometer (BMG Labtech, Offenburg, Germany)

2.14.4 Peptide substrate specificity of *Tb*CATL

In order to determine the hydrolysis of AMC-labelled peptide substrates by *Tb*CATL, an AMC standard curve was constructed. A series of AMC dilutions (5 x 10^{-3}-15 µM) were prepared in triplicate with dH_2O and incubated with *Tb*CATL assay buffer for 10 min at 37 ºC. Fluorescence was measured ($EX_{360\ nm}$; $EM_{460\ nm}$) using the FLUORStar Optima spectrophotometer. The AMC fluorescence standard curve is shown in Figure 2.7.

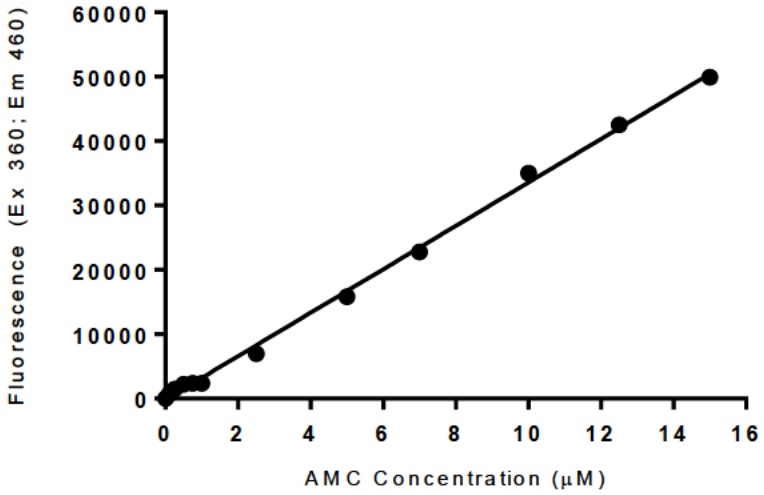

Figure 2.7: The AMC standard curve for determination of the relative amount of fluorescence released upon hydrolysis of the AMC substrates. The AMC dilutions (5 x 10^{-3}-15 µM) were incubated at 37 oC for 10 min. The fluorescence was measured at (EX360 nm; EM460 nm), and the equation of the trendline was Fluorescence= 3373[AMC] - 153.8 with a correlation coefficient of 0.998.

Using different concentrations of various fluorogenic (AMC-labelled) peptide substrates, makes it possible to determine the enzyme's maximum velocity (V_{max}) and its affinity for each substrate (K_m). This is achieved by measuring the initial velocity (v_0) for hydrolysis of each substrate at different concentrations [S] and using the Michaelis-Menten equation ($v_0 = V_{max}[S]/[S] + K_m$) as revised by Briggs and Haldane (1925). The turnover number of the enzyme, k_{cat}, is related to the active concentration, $[E_0]$, of an enzyme: $k_{cat} = V_{max}/[E_0]$. The following fluorogenic peptide substrates were used to determine the substrate specificity of

recombinant *Tb*CATL: Z-Phe-Arg-AMC, Z-Leu-Arg-AMC, H-Pro-Phe-Arg-AMC, H-Ala-Phe-Lys-AMC, Z-Arg-Arg-AMC and Z-Gly-Gly-Arg-AMC (Bachem, Bubendorf, Switzerland). Substrates at concentrations ranging from 10 – 120 µM were added after the incubation of 1 µM recombinant *Tb*CATL in *Tb*CATL assay buffer with 6 mM DTT. Fluorescence was measured at ($Ex_{360\ nm}$; $Em_{460\ nm}$) continuously using the FLUORStar Optima spectrophotometer to determine the initial steady-state velocity (V_0) of *Tb*CATL. The turnover number of the enzyme k_{cat} and V_{max} were determined using GraphPad Prism version 7.

2.14.5 Class-specific inhibitors

The recombinant *Tb*CATL was incubated with the effective concentration of cysteine protease inhibitors (E-64; 10 µM) cystatin (1 µM) and iodoacetate (100 µM), serine protease inhibitor soybean trypsin inhibitor (SBTI; 10 mM), cysteine/serine protease inhibitors (antipain, 10 µM), leupeptin, 100 µM), chymostatin, 10 µM), metallo protease inhibitor (EDTA, 10 mM) and aminopeptidase inhibitor (bestatin, 10 µM) according to Salvesen and Nagase (2001). The recombinant *Tb*CATL (1 µM E-64 titrated active enzyme) and varying concentrations of different inhibitors were incubated at 37 ^{o}C for 15 min in *TbCATL* assay buffer. Following the incubation, 20 µM of substrate H-Pro-Phe-Arg-AMC was added to the black FluoroNunc™ 96-well plate, and fluorescence was measured ($Ex_{360\ nm}$; $Em_{460\ nm}$) using the FLUORStar Optima spectrophotometer, every 30 s for 15 min. The AMC standard curve was used to determine the resultant inhibition of the enzyme, which was presented in percentage inhibition compared to that of the uninhibited enzyme.

2.15 Isolation of genomic DNA from fish blood

Blood samples were collected from 24 samples of *Clarias gariepinus* from 6 different sites along the Vaal River by Professor Annemariè Avenant-Oldewage, Department of Zoology, University of Johannesburg (Table 2.2). Genomic DNA was isolated from fish blood using Mag-Blind® Universal Pathogen 96 Kit (Omega Bio-Tek Inc, Norcross, United States), as per the manufacturer's protocol. Briefly, blood samples were thawed on ice, and the distributor tube was centrifuged (12000 x *g*, 1 min, RT) to remove beads from the wall of the tube. A volume of 250 µL of blood was added to the distributor tubes. A volume of 275 µL of SLX-MIus buffer was added to the distributor tubes and mixed thoroughly by vortexing for 5 min. A volume of 50 µL of DS buffer and 20 µL of the proteinase K solution (Omega Bio-Tek Inc, Norcross, United States) was added into the homogenised samples and mixed by vortexing for 1 min. The mixtures were incubated at 70 ^{o}C for 15 min. After the incubation, 200 µL of PCP buffer was added, and samples were incubated on ice for 5 min. The mixture was centrifuged (12000 x *g*, 10 min, RT), and 300 µL of the supernatant was transferred into a 1.5 mL microcentrifuge tube. A volume of 600 µL was added into the supernatant and mixed

thoroughly by vortexing for 3 min. The mixture was incubated at RT for 5 min. A volume of 700 µL of the lysate was transferred into the MicroElute® Le DNA column and centrifuged (12000 x *g*, 1 min, RT). The filtrate was discarded together with the collection tube, and the column was placed into a new collection tube. A volume of 500 µL of HBC buffer was added into the column, and centrifuged (12000 x *g*, 30 s, RT). The collection tube was discarded, and the column was transferred into the new collection tube. A volume of 700 µL DNA wash buffer was added and centrifuged (12000 x *g*, 30 s, RT), and the collection tube was discarded, and the column was placed into the new collection tube and centrifuged (12000 x *g*, 2 min, RT) to remove excess wash buffer. The column was put into a nuclease-free microcentrifuge tube, and 50 µL of pre-heated (70 ^{o}C) elution buffer was added directly into the membrane column and incubated at RT for 2 min. After incubation, DNA was eluted by centrifugation (12000 x *g*, 1 min, RT). Before PCR, the quality of isolated DNA was visualised by adding 5 µL of DNA extract to 1µl of gel loading dye and electrophoresed on a 1% (w/v) agarose gel containing 10 µg/mL SYBR Safe in 1xTAE buffer, as detailed in Section 2.3.

Table 2.2: Fish blood samples from *Clarias gariepinus* and their codes collected from Vaal River and used for DNA isolation and PCR amplification of trypanosomes SSU rRNA gene.

Fish samples	Codes
Standerton C7	C7-St
Bloemhof Dam C8	C8-Bh
Bloemhof Dam C3	C3-Bh
Bloemhof Dam C11	C11-Bh
Vaal Dam C7	C7-Vd
Douglas Weir C2	C2-Dw
Vaal River Barrage C14	C14-VRB
Vaal Dam C6	C6-Vd
Vaal River Barrage C5	C5-VRB
Vaal Harts Weir C11	C11-Vh
Douglas Weir C7	C7-Dw
Douglas Weir C6	C6-Dw
Bloemhof Dam C5	C5-Bh
Vaal River Barrage C4	C4-VRB
Vaal Dam C2	C2-Vd
Vaal Harts Weir C10	C10-Vh
Vaal River Barrage C8	C8-VRB
Standerton C3	C3-St
Vaal River Barrage C6	C6-VRB
Vaal River Barrage C62	C62-VRB
*	C61
*	CG6-VRB
*	CS4-VPB
*	CG8-UFS

C, *Clarias,* and the number represents the specific fish specimen
*The unnamed fish samples

Trypanosome-specific primers (Sehgal *et al.*, 2001) were used to detect parasites in *Clarias* samples. Two rounds of PCR were employed in a nested protocol to maximise sensitivity. In the first round of PCR, primers S-762 (GACTTTTGCTTCCTCTA(A/T)TG) and S-763 (CATATGCTTGTTTCAAGGAC) were used to amplify a small section of the 5'- to 3'- end of the small subunit ribosomal RNA (SSU rRNA) gene. Briefly, the PCR reaction included 1.25 µM of each primer, 1 U DreamTaq green DNA polymerase master mix (Thermo Fisher Scientific, Waltham, MA USA), 25 mM$MgCl_2$, 10 µg/µL BSA and 0.5 µL of DNA template to a

total volume of 10 µL reaction. The DNA template was not included in the negative control. Cycling conditions were as follows: incubation at 94 °C for 30 s as the initial denaturation, followed by four cycles of 94 °C for 30 s, 50 °C for 30 and 62 °C for 2 min, followed by 94 °C for 30 s and 34 cycles of 55 °C for 30 s, 68 °C for 2 min and a final extension at 68 °C for 5 min. In the second round, 0.5 µL of the PCR product (DNA template) was added to 1 U of DreamTaq green DNA Polymerase Master mix, 25 m$MgCl_2$, 10 µg/µL BSA and 1.25 µM of each primer S-755 (CTACGAACCCTTTAACAGCA) and S-823 (CGAA(T/C)AACTGC(C/T)CTATCAGC) to a total volume of 10 µL. DNA template was not included in the negative control. The cycling conditions were as follows: incubation at 94 °C for 30 s, followed by 34 cycles of 94 °C for 30 s, 60 °C, 68 °C 30 s and a final extension at 68 °C, 5 min. A sample of 2.5 µL of the PCR product was added to 1 µL loading dye subsequently electrophoresed on 1 %(w/v) agarose containing 10 µg/mL SYBR Safe in 1 x TAE buffer, as described in Section 2.3. Amplicons were excised from the gel, and were sequenced at the Central Analytical Facility (CAF), Stellenbosch University

2.16 Sequence alignment and phylogenetic analysis

The electropherograms of the sequences received from CAF were viewed and edited. All sequences were BLASTed against NCBI GenBank (Table 2.3) to confirm sequence identity. The sequence data generated in the present study and sequences obtained from GenBank were combined and aligned using ClustalW version 2.0 (Larkin *et al.*, 2007). Alignments were optimised manually using BioEdit version 7.2.5 (Hall, 1999) to ensure homology. The best-fit substitutions model was estimated using the Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc) in the JModelTest version 2.1.10 (Darriba *et al.*, 2012). The best-fit substitution models were used in both the maximum likelihood (ML) and Bayesian inference (BI). The ML analysis was conducted using Garli 2.0 (Zwickl, 2006), and Bayesian analysis was performed using MrBayes version 3 (Ronquist and Huelsenbeck, 2003) on CIPRES. All trees were midpoint rooted. Nodal support for the most likely tree was assessed using 1000 bootstraps replicates. The consensus tree was constructed using the program CONSENSE, which is part of the PHYLP package version 3.6. a3 (Felsenstein, 2002). Two independent BI runs were conducted, each comprising 20 million generations. The convergence of MCMC chains was assessed in Tracer version 1.7.1 (Rambaut *et al.*, 2018). All Effective Sample Size (ESS) values were > 200, and the first 20% of trees were removed from the tree file as burn-in. A 50% majority-rule consensus tree was constructed in CONSENSE. The ML and BI trees were compared for consistency, and thereafter, the posterior probability values were annotated onto the most likely tree.

Table 2.3: *Trypanosoma* species, accession numbers and fish hosts obtained from GenBank.

***Trypanosoma* Species**	**Host**	**GenBank accession number**	**Reference**
Trypanosoma sp. ex *Sander lucioperca*	*Sander lucioperca*	KJ601723	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma sp. ex *Abramis brama*	*Abramis brama*	KJ601712	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma ophiocephali	*Channa argus*	EU185634	(Gu *et al.*, 2007)
Trypanosoma sp. *Pseudobagri*	*Pseudobagras fulvidraco*	EF375884	(Gu *et al.*, 2007)
Trypanosoma sp. *carpio*	*Cyprinus carpio*	EF375882	(Gu *et al.*, 2007)
Trypanosoma siniperca	*	DQ494415	(Gu *et al.*, 2007)
Trypanosoma sp. Ts-Ab-TB*	*Abramis brama*	AJ620556	(Hamilton *et al.*, 2004)
Trypanosoma sp. CLAR	*Clarias angelensis*	AJ620555	(Hamilton *et al.*, 2004)
Trypanosoma granulosum	*Anguilla anguilla*	AJ620551	(Hamilton *et al.*, 2004)
Trypanosoma sp. Ts-Se-BL*	*Scardinus erythrophthalmus*	AJ620550	(Gibson *et al.*, 2005)
Trypanosoma sp. ex *Silurus glanis*	*Silurus glanis*	KJ601721	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma sp. ex *Scardinius rythrophthalmus*	*Scardinius erythrophthalmus*	KJ601718	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma sp. R6*	*Abramis brama*	AJ620554	(Hamilton *et al.*, 2004)
Trypanosoma sp. 18S*	*Micropterus salmoides*	MH635421	(Jiang *et al.*, 2019)
Trypanosoma sp. ex *Cobitis taenia*	*Cobitis taenia*	KJ601719	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma sp. *fulvidraco*	*Pseudobagras fulvidraco*	EF375883	(Gu *et al.*, 2007)
Trypanosoma sp. *fulvidraco*	*Cyprinus carpio*	AJ620549	(Hamilton *et al.*, 2004)
Trypanosoma granulosum	*Anguilla anguilla*	AJ620552	(Hamilton *et al.*, 2004)
Trypanosoma pleuronectidium	*Melanogrammus aeglefinus*	DQ016618	(Karlsbakk and Nylund, 2006)
Trypanosoma pleuronectidium	*Gadus morhua*	DQ016614	(Karlsbakk and Nylund, 2006)
Trypanosoma pleuronectidium	*Gadus morhua*	DQ016613	(Karlsbakk and Nylund, 2006)
Trypanosoma sp. ex *Perca fluviatilis*	*Perca fluviatilis*	KJ601724	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma sp. ex *Carassius Carassius*	*Carassius carassius*	KJ601715	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma carassii	*Carassius auratus*	AJ271907	(Overath *et al.*, 1998)
Trypanosoma sp. ex *Esox Lucius*	*Esox lucius*	KJ601714	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma sp. Ts-Tt-HOD*	*Tinca tinca*	AJ620553	(Hamilton *et al.*, 2004)
Trypanosoma sp. ex *Abramis brama*	*Abramis brama*	KJ601716	(Grybchuk-Ieremenko *et al.*, 2014)
Trypanosoma pleuronectidium	*Lepidorhombus whiffiagonis*	DQ016617	(Karlsbakk and Nylund, 2006)
Trypanosoma sp. isolate bar_C04_TV_B*	*	KX866929	(Jesus *et al.*, 2018)
Trypanosoma sp. isolate bar_A04_TB_B*	*	KX866927	(Jesus *et al.*, 2018)
Trypanosoma murmanensis	*Hippoglossus hippoglossus*	DQ016616	(Karlsbakk and Nylund, 2006)

*The unnamed *Trypanosoma* species and fish hosts

3 CHAPTER 3: RESULTS

3.1 Genomic DNA isolation from *T. b brucei*

The full-length sequence mRNA (Appendix 1) for *T. b. brucei* cysteine proteases (*TbCATL*, GenBank accession: X16465.1) was identified. Genomic DNA was isolated from cultured *T. b brucei* using the Qiagen DNA extraction kit to obtain enough DNA to be used as a template for the PCR amplification of *TbCATL*. As shown in Figure 3.1, the genomic DNA size (3.5 x 10^7 bp; Melville et al, 2004) was larger than the molecular weight marker band (10000 bp).

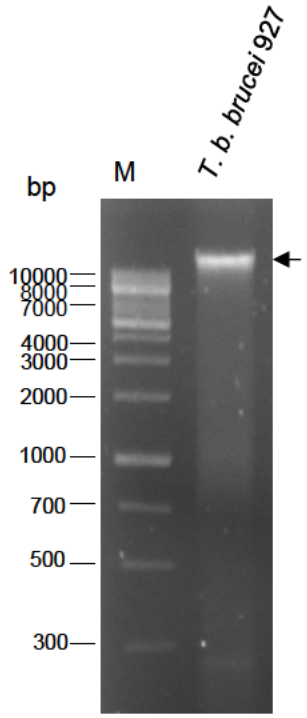

Figure 3.1: Analysis of genomic DNA isolated from *T. b brucei*. Genomic DNA isolated from *T. b brucei* 927 was electrophoresed in a 1% (w/v) agarose gel, containing 10 µg/mL CYBR Safe. M: Solis BioDyne. Arrow shows isolated DNA.

3.2 Amplification of *TbCATL* from genomic DNA

The genomic DNA from *T. b brucei* was used as the template for PCR amplification of the *TbCATL* gene using primers specifically designed for this purpose (Table 2.1). A high-fidelity PCR enzyme with an ability to proofread was used and resulted in a single band at the expected size of ~921 bp (Figure 3.2).

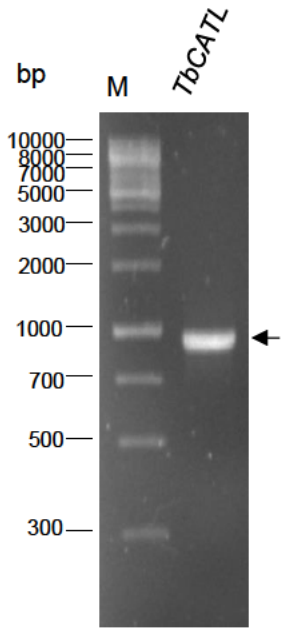

Figure 3.2: Analysis of TbCATL amplified from *T. b brucei* genomic DNA. *TbCATL* was amplified from the genomic DNA of *T. b brucei* 927. The PCR product was electrophoresed in 1 % (w/v) agarose gel, containing 10 µg/mL CYBR Safe. M: Solis BioDyne. Arrow shows amplicon at ~921 bp.

3.3 Cloning of *TbCATL* into T-vector pGEM®-T Easy and colony PCR

The amplified ~921 bp *TbCATL* PCR product was purified using ZymoResearch clean and concentrator™ kit. Before sequencing and subcloning into the expression vectors, the purified PCR product was ligated into pGEM®-T Easy cloning vector and transformed into *E. coli* JM 109 cells that resulted in non-recombinant and recombinant colonies due to blue/white colony screening. Colony PCR was performed to amplify the 921 bp *TbCATL* gene after the isolation of plasmid DNA from recombinant white colonies. PCR amplification using *TbCATL* forward and reverse gene primers resulted in six recombinant pGEM®-T Easy *TbCATL* colonies being identified, shown by the presence of a 921 bp product (Figure 3.3, panel A).

Plasmid DNA was isolated from two positive pGEM®-T Easy *TbCATL* recombinants using a NucleoSpin® kit. These were digested with restriction enzymes BamHI and XhoI to ensure that the *TbCATL* insert DNA was correctly in place and released from pGEM®-T Easy before ligation into an expression vector. After the restriction digestion, each reaction resulted in two bands at 3015 bp and 921 bp, the correct size of the pGEM®-T Easy cloning vector and the *TbCATL* insert, respectively, as shown in Figure 3.3, panel B, after being analysed by agarose gel electrophoresis. The *TbCATL* inserts were purified by gel extraction in preparation for sub-cloning into expression vectors (Figure 3.4).

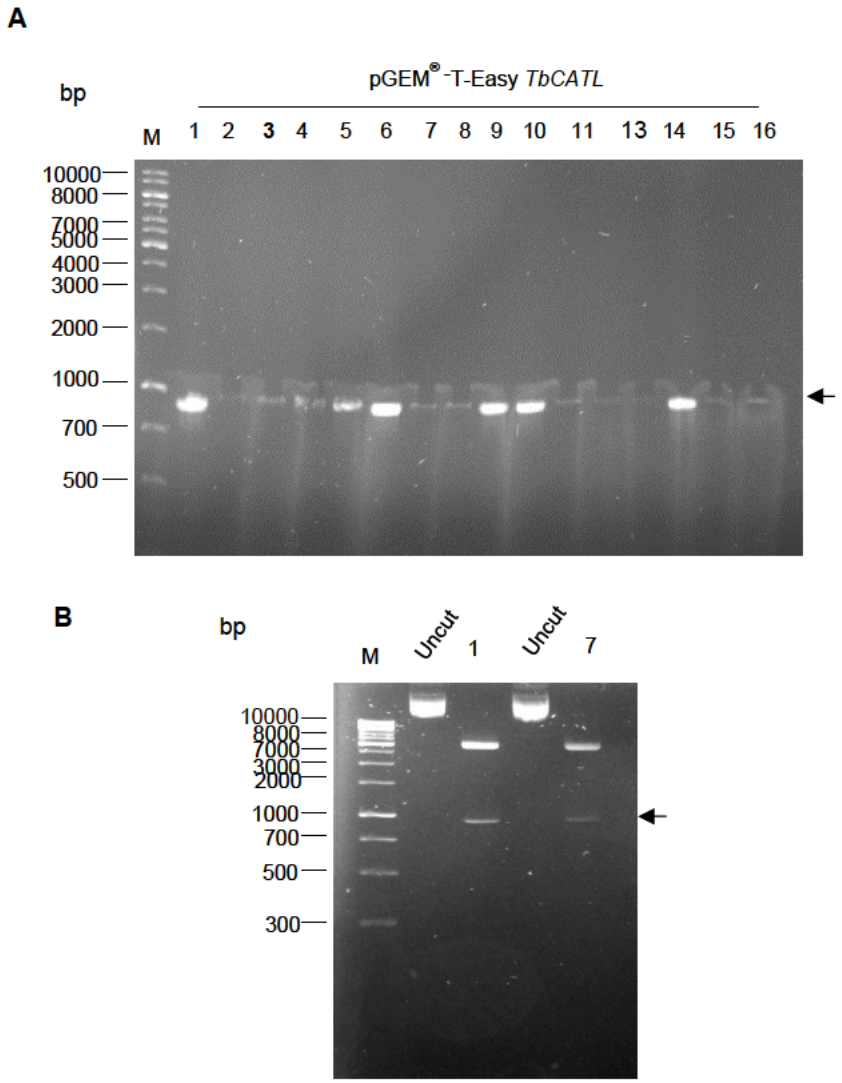

Figure 3.3: Screening for recombinant *TbCATL* clones ligated into pGEM®-T Easy cloning vector by PCR amplification and gel extraction of positive recombinant *TbCATL* clones. After the transformation of the **(A)** pGEM®-T Easy *TbCATL* ligation mixture into competent *E. coli* JM 109 cells, DNA plasmid isolated from white colonies was subjected to PCR using *TbCATL* forward and reverse gene primers and **(B)** positive recombinant *TbCATL* clones (1 & 7) were digested with restriction digestion enzymes BamHI and XhoI at 37 ºC for 2 h. Followed by excision of the *TbCAT*L insert from agarose and purified with ZymoResearch gel extraction kit. Samples were electrophoresed in 1% (w/v) agarose gel containing 10 µg/mL CYBR Safe. M: Solis BioDyne. Arrow shows amplicon at ~921 bp.

3.4 Sub-cloning of *TbCATL* gene into bacterial pET28a and pET32a expression vectors

The pET28a and pET32a expression vectors were linearised with the same enzymes (BamHI and XhoI). Both vectors were purified by agarose gel extraction after being analysed for purity by agarose gel electrophoresis. The linearised vectors revealed a single band. As shown in Figure 3.4, the linearised plasmids' sizes were larger than the un-linearised vectors. *TbCATL* inserts were ligated into linearised pET32a and pET28a expression vectors, respectively.

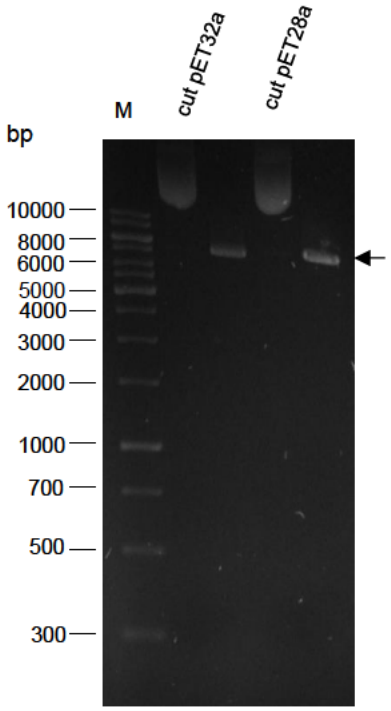

Figure 3.4: Gel extraction of pET32a and pET28a expression vectors. After digestion of pET32a and pET28a expression vectors with restriction enzymes BamHI and XhoI, they were excised and purified from 1% (w/v) agarose gel containing 10 µg/mL CYBR Safe. M: Solis BioDyne. Arrow shows linearised vectors.

The ligated pET28a-*TbCATL* and pET32a-*TbCATL* were transformed into *E. coli* JM 109 cells using the $CaCl_2$ transformation method. The plasmid DNA of the resulting recombinant pET28a-*TbCATL* and pET32a-*TbCATL* was isolated and used as the template for colony PCR amplification of the 921 bp gene. PCR amplification using forward and reverse gene primers resulted in eight recombinant pET28a-*TbCATL* and pET32a-*TbCATL* colonies being identified, shown by the presence of an approximately 921 bp product (Figure 3.5). Recombinant clones were selected and given new names, A1 and A7. Therefore, A1, A7pET28a *TbCATL* and A1, A7pET32a *TbCATL*.

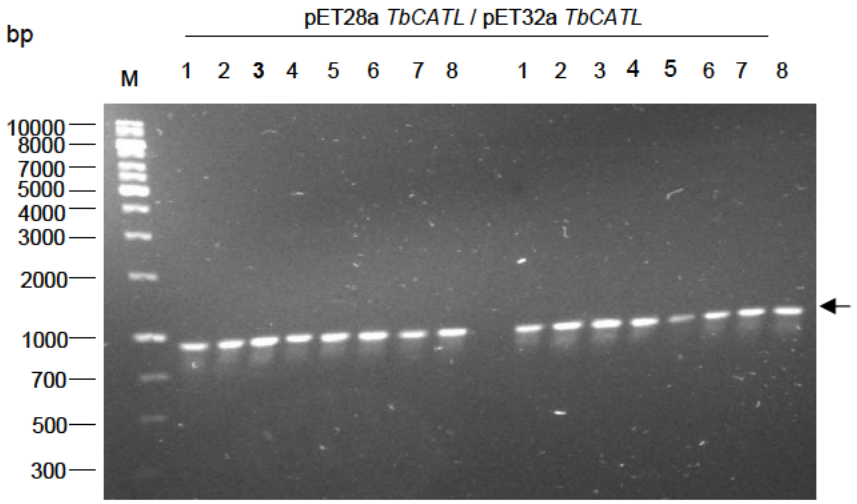

Figure 3.5: Screening for recombinant *TbCATL* clones, following ligation of *TbCATL* into pET28a and pET32a expression vectors, from the isolated plasmid DNA by PCR amplification. Following transformation of, A1pET28a *TbCATL* and A7pET32a *TbCATL* ligation mixtures into competent *E. coli* JM 109 cells, the plasmid DNA isolated from each clone was subjected to PCR amplification using *TbCATL* forward and reverse gene primers. PCR products were electrophoresed on a 1% (w/v) agarose gel containing 10 µg/mL CYBR Safe. M: Solis BioDyne. Arrow shows amplicon at ~921 bp.

The plasmids isolated from selected clones were transformed into *E. coli* BL21 (DE3) cells for protein expression. The plasmid DNA of the resulting recombinant of pET28a *TbCATL* and pET32a *TbCATL* was isolated and used as the template for colony PCR amplification of the 921 bp gene. PCR amplification using forward and reverse gene primers resulted in two recombinant pET28a *TbCATL* and pET32a *TbCATL*, respectively (Figure 3.6).

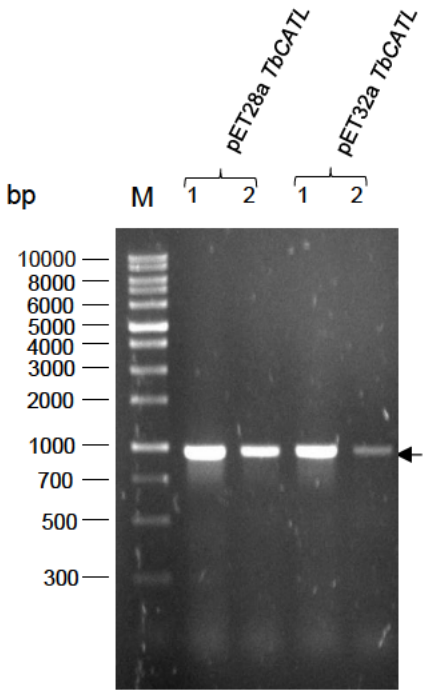

Figure 3.6: Screening for recombinant *TbCATL* clones, following ligation of *TbCATL* into pET28a and pET32a expression vectors, from the isolated plasmid DNA, by PCR amplification. Following transformation of *E. coli* BL 21 (DE3) cells with pET28a *TbCATL* and pET32a *TbCATL* the plasmid DNA isolated from each clone was subjected to PCR amplification using *TbCATL* forward and reverse gene primers. PCR products were electrophoresed on a 1% (w/v) agarose gel containing 10 µg/mL CYBR Safe. M: Solis BioDyne. Arrow shows amplicon at ~921 bp.

3.5 Recombinant expression, solubilising and purification of *TbCATL*

The amino acid sequence was used to estimate the molecular weight and pI of a protein using ExPASY computer software (Gasteiger *et al.*, 2005). The molecular weight of the protein was 33.74 kDa and the pI 4.98. Expression from A1pET28a-*TbCATL* clone 2 and A7pET32a *TbCATL* clone 7 was performed at 16 ºC with IPTG induction. After the expression, *Tb*CATL protein bands at approximately 43 kDa from A1pET28a-*TbCATL* and 55 kDa from A7pET32a-*TbCATL*, respectively, were observed after staining the reducing SDS-PAGE gel with Coomassie blue R-250 (Figure 3.7, panels A and B). The identity of the His-tagged recombinant protein was confirmed by western blotting using mouse anti-His antibodies and the goat anti-mouse-HRPO with 4-chloro-1-naphthol/H_2O_2 detection system (Figure 3.7, panel C).

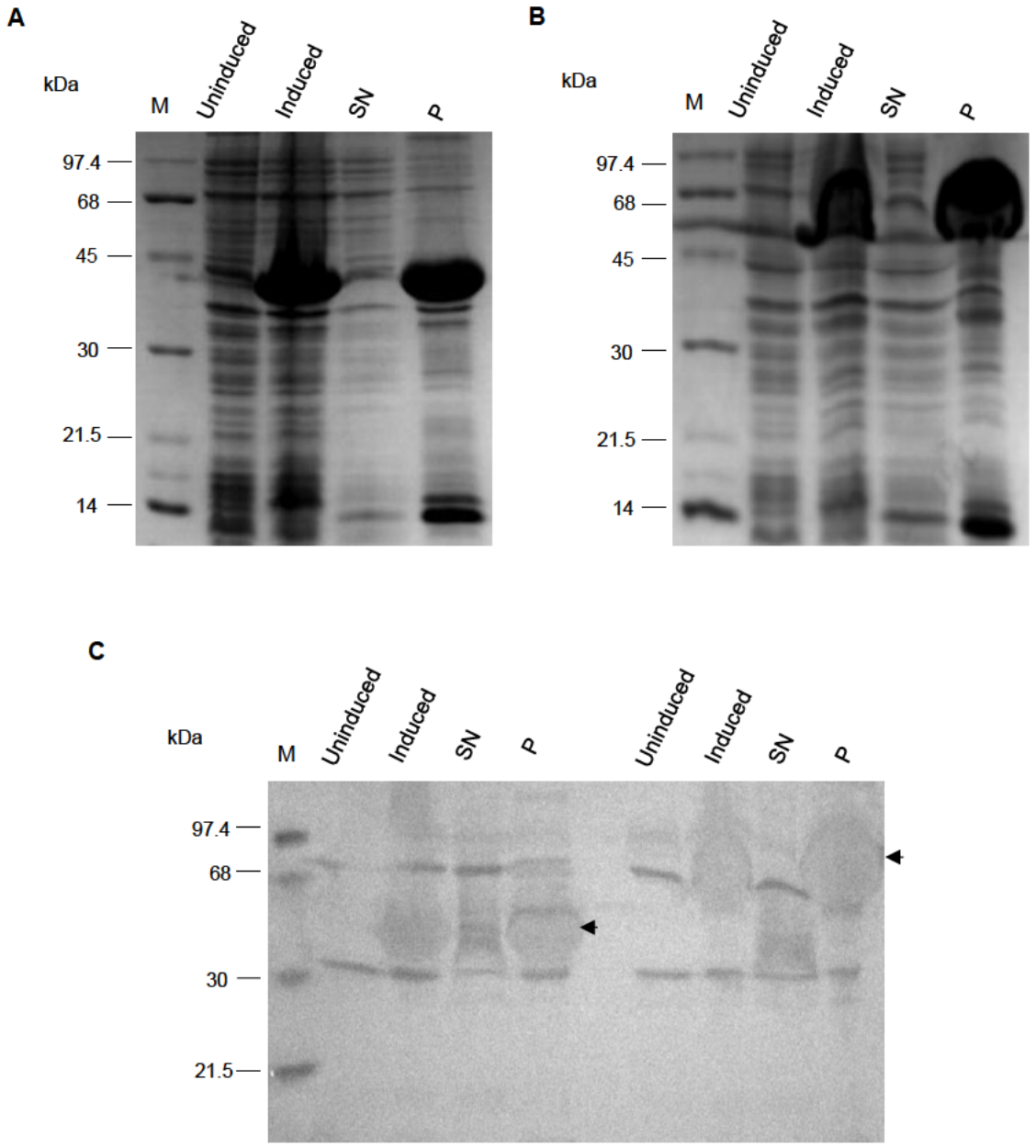

Figure 3.7: Reducing SDS-PAGE and western blot analysis of recombinantly expressed *Tb*CATL in pET28a and pET32a. Expression samples from the expression of **(A)** A1pET28a-*TbCATL* clone 2 and **(B)** A7pET32a-*TbCATL* clone 7 at 16 ºC, induced with 0.3 mM IPTG, were electrophoresed on 12.5% reducing SDS-PAGE gels with both **(A)** & **(B)** stained with Coomassie blue R-250 and proteins from the other **(C)** transferred onto nitrocellulose, blocked with 0.5% (w/v) milk-TBS and incubated with mouse anti-His antibodies [1:1500 in 0.5% (w/v) BSA-TBS]. Goat anti-mouse HRPO conjugate [1:5000 in 0.5% (w/v) BSA-PBS] and 4-chloro-1-naphthol/H_2O_2 were used to identify the expressed protein. M: Spectra™ unstained Broad Range Protein Ladder. SN: supernatant. P: pellet.

Recombinant *Tb*CATL expressed predominantly within the inclusion bodies (Figure 3.7, panel A and B). Therefore, the inclusion bodies from pET28a-*TbCATL* were solubilised using the sarkosyl method, followed by refolding, and purifying the recombinant *Tb*CATL on a cobalt IMAC column. After collection of the unbound proteins, the N-terminal 6 x His tagged *Tb*CATL

only eluted upon the application of the elution buffer after extensive washing of the cobalt resin. Contaminants co-eluted in fraction 2, but the rest of the eluted fractions were relatively pure without bacterial contamination (Figure 3.8, panel A, lanes3-10). The western blot with anti-His antibodies confirmed the identity of the purified proteins (Figure 3.8, panel B).

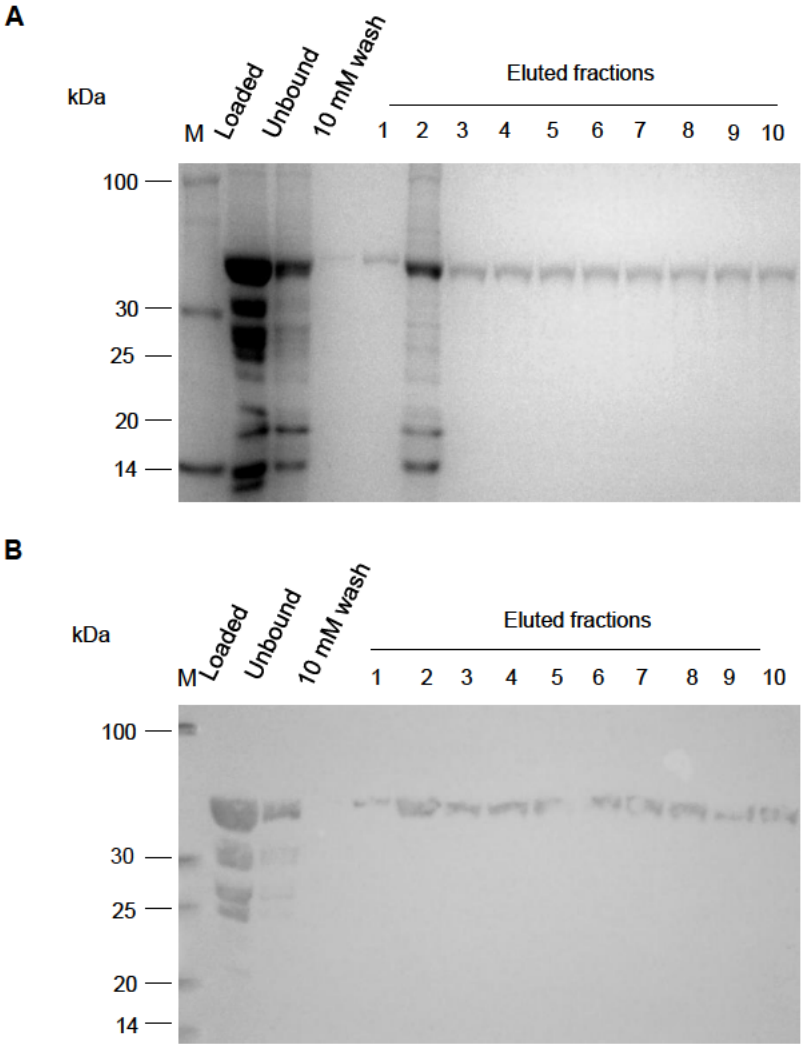

Figure 3.8: Cobalt IMAC on-column refolding and purification of sarkosyl-solubilised recombinant *Tb*CATL. Samples from insoluble expression and fractions solubilised with the sarkosyl method and purified using a cobalt IMAC purification were electrophoresed on 12.5% reducing SDS-PAGE gels and **(A)** stained with Coomassie blue R-250. proteins from the other gel **(B)** were transferred onto nitrocellulose, blocked with 5% (w/v) milk-TBS, and incubated with mouse anti-His antibodies [1:1500 in 0.5% (w/v) BSA-PBS]. The goat anti-mouse HRPO conjugate [1:5000 in 0.5% (w/v) BSA-PBS] and 4-chloro-1-naphthol/H_2O_2 were used to identify the purified protein. M: Spectra™ unstained Broad Range Protein Ladder.

Following the cobalt IMAC purification, fractions 3-9 were pooled, dialysed, and concentrated with PEG (20 kDa). The concentrated sample showed a single band at 45 kDa in a reducing

SDS-PAGE gel (Figure 3.9A). The identity of the 45 kDa protein was confirmed with the anti-His antibodies on a western blot (Figure 3.9B).

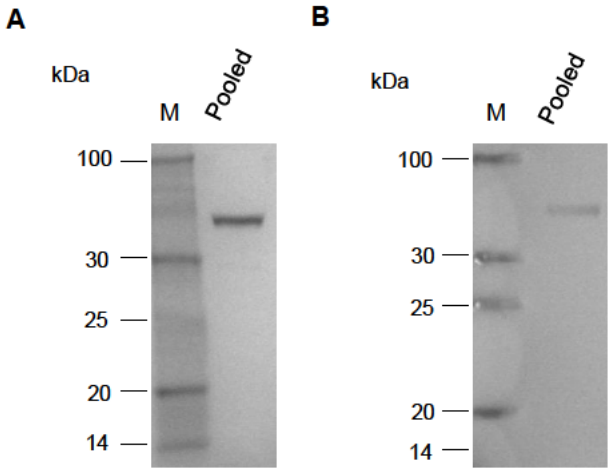

Figure 3.9: Concentration of recombinant *Tb*CATL by dialysing against PEG20kDa. After purification of recombinant T*b*CATL the cobalt metal-chelate affinity column, homogenous fractions were pooled, dialysed, and concentrated against PEG20 kDa. The PEG concentrated sample was electrophoresed on 10% SDS-PAGE gel one **(A)** stained with Coomassie blue R-250. The other gel **(B)** transferred into nitrocellulose, blocked with 5% (w/v) milk-TBS, and incubated with anti-His [1:1500 in 0.5% (w/v) BSA-TBS]. Goat anti-mouse HRPO conjugate [1:5000 in 0.5% (w/v) BSA-PBS] and 4-chloro-1-naphthol H_2O_2 used to identify the purified protein. M: Spectra™ unstained Broad Range Protein Ladder.

3.6 Enzymatic characterisation of *Tb*CATL

3.6.1 Gelatin SDS-PAGE

After purification of the recombinant *Tb*CATL using cobalt IMAC chromatography, the gelatinolytic activity of the recombinant *Tb*CATL was observed through hydrolysis of gelatin contained in a non-reducing SDS-PAGE gel as evidenced by a clear zone in the amido black-stained background on the gel (Figure 3.10). The recombinant *Tb*CATL was active in the gelatin substrate SDS-PAGE gel after incubation overnight at 37 ºC, in *Tb*CATL assay buffer containing reducing agent at pH 5. The gel showed that the non-reducing conditions resulted in the migration of *TbCATL* at both a higher molecular weight size of approximately 100 kDa as compared to its size under reducing conditions (2-mercaptoethanol and boiling samples) as well as activity at 45 kDa (Figure 3.9).

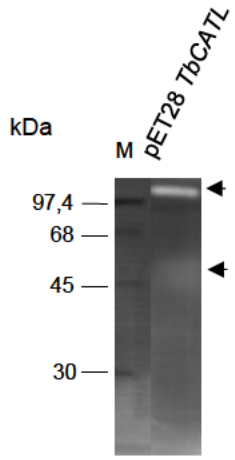

Figure 3.10: Gelatin substrate-containing non-reducing 12.5% SDS-PAGE to show the activity of recombinantly expressed and purified *Tb*CATL. Pooled fractions from cobalt metal-chelate affinity purification were dialysed, concentrated and electrophoresed on a non-reducing 12.5% SDS-PAGE containing 1% (w/v) porcine gelatin. After washing twice with 2.5% (v/v) Triton X-100 and incubation with *Tb*CATL assay buffer, pH 5 the gel was stained with amido black. M: Spectra™ unstained Broad Range Protein Ladder (please note that molecular sizes are estimations on a non-reducing SDS-PAGE gel). Arrows indicate proteolytic digestion.

Replicate samples of recombinant *Tb*CATL separated on a gelatin containing SDS-PAGE gel were incubated with different protease catalytic class specific inhibitors, both reversible and irreversible, to assess which would affect the activity of the enzyme. As shown in Figure 3.11, only the cysteine (cystain, E-64) and dual specificity cysteine/serine (antipain, chymostatin and leupeptin) protease inhibitors were able to inhibit the activity of recombinant *Tb*CATL. The aminopeptidase inhibitor (bestatin) serine protease inhibitor (SBTI) and metalloprotease inhibitor (EDTA) did not inhibit the gelatinolytic activity of *Tb*CATL.

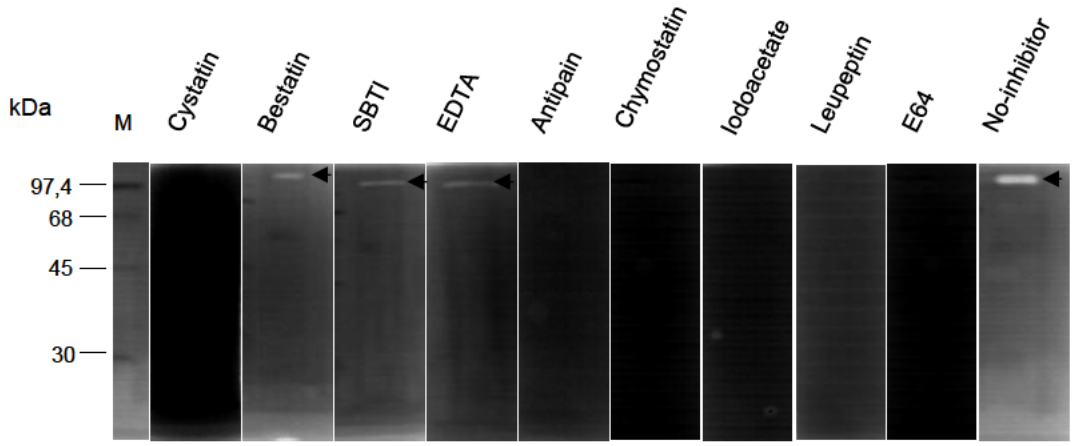

Figure 3.11: The Effect of inhibitors on the gelatinase activity of recombinant *Tb*CATL. PEG concentrated sample was electrophoresed on non-reducing 12.5% SDS-PAGE containing 1% (w/v) procine gelatin. After washing twice with 2.5% (v/v) Triton X-100 gel strips were incubated overnight at 37 °C in *TbCATL* assay buffer, pH 5 supplemented with respective inhibitors and one without inhibitors (positive control). Different concentrations of inhibitors used: cystatin (1µM), bestatin (10 µM), SBTI, EDTA (10 mM), antipain (10 µM), chymostatin (10 µM), iodoacetate (100 µM), leupeptin (100 µM), E-64 (10 µM). M: Spectra™ unstained Broad Range Protein Ladder.

3.6.2 pH optimum

The pH optimum of the purified recombinant *Tb*CATL was observed via hydrolysis of the gelatin substrate and measuring the rate of hydrolysis of Z-Phe-Arg-AMC in constant ionic strength AMT buffers at different pH values. The activity was visualised by the digestion of the gelatin substrate during incubation at pH 4 to pH 9, shown as clear zones in the amido black stained gelatin-containing non-reducing SDS-PAGE gel (Figure 3.12). The recombinant *Tb*CATL was therefore active across a wide pH range (4-9).

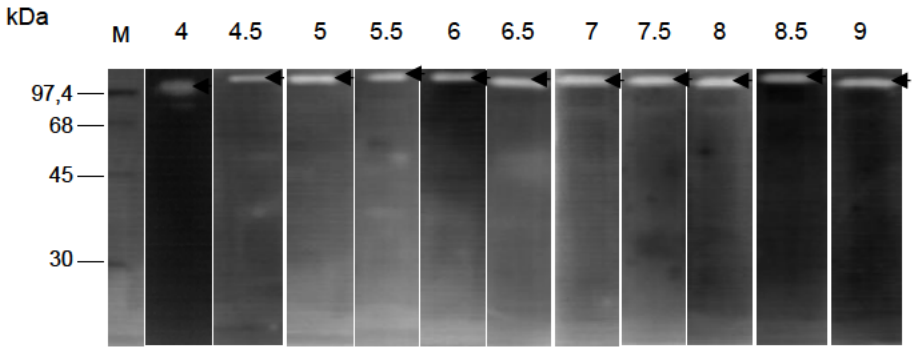

Figure 3.12: The effect of pH on the gelatinase activity of recombinant *Tb*CATL. The purified recombinant *Tb*CATL was electrophoresed on non-reducing 12.5% SDS-PAGE containing 1% (w/v) porcine gelatin. After washing twice with Triton X-100 (v/v) and incubation with the respective AMT buffers and stained with amido black. M: Spectra™ unstained Broad Range Protein Ladder.

The hydrolysis of the substrate Z-Phe-Arg-AMC by the recombinant showed maximal activity at pH 5, with approximately 80% of maximal activity observed at pH 5.5 (Figure 3.13). About 50% of maximal activity was observed at pH 4, 4.5 and 6, but at neutral and alkaline pH values *Tb*CATL showed no peptidolytic activity.

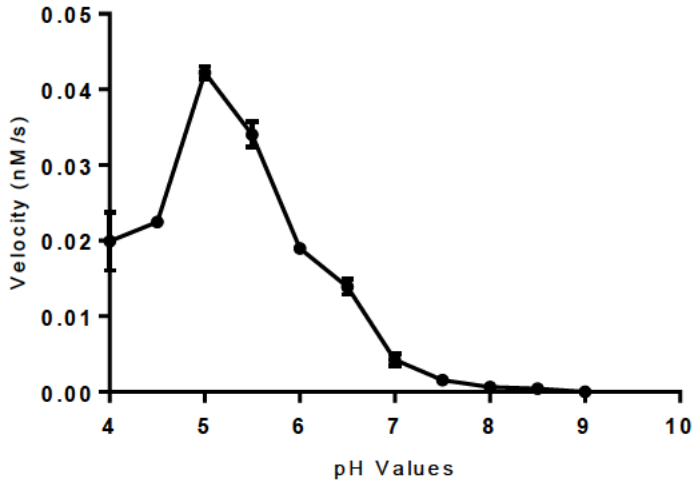

Figure 3.13: The pH profile of recombinant *Tb*CATL. Duplicate samples of recombinant *TbCATL* were incubated in constant ionic strength AMT buffers containing 6 mM DTT and incubated at 37 °C for 10 min, followed by the addition of 20 µM Z-Phe-Arg-AMC, before the hydrolysis of AMC was measured, over 15 min at Ex_{360} and Em_{460}.

3.6.3 Substrate specificity

The ability of the recombinant *Tb*CATL to hydrolyse different fluorogenic peptide substrates was observed through the reaction of *Tb*CATL with a range of substrates that vary in the amino acid residues in P_1, P_2 and P_3. Table 3.1 shows that Z-Phe-Arg-AMC is the best substrate for *Tb*CATL, as indicated by the lower K_m value of 18.6 µM and higher catalytic efficiency (k_{cat}/K_m) value of 5.87 s^{-1} mM^{-1}. Z-Leu-Arg-AMC (k_{cat}/K_m = 8.39 s^{-1} mM^{-1}) and H-Pro-Phe-Arg-AMC (k_{cat}/K_m = 5.4 s^{-1} mM^{-1}) are also good substrates for *Tb*CATL. H-Ala-Phe-Lys-AMC had a relatively low catalytic efficiency (k_{cat}/K_m) value of 3.83 s^{-1} mM^{-1} and a higher K_m value of 75.67 µM. These results show pronounced hydrolysis of the peptide substrates by *Tb*CATL if the basic amino acids Arg or Lys occur at the P_1 substrate position that binds to the S_1 position of the active site cleft and large hydrophobic amino acids such as Phe or Leu at the P_2 substrate position, which binds to the hydrophobic S_2 of the active site cleft.

The recombinant *Tb*CATL could not hydrolyse peptide substrates containing the small hydrophobic amino acids such as Gly or Val at P_3 and an arginine-rich substrate Z-Arg-Arg-AMC. Therefore, *Tb*CATL preferred to cleave peptide bonds when the P_2 position is occupied by hydrophobic amino acid residues such as Phe or Leu, and P_1 preference is Arg or Lys.

Table 3.1: Substrate specificity of recombinant *Tb*CATL.

Substrate	K_m (µM)	V_{max}	k_{cat} (S^{-1})	k_{cat}/K_m (S^{-1} mM^{-1})
Z-Phe-Arg-AMC	18.60	0.02504	0.1091	5.87
Z-Leu-Arg-AMC	36.60	0.06954	0.3030	8.39
H-Pro-Phe-Arg-AMC	59.13	0.07333	0.3195	5.40
H-Ala-Phe-Lys-AMC	75.67	0.06651	0.2898	3.83

Substrates with no activity even after extended incubation of the enzyme at 37 °C: Z-Arg-Arg-AMC, Z-Gly-Gly-Arg-AMC, Z-Pyr-Gly-Arg-MCA

3.6.4 Effect of inhibitors on Z-Phe-Arg-AMC hydrolysis by *Tb*CATL

The recombinant *Tb*CATL was incubated with both reversible and irreversible inhibitors to determine which would influence enzyme activity. As shown by Figure 3.14, only the cysteine and cysteine/serine protease specific inhibitors, antipain, leupeptin, chymostatin, cystatin, E-64, iodoacetate, were able to inhibit the activity of recombinant *Tb*CATL by more than 90%. A serine specific inhibitor, soybean trypsin inhibitor and a metallo-specific inhibitor, EDTA, inhibited the recombinant *Tb*CATL by more than 50%. The aminopeptidase specific inhibitor, bestatin, only showed less than 20% inhibition of the recombinant *Tb*CATL.

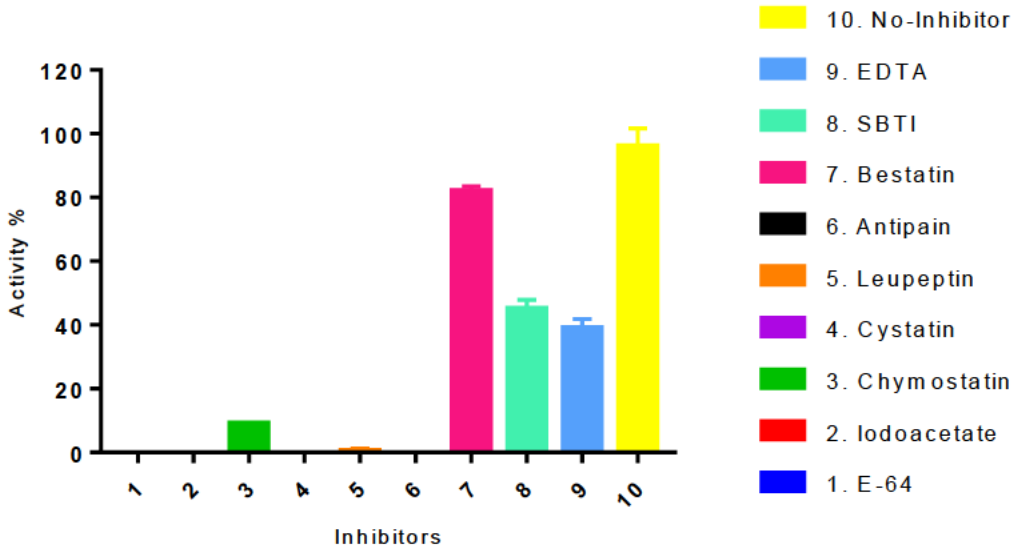

Figure 3.14: Protease class-specific inhibition of recombinant *Tb*CATL. Recombinant *Tb*CATL (1 µM) was incubated in *Tb*CATL assay buffer containing different protease inhibitors: cystatin (1 µM), bestatin (10 µM), SBTI, EDTA (10 mM), antipain (10 µM), chymostatin (10 µM), iodoacetate (100 µM), leupeptin (100 µM), E-64 (10 µM). Substrate, 20 µM Z-Phe-Arg-AMC, was added before the hydrolysis was measured, over 15 min at $Ex_{360\ nm}$ and $Em_{460\ nm}$. Percentage inhibition was represented using assays in the presence of no inhibitor as 100% activity.

3.7 DNA isolation and amplification

Good quality genomic DNA (concentration, not shown, was determined on a Nano drop) was isolated from fish blood samples using a DNA extraction kit. The isolated genomic DNA samples were used as a template for PCR amplification of the trypanosome SSU rRNA gene using primers designed by Sehgal *et al.* (2001). As shown in Figure 3.15, panels A and B, the genomic DNA size was larger than the molecular weight marker band (10000 bp).

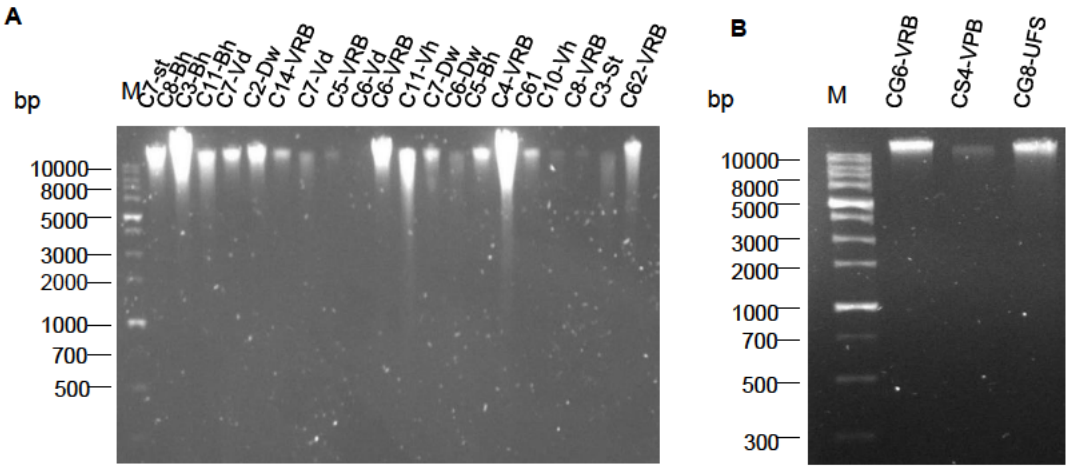

Figure 3.15: Genomic DNA isolated from fish blood samples. Following the isolation of genomic DNA from 23 fish blood samples. Genomic DNA samples were electrophoresed on a 1% (w/v) agarose gel containing 10 µg/mL CYBR safe. M: Solis BioDyne DNA ladder.

The PCR amplification of the SSU rRNA gene from the isolated DNA samples resulted in a single band with a size of approximately 340 bp (Figure 3.16, panels A and B). There was no amplification observed in the negative control. Non-specific binding was detected in 17 of 23 amplified samples.

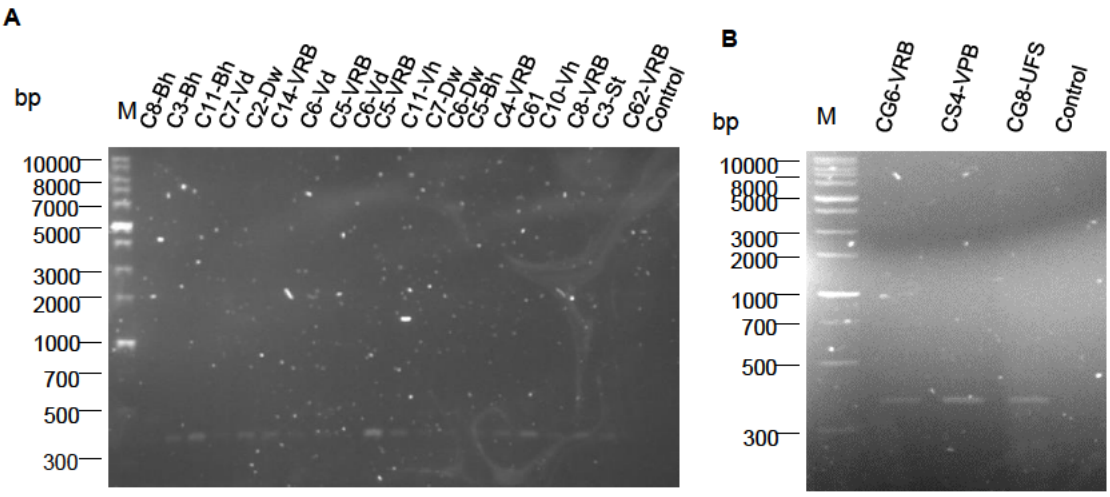

Figure 3.16: Screening for trypanosomes from fish genomic DNA by PCR amplification. The SSU rRNA gene of trypanosomes was amplified from fish genomic DNA using primers designed by Sehgal et al. (2001). Samples were electrophoresed on 1% (w/v) agarose gel containing 10 µg/mL CYBR safe. M: Solis BioDyne DNA ladder.

3.8 Phylogenetic analysis

Out of 24 samples sent for sequencing at CAF, only 6 were usable for phylogenetic analysis. The final alignment contained 38 sequences including those retrieved from GenBank, and each sequence had 323 base pairs. The best-fit substitution model used in ML and BI was GTR, (partition = 012345, -InL = 732.9460, K = 82). The trees produced by ML and BI were consistent and no conflict was detected. The ML bootstrap values and BI posterior probability values were both annotated onto the most likely tree (Figure 3.17). Only bootstrap (BS) values > 60% and posterior probability (PP) > 0.60 are annotated onto the most likely tree. The phylogenetic analyses recovered two main lineages. One lineage containing trypanosomes collected from freshwater fish species (branch support: ML bootstrap = 89%; BI posterior probabilities = 0.94) and a second lineage including parasites collected from marine fish (branch support: ML bootstrap = 100%; BI posterior probabilities = 0.94). All trypanosomes collected from *Clarias gariepinus* fell within the freshwater clade but did not form a well-supported monophyletic clade. The six sequences were not identical, and four haplotypes were recovered. Only one of these haplotypes (C11_VH) was also shared with a trypanosome previously isolated from *Sander lucioperca* (KJ601723).

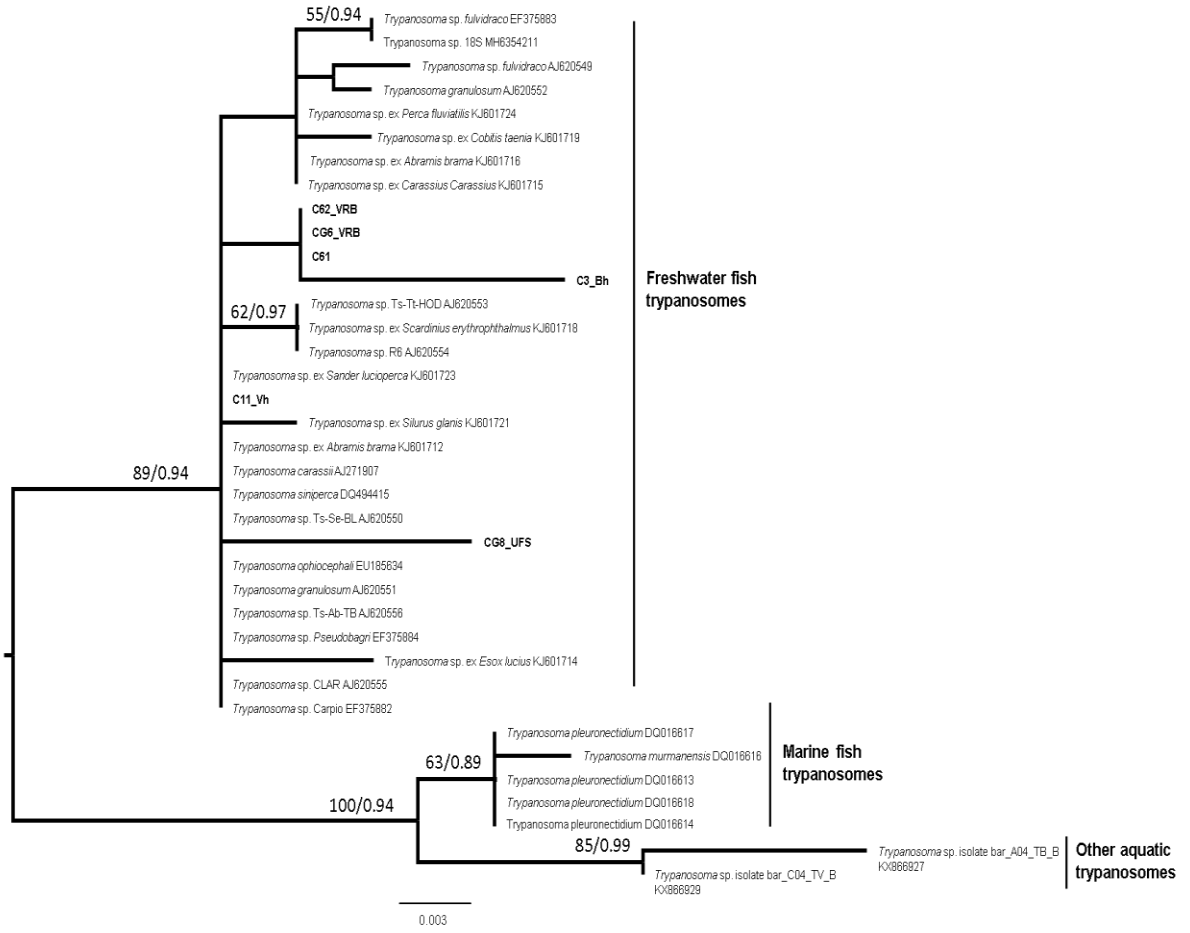

Figure 3.17: Maximum likelihood phylogeny of fish trypanosomes constructed from small subunit ribosomal RNA (SSU rRNA) sequences. Maximum likelihood (1000 replicates) bootstrap support values and Bayesian posterior probabilities (20 million generations) are shown on the nodes. GenBank accession numbers and species names of the present study (bold) are annotated onto the branches.

4 CHAPTER 4: GENERAL DISCUSSION

Trypanosomes infect all classes of vertebrates, including humans, animals, fish and birds. Whereas leeches transmit fish trypanosomes, in sub-Saharan Africa tsetse flies (*Glossina* spp.) transmit trypanosomes to their mammalian hosts (Grybchuk-Ieremenko *et al.*, 2014; Büscher *et al.*, 2017), resulting in health issues to humans and affecting livestock production. African trypanosomes' ability to periodically switch their VSGs, allow them to evade their hosts' immune system and develop resistance to the drugs (Nok, 2009; Horn, 2014). Over the years, controlling trypanosomiasis has been based on vector control and therapeutic treatment of the infected host, however, all these techniques are either expensive, or parasites are resistant to the administered drugs (Baral, 2010). A major factor that contributes to the pathogenesis of trypanosomes is the activity of proteases that are involved in different biological processes in parasites. They have been identified as promising candidates for developing chemotherapeutic trypanosome-specific medicines (Steverding, 2013).

The cathepsin L-like protease from *T. b brucei* and the phylogeny of fish trypanosomes were the focus of the present study. Cathepsin L belongs to clan CA of the C1 peptidase family. The catalytic activity of clan CA peptidases depend on the dyad of Cys and His residues in the active site (Caffrey and Steverding, 2009). However, Asn or Asp residues play a vital role to orientate the imidazole ring of the His residue. Clan CA consists of more than 20 families, and they are grouped into clans based on their structures and similarities in the sequence motifs around the catalytic residues. The C1 family is subdivided into C1A (papain subfamily) and C1B (bleomycin hydrolase subfamily). The C1A family consists of endopeptidases, mostly lysosomal enzymes present in viruses, protozoa, plants and animals. The exopeptidases are found in Gram-positive bacteria, fungi and animals (Rawlings and Barrett, 2013).

Cysteine proteases have been implicated as pathogenic factors during trypanosome infection, and they are involved in the manifestation of disease symptoms (Caffrey and Steverding, 2009). Cathepsin L-like and Cathepsin B-like cysteine proteases are present in *Trypanosoma* parasites (O'Brien *et al.*, 2008). The cathepsin L-like cysteine protease from *T. brucei* (*Tb*CATL) and *T. cruzi* (*Tcr*CATL) have been shown to express in high levels in the bloodstream forms and during all life-cycle stages, respectively. *Tb*CATL plays a crucial role in the survival of the parasites (Caffrey and Steverding, 2009; Siqueira-Neto *et al.*, 2018) and is involved in virulence, host immune evasion, and allows crossing the human host's blood-brain barrier, causing the disease's late-stage symptoms (Steverding, 2013; Roberta *et al.*, 2016). *Tcr*CATL of *T. cruzi*, an intracellular parasite that causes Chagas disease in South America, is essential in pathogenesis, host immune evasion, invasion of host cells and parasite differentiation (Siqueira-Neto *et al.*, 2018). *Tco*CATL of the animal-infective *T.*

congolense is expressed throughout the life-cycle and is responsible for the parasites' pathogenesis and survival (Lalmanach *et al.*, 2002). *Tb*CATB of *T. brucei* is expressed in high levels in the bloodstream stage and is involved in the degradation of host transferrin, essential for iron acquisition by the parasite (McKerrow, 2006; Roberta *et al.*, 2016). *Tco*CATB of *T. congolense* is found in the bloodstream forms, and it plays a vital role in the pathogenesis and the life-cycle of parasites (Mendoza-Palomares *et al.*, 2008). Several studies have been conducted on cysteine proteases of trypanosomes in order to find possible inhibitors that could be used as drugs against these disease-causing pathogens. An example is the diazomethane inhibitor, Z-Phe-Ala-CHN_2 that is an inhibitor of cathepsin L-like and cathepsin B-like proteases and kills trypanosomes (O'Brien *et al.*, 2008). Although possible inhibitors and drugs of trypanosomes were identified, more studies on proteases still need to be conducted to find effective and affordable drugs, especially for the livestock disease, nagana (Steverding, 2013).

Cathepsin L is usually produced as a zymogen, containing a prodomain and a mature domain. The prodomain has different functions, such as assisting with protein folding and regulating the enzyme by inhibiting the enzyme's activity. The prodomain keeps the enzyme in an inactive state until it reaches its destination. For a cathepsin L to be fully activated, it requires removal of the prodomain (Yamamoto *et al.*, 2002; Brömme *et al.*, 2004). The prodomain of *T. cruzi* and *Leishmania* cathepsin L is removed in the lysosomes or in endosomal compartments (Koeller and Bangs, 2019). The prodomain can be removed through different methods, such as by another protease or autocatalitically by acidic pH, in the presence of a reducing agent (Brömme *et al.*, 2004). In *Pichia pastoris*, the cysteine protease of *T. brucei* and *T. congolense* is recombinantly produced in a soluble but inactive form, requiring purification and activation by acidic pH in the presence of a reducing agent (Caffrey *et al.*, 2001; Boulangé *et al.*, 2011).

In the present study, the gene coding for the full-length CATL from *T. b brucei* (*Tb*CATL) was cloned into an *E. coli* pET-28a expression vector and expression induced at 16 ^{o}C. The *E. coli* BL21 DE3 strain contains an IPTG inducible chromosomal copy of the T7 RNA polymerase gene (DE3). At lower temperatures, such as 16 ^{o}C, the slower rate of proteins expression gives the proteins time to refold properly, and this results in proteins being produced in a soluble form (Lebendiker and Danieli, 2014). *Tb*CATL was successfully expressed as an approximately 45 kDa protein in the *E. coli* (BL21 DE3) expression system. However, the enzyme was expressed in an insoluble form as inclusion bodies. When proteins are expressed in high amounts, it results in proteins being expressed within inclusion bodies to protect the host organism from potentially toxic proteins. This is usually the case when expressing cysteine proteases in the *E. coli* (BL21 DE3) expression system (Brömme *et al.*, 2004). The recombinant *Tb*CATL is insoluble, like the *Tc*CATL requires solubilisation by denaturation before being refolded and purified on a suitable affinity chromatography column (Yamaguchi

and Miyazaki, 2014; Koeller and Bangs, 2019). A recombinant cysteine protease of *Leishmania mexicana* has been produced in an insoluble form in *E. coli* and required solubilisation under denaturing conditions, refolding, and purification on-column in the presence of a reducing agent, prior activation by acidic pH environment (Sanderson *et al.*, 2000).

A fusion tag is often used for the recombinant expression of proteins to aid in affinity purification and improve the protein's solubility (Costa *et al.*, 2014). A 6-His tagged expression vector was used in the present study to aid protein purification. Fortuitously, IMAC chromatography that is used to purify His-tagged proteins is compatible with denaturation agents (Petty, 2001) and can therefore be used directly after solubilisation of a recombinant protein. Several different metal ions can be used to purify the 6-His tagged proteins using IMAC (Kimple *et al.*, 2013). In the present study, recombinant *Tb*CATL was purified using Co^{2+} IMAC, and most of the impurities were removed after solubilisation of the protein from the inclusion bodies with the sarkosyl method (Singh *et al.*, 2015). After the purification of the recombinant *Tb*CATL, purity was assessed on a Coomassie blue-stained reducing SDS-PAGE gel and by western blotting using anti-His antibodies that would only detect His-tagged recombinant protein (Pamer *et al.*, 1991; Kimple *et al.*, 2013). A homogenous preparation of *Tb*CATL was shown at approximately 45 kDa in both the reducing SDS-PAGE gel and western blot. These results are comparable to those obtained by Abdulla *et al.* (2008), who demonstrated the native *Tb*CATL on western blot at approximately 45 kDa and 47 kDa. However, the results obtained from SDS-PAGE gel analysis of the recombinant *Tb*CATL in the present study, were inconsistent with those shown by Caffrey *et al.* (2001), who demonstrated expression of the full length form from the soluble fraction, with approximately 36 kDa.

Following desalting of recombinant *Tb*CATL, biochemical studies were conducted on the purified enzyme using various enzyme assays to determine gelatinolytic and peptidolytic activity, optimum pH for hydrolysis of protein and peptide substrates, peptide substrate specificity and sensitivity of *Tb*CATL to various protease class-specific inhibitors. In the present study, the recombinant *Tb*CATL indicated that, even though the enzyme was expressed in an insoluble form, it can hydrolyse different peptide substrates after solubilisation, and following activation showed maximum activity at acidic pH, like the native *Tb*CATL reported by Troeberg *et al.* (1996). The activity of recombinant *Tb*CATL was induced by acidic pH in the presence of DTT, which is responsible for maintaining the active site cysteine residue in a reduced state (Mbawa *et al.*, 1992). In the present study, the recombinant *Tb*CATL showed protease activity when analysed on a gelatin-containing SDS-PAGE gel, like the cysteine protease of *Leishmania* (Sanderson *et al.*, 2000). However, the size of the

recombinant *Tb*CATL was higher than when analysed on the reducing SDS-PAGE gel (approximately 45 kDa). This may be due to the gelatin restricting the mobility of the protein. It is also known that under non-reducing conditions, the mobility of proteins depend on their charge to mass ratio and not exclusively on their size because the SDS is not able to bind uniformly to the unreduced protein (Heussen and Dowdle, 1980). The recombinant cysteine protease from *T. b. rhodesiense* also demonstrated gelatinolytic activity and hydrolysis of different peptide substrates, including Z-Phe-Arg-AMC (Pamer *et al.*, 1991).

In the present study, the optimum pH for the hydrolysis of gelatin and Z-Phe Arg-AMC by recombinant *Tb*CATL was determined using the constant ionic strength AMT buffers to ensure the effect of ionic strength, that could be considerable, does not mask the effect of pH as shown for mammalian cathepsin L (Dehrmann *et al.*, 1995). The recombinant *Tb*CATL showed activity across a wide pH range (4 to 9) on the gelatin-containing gel, but maximal activity against Z-Phe-Arg-AMC was observed in a much narrower pH range of 4.5 – 5.5, with optimal activity at pH 5. However, it is important to note that the activity against the protein substrate extended across a wider pH range, including physiological pH. These results are consistent with those obtained by Caffrey *et al.* (2001), where the pH optimum for Z-Phe-Arg-AMC hydrolysis by the recombinantly expressed catalytic domain of a cysteine protease from *T. b. rhodesiense* was at pH 5, with 50% activity remaining at physiological pH. These results show that CATL can be active and retain high activity in acidic and physiological environments (Caffrey *et al.*, 2001; Mendoza-Palomares *et al.*, 2008). The activity of CATL in acidic and physiological pH shows that it would be active in the bloodstream of the infected host and in the lysosomal compartments, as demonstrated by Mbawa *et al.* (1992). In *T. congolense* (*Tco*CATL) infected rats' bloodstream, the bloodstream and metacyclic forms expressed high levels of cysteine protease (Mbawa *et al.*, 1992).

In the present study, the substrate specificity of recombinant *Tb*CATL was investigated using fluorogenic peptide substrates with different residues in the P1 and P2 positions. The recombinant *Tb*CATL was found to hydrolyse substrates with hydrophobic amino acids Phe or Leu at the P2 position, which is typical for most mammalian or protozoan family C1 clan CA cysteine proteases. A further preference for the basic residues, Arg or Lys in P1 shows that the enzyme is more like cathepsin L than cathepsin B (Mbawa *et al.*, 1992; Caffrey *et al.*, 2001).

The specificity of papain-like proteases is determined by the interaction of the substrate's P2 and the enzyme's S2 binding pocket. The cathepsin-like protease's S2 binding pocket prefers bulky, hydrophobic amino acids of the substrate such as Phe or Leu in the P2 position (Steverding, 2015). In contrast, cathepsin B prefers Arg at P2 of the substrate (Mbawa *et al.*,

1992). The preference for a bulky, hydrophobic amino acid residue in P2 of the substrate is supported by the inability of the recombinant *Tb*CATL to hydrolyse peptide substrates with the basic Arg or small hydrophobic Gly amino acid residues in P2 of the substrate. Therefore, the active site of recombinant *Tb*CATL preferred Z-Phe-Arg-AMC, Z-Leu-Arg-AMC, H-Pro-Phe-Arg-AMC, and H-Ala-Phe-Lys-AMC shown by a relatively high k_{cat}/K_m ratios. The substrate containing Leu at P2 had a higher k_{cat}/K_m ratio over all the substrates used, and the substrate containing Ala at P2 had a lower k_{cat}/K_m value. All substrates containing Arg at P1 showed higher k_{cat}/K_m values than those containing Lys at P1, showing that CATL has a higher preference for Arg or Lys at P1. The inability of the recombinant *Tb*CATL to cleave Z-Arg-Arg-AMC shows that this protease is not a cathepsin B (Mbawa *et al.*, 1991).

In the present study, the inhibition of the recombinant *Tb*CATL was analysed using gelatin-containing non-reducing SDS-PAGE and H-Pro-Phe-Arg-AMC and different reversible and irreversible class-specific inhibitors: antipain, chymostatin, bestatin, leupeptin, SBTI, EDTA, cystatin, iodoacetate and E-64. The gelatinase and fluorogenic substrate activity of recombinant *Tb*CATL was inhibited by antipain, chymostatin, and leupeptin containing a basic amino acid Arg at P1. This confirms that recombinant *Tb*CATL prefers a basic Arg residue at P1. This inhibition is also reported for *Tco*CATL (Boulangé *et al.*, 2011). The activity of recombinant *Tb*CATL on fluorogenic substrates and gelatin was not inhibited by bestatin which contains Phe at P1, suggesting that Phe is not a preferred residue at P1. This suggests that recombinant *Tb*CATL is not an aminopeptidase. The enzyme showed a slight sensitivity towards the addition of the metalloprotease inhibitor EDTA, but it was not inhibited completely. The serine specific inhibitor SBTI also showed a slight effect on the recombinant *Tb*CATL activity. The reason behind this is unclear. The effective inhibition of recombinant *Tb*CATL hydrolytic activity against the fluorogenic substrates and gelatin by cysteine specific inhibitors such as E-64, iodoacetate and cystatin further confirms that recombinant *Tb*CATL is a cysteine protease.

Fish trypanosomes were among the first *Trypanosoma* species described, and they pose a severe threat to aquaculture production (Rodrigues *et al.*, 2018). However, more attention has been given to mammalian trypanosomes. The lack of biological information about fish trypanosomes is a major issue, especially because these parasites result in about 70 to 100% of the infected fish hosts' deaths (Grybchuk-Ieremenko *et al.*, 2014). Fish trypanosomes are close phylogenetic relatives of African trypanosomes. A taxonomic analysis of trypanosomes used to be based on two criteria: the position of the kinetoplast and nucleus and host specificity (Grybchuk-Ieremenko *et al.*, 2014; Hayes *et al.*, 2014). However, these characteristics are unreliable in the phylogenetic analysis because the kinetoplast or nucleus position may change as parasites grow. A single trypanosome species can infect different fish species (Gu

et al., 2010; Grybchuk-Ieremenko *et al.*, 2014) In this study, the genetic diversity of trypanosomes found in *Clarias gariepinus* collected from a single river system was investigated. In particular, the phylogenetic analysis was used to examine the host specificity of *Clarias gariepinus* trypanosomes. An important question addressed was whether parasites that are highly host-specific would occur in a single host species, whereas generalist parasites will be found in hosts from several different fish species.

The issue of inconsistency in results of trypanosomes taxonomy obtained from the use of morphological features or host specificity can be solved by molecular taxonomy. The SSU rRNA gene proves to be a more reliable biomarker in the phylogenetic analysis of trypanosomes. The SSU rRNA gene is widely used in the molecular and phylogenetic analysis of fish trypanosomes (Grybchuk-Ieremenko *et al.*, 2014; Lemos *et al.*, 2015). Genomic DNA was isolated from 24 *Clarias gariepinus* blood samples, and the SSU rRNA *Trypanosoma* gene was amplified from the isolated genomic DNA. Only 6 SSU rRNA sequences obtained from the present study were usable for phylogenetic analysis. The maximum likelihood tree was constructed using the sequences obtained from the present study as well as sequences obtained from GenBank. The bootstrap (BS) values >60% and posterior probability (PP) >0.60 were annotated onto the most likely tree. The phylogenetic tree showed two major lineages, freshwater fish trypanosomes and marine water fish trypanosomes. All trypanosomes collected from *Clarias gariepinus* fell within the freshwater clade. Even though all six *Clarias gariepinus* species were collected from a single river system, trypanosomes infecting these hosts did not form a well-supported monophyletic clade, suggesting that this fish species may be infected by multiple lineages of trypanosomes. These results were inconsistent with the results obtained by Lemos *et al.* (2015), who reported a well-formed monophyletic clade of fish trypanosomes. Genetic variation from all SSU rRNA sequences obtained from the present study was observed, and four haplotypes were recovered.

In order to show the presence of trypanosomes in the fish blood samples, a portion of the SSU rRNA gene that is not very variable was targeted in the present study. However, to be of more value for phylogenetic analyses, primers will in future be designed for amplification of the more variable portion of the gene. The more conserved primers used in the present study may be most useful as qPCR primers to determine level of infection.

In the present study, the gene coding for *Tb*CATL was successfully cloned, recombinantly expressed, and *Tb*CATL purified to homogeneity. Enzymatic characterisation of recombinant *Tb*CATL was conducted to evaluate the pH optimum, substrate specificity and effect of class-specific inhibitors on substrate hydrolysis. From the results obtained from this work, it can be concluded that recombinant *Tb*CATL has a pronounced preference for hydrophobic residues

Phe or Leu at the P2 position and that all the cysteine protease-specific inhibitors inhibited the activity. The findings of this study suggest that recombinant *Tb*CATL is more likely to be a cysteine protease. The SSU rRNA was amplified from DNA isolated from *Clarias gariepinus,* and phylogenetic analysis of fish trypanosomes performed. All six trypanosomes of the present study belong to freshwater fish trypanosomes, and they formed a not a well-supported monophyletic clade. Four of six are originated from a single parent, and they infect the same host species of the same region, and therefore, they are likely to be host-specific. This study laid a foundation for conducting further studies on *Tb*CATL as a chemotherapeutic and a diagnostic target for AT. This preliminary study of the phylogeny of *Clarias gariepinus* trypanosomes provides a foundation for future studies of the evolution and diversification of these parasites in fishes.

References

Abdulla, M.-H., O'Brien, T., Mackey, Z. B., Sajid, M., Grab, D. J. & 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,** e298.

Alizadehrad, D., Krüger, T., Engstler, M. & Stark, H. (2015). Simulating the complex cell design of *Trypanosoma brucei* and its motility. *PLOS Computational biology,* **11,** e1003967.

Antoine-Moussiaux, N., Büscher, P. & Desmecht, D. (2009). Host-parasite interactions in trypanosomiasis: on the way to an antidisease strategy. *Infection and immunity,* **77,** 1276-1284.

Baral, T. N. (2010). Immunobiology of African trypanosomes: need of alternative interventions. *Journal of biomedicine & biotechnology,* **2010,** 389153-389153.

Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M. & Hanada, K. (1982). L-trans-epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. *The biochemical journal,* **201,** 189-98.

Barrett, A. J. & Kirschke, H. (1981). Cathepsin B, cathepsin H, and cathepsin L. *Methods in enzymology,* **80 Pt C,** 535-61.

Barrett, M. P., Burchmore, R. J., Stich, A., Lazzari, J. O., Frasch, A. C., Cazzulo, J. J. & Krishna, S. (2003). The trypanosomiases. *The Lancet,* **362,** 1469-1480.

Bastos, I. M. D., Motta, F. N., Grellier, P. & Santana, J. M. (2013). Parasite prolyl oligopeptidases and the challenge of designing chemotherapeuticals for chagas disease, leishmaniasis and African trypanosomiasis. *Current medicinal chemistry,* **20,** 3103-3115.

Borges, A. R., Lemos, M., Morais, D. H., Souto-Pedrón, T. & D'agosto, A. (2016). In vitro culture and morphology of fish trypanosomes from South American wetland areas. *SOJ microbiology & infectious diseases,* **4,** 1-5.

Boulangé, A., Pillay, D., Chevtzoff, C., Biteau, N., Comé de Graça, V., Rempeters, L., Theodoridis, D. & Baltz, T. (2017). Development of a rapid antibody test for point-of-care diagnosis of animal African trypanosomosis. *Veterinary parasitology,* **233,** 32-38.

Boulangé, A. F., Khamadi, S. A., Pillay, D., Coetzer, T. H. & Authié, E. (2011). Production of congopain, the major cysteine protease of *Trypanosoma* (*Nannomonas*) *congolense,* in *Pichia pastoris* reveals unexpected dimerisation at physiological pH. *Protein expression and purification,* **75,** 95-103.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry,* **72,** 248-254.

Briggs, G. E. & Haldane, J. B. (1925). A note on the kinetics of enzyme action. *The biochemical journal,* **19,** 338-339.

Brock, T. D. & Freeze, H. (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of bacteriology,* **98,** 289.

Brömme, D., Nallaseth, F. S. & Turk, B. (2004). Production and activation of recombinant papain-like cysteine proteases. *Methods,* **32,** 199-206.

Brun, R., Blum, J., Chappuis, F. & Burri, C. (2010). Human African trypanosomiasis. *The Lancet,* **375,** 148-159.

Burleigh, B. A. & Andrews, N. W. (1995). A 120-kDa alkaline peptidase from *Trypanosoma cruzi* is involved in the generation of a novel Ca2+-signaling factor for mammalian cells. *Journal of biological chemistry,* **270,** 5172-5180.

Büscher, P., Cecchi, G., Jamonneau, V. & Priotto, G. (2017). Human african trypanosomiasis. *The Lancet,* **390,** 2397-2409.

Caffrey, C. R., Hansell, E., Lucas, K. D., Brinen, L. S., Hernandez, A. A., Cheng, J., Gwaltney II, S. L., Roush, W. R., Stierhof, Y.-D. & Bogyo, M. (2001). Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*. *Molecular and biochemical parasitology,* **118,** 61-73.

Caffrey, C. R. & Steverding, D. (2009). Kinetoplastid papain-like cysteine peptidases. *Molecular and biochemical parasitology,* **167,** 12-19.

Cayla, M., Rojas, F., Silvester, E., Venter, F. & Matthews, K. R. (2019). African trypanosomes. *Parasites & vectors,* **12,** 190.

Chappuis, F., Loutan, L., Simarro, P., Lejon, V. & Büscher, P. (2005). Options for field diagnosis of human African trypanosomiasis. *Clinical microbiology reviews,* **18,** 133-146.

Coetzer, T. H., Goldring, J. D. & Huson, L. E. (2008). Oligopeptidase B: a processing peptidase involved in pathogenesis. *Biochimie,* **90,** 336-344.
Compton, S. J. & Jones, C. G. (1985). Mechanism of dye response and interference in the Bradford protein assay. *Analytical biochemistry,* **151,** 369-374.
Corrêa, L. L., Oliveira, M. S. B., Tavares-Dias, M. & Ceccarelli, P. S. (2016). Infections of *Hypostomus* spp. by *Trypanosoma* spp. and leeches: a study of hematology and record of these hirudineans as potential vectors of these hemoflagellates. *Revista Brasileira de parasitologia veterinária,* **25,** 299-305.
Costa, S., Almeida, A., Castro, A. & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Frontiers in microbiology,* **5,** 63.
Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods,* **9,** 772.
Davis, L. G., Dibner, M. D. & Battey, J. F. (1986). Section 5-5 - agarose gel electrophoresis. *In:* Davis, L. G., Dibner, M. D. & Battey, J. F. (eds.) *Basic methods in molecular biology.* Elsevier.
Dehrmann, F. M., Coetzer, T. H., Pike, R. N. & Dennison, C. (1995). Mature cathepsin L is substantially active in the ionic milieu of the extracellular medium. *Archives of Biochemistry and Biophysics,* **324,** 93-8.
Desquesnes, M., Bengaly, Z., Millogo, L., Meme, Y. & Sakande, H. (2001). The analysis of the cross-reactions occurring in antibody-ELISA for the detection of trypanosomes can improve identification of the parasite species involved. *Annals of tropical medicine and parasitolog,* **95,** 141-55.
Desquesnes, M., Holzmuller, P., Lai, D. H., Dargantes, A., Lun, Z. R. & Jittaplapong, S. (2013). *Trypanosoma evansi* and surra: a review and perspectives on origin, history, distribution, taxonomy, morphology, hosts, and pathogenic effects. *Biomed research international,* **2013,** 194176.
Donovan, R. S., Robinson, C. W. & Glick, B. R. (1996). Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter. *Journal of industrial microbiology,* **16,** 145-54.
Dóró, É., Jacobs, S. H., Hammond, F. R., Schipper, H., Pieters, R. P., Carrington, M., Wiegertjes, G. F. & Forlenza, M. (2019). Visualizing trypanosomes in a vertebrate host reveals novel swimming behaviours, adaptations and attachment mechanisms. *eLife,* **8,** e48388.
Dos Reis, F. C., Júdice, W. A., Juliano, M. A., Juliano, L., Scharfstein, J. & De A. Lima, A. P. C. (2006). The substrate specificity of cruzipain 2, a cysteine protease isoform from *Trypanosoma cruzi*. *FEMS microbiology letters,* **259,** 215-220.
Duschak, V. G. & Couto, A. S. (2009). Cruzipain, the major cysteine protease of *Trypanosoma cruzi*: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target. A review. *Current medicinal chemistry,* **16,** 3174-202.
Ellis, K. J. & Morrison, J. F. (1982). [23] Buffers of constant ionic strength for studying pH-dependent processes. *In:* Purich, D. L. (ed.) *Methods in Enzymology.* Academic Press.
Erez, E., Fass, D. & Bibi, E. (2009). How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature,* **459,** 371-8.
Eyford, B. A., Ahmad, R., Enyaru, J. C., Carr, S. A. & Pearson, T. W. (2013). Identification of trypanosome proteins in plasma from African sleeping sickness patients infected with *T. b. rhodesiense*. *PLOS One,* **8,** e71463.
Eyssen, L. E. & Coetzer, T. H. (2019). Expression, purification and characterisation of *Trypanosoma congolense* metacaspase 5 (*Tco*MCA5)-a potential drug target for animal African trypanosomiasis. *Protein expression and purification,* **164,** 105465.
Eyssen, L. E. A., Vather, P., Jackson, L., Ximba, P., Biteau, N., Baltz, T., Boulangé, A., Büscher, P. & Coetzer, T. H. T. (2018). Recombinant and native *Tvi*CATL from *Trypanosoma vivax:* enzymatic characterisation and evaluation as a diagnostic target for animal African trypanosomosis. *Molecular and biochemical parasitology,* **223,** 50-54.
Felsenstein, J. (2002). Phylogeny inference package (PHYLIP), Version 3.6 a3. *Seattle: department of genetics, university of Washington*.
Fermino, B. R., Paiva, F., Soares, P., Tavares, L. E. R., Viola, L. B., Ferreira, R. C., Botero-Arias, R., de-Paula, C. D., Campaner, M. & Takata, C. S. (2015). Field and experimental evidence of a new caiman trypanosome species closely phylogenetically related to fish trypanosomes and transmitted by leeches. *International journal for parasitology: parasites and wildlife,* **4,** 368-378.
Franco, J. R., Simarro, P. P., Diarra, A. & Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Clinical epidemiology,* **6,** 257.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook***,** 571-607.

Gibson, W. C., Lom, J., PeckovÁ, H., Ferris, V. R. & Hamilton, P. B. (2005). Phylogenetic analysis of freshwater fish trypanosomes from Europe using ssu rRNA gene sequences and random amplification of polymorphic DNA. *Parasitology,* **130,** 405-412.

Giordani, F., Morrison, L. J., Rowan, T. G., De Koning, H. P. & Barrett, M. P. (2016). The animal trypanosomiases and their chemotherapy: a review. *Parasitology,* **143,** 1862-1889.

Gooderham, K. (1984). Transfer techniques in protein blotting. *Proteins.* Springer.

Grybchuk-Ieremenko, A., Losev, A., Kostygov, A., Lukeš, J. & Yurchenko, V. (2014). High prevalence of trypanosome co-infections in freshwater fishes. *Folia parasitologica,* **61,** 495-504.

Gu, Z., Wang, J., Ke, X., Liu, Y., Liu, X., Gong, X. & Li, A. (2010). Phylogenetic position of the freshwater fish trypanosome, *Trypanosoma ophiocephali* (Kinetoplastida) inferred from the complete small subunit ribosomal RNA gene sequence. *Parasitology research,* **106,** 1039-1042.

Gu, Z., Wang, J., Li, M., Zhang, J., Ke, X. & Gong, X. (2007). Morphological and genetic differences of *Trypanosoma* in some Chinese freshwater fishes: difficulties of species identification. *Parasitology research,* **101,** 723-730.

Gupta, N. (2006). Historical review of piscine trypanosomiasis and survey of Indian *Trypanosoma*. *Journal of parasitic diseases,* **30,** 101-115.

Gupta, N. & Gupta, D. (2012). Erythropenia in piscine trypanosomiasis. *Trends in parasitology,* **1,** 1-6.

Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series, 1999. [London]: Information Retrieval Ltd., c1979-c2000., 95-98.

Hamill, L., Picozzi, K., Fyfe, J., von Wissmann, B., Wastling, S., Wardrop, N., Selby, R., Acup, C. A., Bardosh, K. L., Muhanguzi, D., Kabasa, J. D., Waiswa, C. & Welburn, S. C. (2017). Evaluating the impact of targeting livestock for the prevention of human and animal trypanosomiasis, at village level, in districts newly affected with *T. b. rhodesiense* in Uganda. *Infectious diseases of poverty,* **6,** 16-16.

Hamilton, P. B., Stevens, J. R., Gaunt, M. W., Gidley, J. & Gibson, W. C. (2004). Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *International journal for parasitology,* **34,** 1393-1404.

Hayes, P. M., Lawton, S. P., Smit, N. J., Gibson, W. C. & Davies, A. J. (2014). Morphological and molecular characterization of a marine fish trypanosome from South Africa, including its development in a leech vector. *Parasites & vectors,* **7,** 50.

Heussen, C. & Dowdle, E. B. (1980). Electrophoretic analysis of plasminogen activators in polyacrylamide gels containing sodium dodecyl sulfate and copolymerized substrates. *Analytical biochemistry,* **102,** 196-202.

Horn, D. (2014). Antigenic variation in African trypanosomes. *Molecular and biochemical parasitology,* **195,** 123-129.

Huang, Q. & Fu, W.-L. (2005). Comparative analysis of the DNA staining efficiencies of different fluorescent dyes in preparative agarose gel electrophoresis. *Clinical chemistry and laboratory medicine : CCLM / FESCC,* **43,** 841-2.

Islam, A. N. & Woo, P. T. (1991). *Trypanosoma danilewskyi* in *Carassius auratus*: the nature of protective immunity in recovered goldfish. *The journal of parasitology***,** 258-262.

Jackson, A. P., Goyard, S., Xia, D., Foth, B. J., Sanders, M., Wastling, J. M., Minoprio, P. & Berriman, M. (2015). Global gene expression profiling through the complete life cycle of *Trypanosoma vivax*. *PLOS neglected tropical diseases,* **9,** e0003975.

Jesus, R. B. d., Gallani, S. U., Valladão, G. M. R., Pala, G., Silva, T. F. A. d., Costa, J. C. d., Kotzent, S. & Pilarski, F. (2018). Trypanosomiasis causing mortality outbreak in Nile tilapia intensive farming: Identification and pathological evaluation. *Aquaculture,* **491,** 169-176.

Jiang, B., Lu, G., Du, J., Wang, J., Hu, Y., Su, Y. & Li, A. (2019). First report of trypanosomiasis in farmed largemouth bass (*Micropterus salmoides*) from China: pathological evaluation and taxonomic status. *Parasitology Research,* **118,** 1731-1739.

Kangethe, R. T., Boulangé, A. F., Coustou, V., Baltz, T. & Coetzer, T. H. (2012). *Trypanosoma brucei brucei* oligopeptidase B null mutants display increased prolyl oligopeptidase-like activity. *Molecular and biochemical parasitology,* **182,** 7-16.

Karlsbakk, E. & Nylund, A. (2006). Trypanosomes infecting cod *Gadus morhua L.* in the North Atlantic: a resurrection of *Trypanosoma pleuronectidium* Robertson, 1906 and delimitation of *T. murmanense* Nikitin, 1927 (emend.), with a review of other trypanosomes from North Atlantic and mediterranean teleosts. *Systematic parasitology,* **65,** 175-203.

Kaufer, A., Stark, D. & Ellis, J. (2020). A review of the systematics, species identification and diagnostics of the Trypanosomatidae using the maxicircle kinetoplast DNA: from past to present. *International journal for parasitology,* **50,** 449-460.
Kaur, J., Kumar, A. & Kaur, J. (2018). Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *International journal of biological macromolecules,* **106,** 803-822.
Kelly, E., Barbosa, A. D., Gibson-Kueh, S. & Lymbery, A. J. (2018). Haematozoa of wild catfishes in northern Australia. *International journal for parasitology: parasites and wildlife,* **7,** 12-17.
Kennedy, P. G. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *Lancet neurology,* **12,** 186-94.
Kerr, I. D., Wu, P., Marion-Tsukamaki, R., Mackey, Z. B. & Brinen, L. S. (2010). Crystal structures of *Tb*CatB and rhodesain, potential chemotherapeutic targets and major cysteine proteases of *Trypanosoma brucei*. *PLoS neglected tropical diseases,* **4,** e701-e701.
Kimple, M. E., Brill, A. L. & Pasker, R. L. (2013). Overview of affinity tags for protein purification. *Current protocols in protein science,* **73,** 9.9. 1-9.9. 23.
Klemba, M. & Goldberg, D. E. (2002). Biological roles of proteases in parasitic protozoa. *Annual review of biochemistry,* **71,** 275-305.
Koeller, C. M. & Bangs, J. D. (2019). Processing and targeting of cathepsin L (*TbCatL*) to the lysosome in *Trypanosoma brucei*. *Cellular microbiology,* **21,** e12980.
Kolev, N. G., Ramey-Butler, K., Cross, G. A., Ullu, E. & Tschudi, C. (2012). Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science,* **338,** 1352-3.
Lacomble, S., Vaughan, S., Gadelha, C., Morphew, M. K., Shaw, M. K., McIntosh, J. R. & Gull, K. (2010). Basal body movements orchestrate membrane organelle division and cell morphogenesis in *Trypanosoma brucei*. *Journal of cell science,* **123,** 2884-91.
Lacount, D., Gruszynski, A., Grandgenett, P., Bangs, J. & Donelson, J. (2003). Expression and function of the *Trypanosoma brucei* major surface protease (GP63) genes. *Journal of biological chemistry,* **278,** 24658-64.
Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature,* **227,** 680-685.
Lalmanach, G., Boulangé, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F. & Authié, E. (2002). Congopain from *Trypanosoma congolense:* drug target and vaccine candidate. *Biological chemistry,* **383,** 739-49.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics,* **23,** 2947-2948.
Latif, A. A., Ntantiso, L. & De Beer, C. (2019). African animal trypanosomosis (nagana) in northern KwaZulu-Natal, South Africa: strategic treatment of cattle on a farm in endemic area. *The onderstepoort journal of veterinary research,* **86,** e1-e6.
Lebendiker, M. & Danieli, T. (2014). Production of prone-to-aggregate proteins. *FEBS Letters,* **588,** 236-246.
Lemos, M., Fermino, B. R., Simas-Rodrigues, C., Hoffmann, L., Silva, R., Camargo, E. P., Teixeira, M. M. & Souto-Padrón, T. (2015). Phylogenetic and morphological characterization of trypanosomes from Brazilian armoured catfishes and leeches reveal high species diversity, mixed infections and a new fish trypanosome species. *Parasites & vectors,* **8,** 573.
Lischke, A., Klein, C., Stierhof, Y. D., Hempel, M., Mehlert, A., Almeida, I. C., Ferguson, M. A. & Overath, P. (2000). Isolation and characterization of glycosylphosphatidylinositol-anchored, mucin-like surface glycoproteins from bloodstream forms of the freshwater-fish parasite *Trypanosoma carassii*. *The biochemical journal,* **345 Pt 3,** 693-700.
MacLean, L., Reiber, H., Kennedy, P. G. E. & Sternberg, J. M. (2012). Stage progression and neurological symptoms in *Trypanosoma brucei rhodesiense* sleeping sickness: role of the CNS inflammatory response. *PLoS neglected tropical diseases,* **6,** e1857-e1857.
Maslov, D. A., Lukeš, J., Jirku, M. & Simpson, L. (1996). Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. *Molecular and biochemical parasitology,* **75,** 197-205.
Matthews, K. R. (2005). The developmental cell biology of *Trypanosoma brucei*. *Journal of cell science,* **118,** 283.
Matthews, K. R., McCulloch, R. & Morrison, L. J. (2015). The within-host dynamics of African trypanosome infections. *Philosophical transactions of the royal society of London. series b, biological sciences,* **370,** 20140288.

Mbawa, Z. R., Gumm, I. D., Fish, W. R. & Lonsdale-Eccles, J. D. (1991). Endopeptidase variations among different life-cycle stages of African trypanosomes. *European journal of biochemistry,* **195,** 183-190.

Mbawa, Z. R., Gumm, I. D., Shaw, E. & Lonsdale-Eccles, J. D. (1992). Characterisation of a cysteine protease from bloodstream forms of *Trypanosoma congolense*. *European journal of biochemistry,* **204,** 371-9.

McKerrow, J. H., C. Caffrey, B. Kelly, P. Loke & M. Sajid (2006). Proteases in parasitic diseases. *Annual review of pathology,* **1,** 497-536.

McLuskey, K., Rudolf, J., Proto, W. R., Isaacs, N. W., Coombs, G. H., Moss, C. X. & Mottram, J. C. (2012). Crystal structure of a *Trypanosoma brucei* metacaspase. *Proceedings of the national academy of sciences of the United States of America,* **109,** 7469-7474.

Melville, S. E., Majiwa, P. A. & Tait, A. (2004). The African Trypanosome Genome. In The Trypanosomiases, (ed. I. Maudlin P. H. Holmes and M. A. Miles), pp. 39-57. Wallingford: CAB International.

Mendoza-Palomares, C., Biteau, N., Giroud, C., Coustou, V., Coetzer, T., Authié, E., Boulangé, A. & Baltz, T. (2008). Molecular and biochemical characterization of a cathepsin B-like protease family unique to *Trypanosoma congolense*. *Eukaryotic cell,* **7,** 684-697.

Meyer, A., Holt, H. R., Selby, R. & Guitian, J. (2016). Past and ongoing tsetse and animal trypanosomiasis control operations in five African countries: a systematic review. *PLOS neglected tropical diseases,* **10,** e0005247.

Minina, E. A., Coll, N. S., Tuominen, H. & Bozhkov, P. V. (2017). Metacaspases versus caspases in development and cell fate regulation. *Cell death and differentiation,* **24,** 1314-1325.

Mitashi, P., Hasker, E., Lejon, V., Kande, V., Muyembe, J.-J., Lutumba, P. & Boelaert, M. (2012). Human African trypanosomiasis diagnosis in first-line health services of endemic countries, a systematic review. *PLoS neglected tropical diseases,* **6**.

Moreira, D., López-García, P. & Vickerman, K. (2004). An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *International journal of systematic and evolutionary microbiology,* **54,** 1861-1875.

Moreno, C. J., Torres, T. & Silva, M. S. (2019). Variable surface glycoprotein from *Trypanosoma brucei* undergoes cleavage by matrix metalloproteinases: an in silico approach. *Pathogens,* **8**.

Morty, R. E., Authié, E., Troeberg, L., Lonsdale-Eccles, J. D. & Coetzer, T. H. T. (1999). Purification and characterisation of a trypsin-like serine oligopeptidase from *Trypanosoma congolense*. *Molecular and biochemical parasitology,* **102,** 145-155.

Mottram, J. C., Helms, M. J., Coombs, G. H. & Sajid, M. (2003). Clan CD cysteine peptidases of parasitic protozoa. *Trends in parasitology,* **19,** 182-187.

Muhanguzi, D., Mugenyi, A., Bigirwa, G., Kamusiime, M., Kitibwa, A., Akurut, G. G., Ochwo, S., Amanyire, W., Okech, S. G., Hattendorf, J. & Tweyongyere, R. (2017). African animal trypanosomiasis as a constraint to livestock health and production in Karamoja region: a detailed qualitative and quantitative assessment. *BMC veterinary research,* **13,** 355.

Mullis, K. B. & Faloona, F. A. (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology,* **155,** 335-350.

Neville Jr, D. M. (1971). Molecular weight determination of protein-dodecyl sulfate complexes by gel electrophoresis in a discontinuous buffer system. *Journal of biological chemistry,* **246,** 6328-6334.

Nok, A. J. (2009). Chapter 62 - African trypanosomiasis. *In:* Barrett, A. D. T. & Stanberry, L. R. (eds.) *Vaccines for biodefense and emerging and neglected diseases.* London: Academic Press.

Nussbaum, K., Honek, J., Cadmus, C. M. & Efferth, T. (2010). Trypanosomatid parasites causing neglected diseases. *Current medicinal chemistry,* **17,** 1594-617.

O'Brien, T. C., Mackey, Z. B., Fetter, R. D., Choe, Y., O'Donoghue, A. J., Zhou, M., Craik, C. S., Caffrey, C. R. & McKerrow, J. H. (2008). A parasite cysteine protease is key to host protein degradation and iron acquisition. *Journal of biological chemistry,* **283,** 28934-28943.

Oladiran, A. & Belosevic, M. (2012). Immune evasion strategies if trypanosomes: A review. *The journal of parasitology,* **98,** 284-292.

Ooi, C.-P. & Bastin, P. (2013). More than meets the eye: understanding *Trypanosoma brucei* morphology in the tsetse. *Frontiers in cellular and infection microbiology,* **3**.

Ornstein, L. (1964). Disc electrophoresis. I. Background and theory. *Ann. NY Acad. Sci,* **121,** 321-349.

Osório, A. L., Madruga, C. R., Desquesnes, M., Soares, C. O., Ribeiro, L. R. & Costa, S. C. (2008). *Trypanosoma* (*Duttonella*) *vivax*: its biology, epidemiology, pathogenesis, and introduction in the New World--a review. *Memórias do instituto Oswaldo Cruz,* **103,** 1-13.

Overath, P., Haag, J., Lischke, A. & O'HUigin, C. (2001). The surface structure of trypanosomes in relation to their molecular phylogeny. *International journal for parasitology,* **31,** 468-471.

Overath, P., Haag, J., Mameza, M. & Lischke, A. (1999). Freshwater fish trypanosomes: definition of two types, host control by antibodies and lack of antigenic variation. *Parasitology,* **119,** 591-601.
Overath, P., Ruoff, J., Stierhof, Y.-D., Haag, J., Tichy, H., Dyková, I. & Lom, J. (1998). Cultivation of bloodstream forms of *Trypanosoma carassii*, a common parasite of freshwater fish. *Parasitology research,* **84,** 343-347.
Pamer, E., Davis, C. & So, M. (1991). Expression and deletion analysis of the *Trypanosoma brucei rhodesiense* cysteine protease in *Escherichia coli*. *Infection and immunity,* **59,** 1074-1078.
Peacock, L., Cook, S., Ferris, V., Bailey, M. & Gibson, W. (2012). The life cycle of *Trypanosoma* (*Nannomonas*) *congolense* in the tsetse fly. *Parasites & vectors,* **5,** 109.
Penchenier, L., Grébaut, P., Njokou, F., Eboo Eyenga, V. & Büscher, P. (2003). Evaluation of LATEX/*T.b.gambiense* for mass screening of *Trypanosoma brucei gambiense* sleeping sickness in Central Africa. *Acta tropica,* **85,** 31-7.
Petty, K. J. (2001). Metal-chelate affinity chromatography. *Current protocols in molecular biology,* **Chapter 10,** Unit 10.11B.
Phan, T. C., Nowak, K. J., Akkari, P. A., Zheng, M. H. & Xu, J. (2003). Expression of caltrin in the baculovirus system and its purification in high yield and purity by cobalt (II) affinity chromatography. *Protein expression and purification,* **29,** 284-290.
Pillay, D., Izotte, J., Fikru, R., Büscher, P., Mucache, H., Neves, L., Boulangé, A., Seck, M. T., Bouyer, J., Napier, G. B., Chevtzoff, C., Coustou, V. & Baltz, T. (2013). *Trypanosoma vivax* GM6 antigen: a candidate antigen for diagnosis of African animal trypanosomosis in cattle. *PLoS One,* **8,** e78565.
Ponte-Sucre, A. (2016). An overview of *Trypanosoma brucei* infections: an intense host–parasite interaction. *Frontiers in microbiology,* **7**.
Radwanska, M. (2010). Emerging trends in the diagnosis of human African trypanosomiasis. *Parasitology,* **137,** 1977-1986.
Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology,* **67,** 901.
Rawlings, N. & Barrett, A. J. (2013). Introduction: The clans and families of cysteine peptidases. *Handbook of proteolytic enzymes,* **2,** 1743-1773.
Rawlings, N. D. (2020). Twenty-five years of nomenclature and classification of proteolytic enzymes. *Biochim biophys acta proteins proteom,* **1868,** 140345.
Read, S. M. & Northcote, D. (1981). Minimization of variation in the response to different proteins of the Coomassie blue G dye-binding assay for protein. *Analytical biochemistry,* **116,** 53-64.
Reynolds, J. A. & Tanford, C. (1970). Binding of dodecyl sulfate to proteins at high binding ratios. Possible implications for the state of proteins in biological membranes. *Proceedings of the National Academy of Sciences,* **66,** 1002-1007.
Roberta, E., Santo, P., Lucia, T., Gregorio, C., Silvana, G. & Maria, Z. (2016). The inhibition of cysteine proteases rhodesain and *TbCatB*: a valuable approach to treat human African trypanosomiasis. *Mini-reviews in medicinal chemistry,* **16,** 1374-1391.
Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S. K., Dryden, D. T. & Dybvig, K. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic acids research,* **31,** 1805-1812.
Rodrigues, R. N., Oliveira, M. S. B., Tavares-Dias, M. & Corrêa, L. L. (2018). First record of infection by *Trypanosoma* sp. of *Colossoma macropomum* (Serrasalmidae), a neotropical fish cultivated in the Brazilian Amazon. *Journal of applied aquaculture,* **30,** 29-38.
Rojas, F., Silvester, E., Young, J., Milne, R., Tettey, M., Houston, D. R., Walkinshaw, M. D., Pérez-Pi, I., Auer, M., Denton, H., Smith, T. K., Thompson, J. & Matthews, K. R. (2019). Oligopeptide signaling through *Tb*GPR89 drives trypanosome quorum sensing. *Cell,* **176,** 306-317.e16.
Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* **19,** 1572-1574.
Rotureau, B. & Van Den Abbeele, J. (2013). Through the dark continent: African trypanosome development in the tsetse fly. *Frontiers in cellular and infection microbiology,* **3,** 53.
Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology,* **132,** 365-86.
Ruszczyk, A., Forlenza, M., Savelkoul, H. F. & Wiegertjes, G. F. (2008). Molecular cloning and functional charactrisation of a cathepsin L-like proteinase from the fish kinetoplastid parasite *Trypanosoma carassii*. *Fish & shellfish immunology,* **24,** 205-214.

Salvesen,G. S and Nagase, H. (2001) Inhibition of Proteolytic Enzymes In Proteolytic Enzymes A Practical Approach (1st Edition) Rob Beynon and J. S. Bond (Eds), Oxford University Press,Oxford pp.83-104; 319-330

Sanderson, S. J., Pollock, K. G., Hilley, J. D., Meldal, M., St Hilaire, P., JULIANO, M. A., Juliano, L., Mottram, J. C. & Coombs, G. H. (2000). Expression and characterization of a recombinant cysteine proteinase of *Leishmania mexicana*. *Biochemical Journal,* **347,** 383-388.

Schägger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical biochemistry,* **166,** 368-79.

Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and biophysical research communications,* **27,** 157-162.

Schlager, B., Straessle, A. & Hafen, E. (2012). Use of anionic denaturing detergents to purify insoluble proteins after overexpression. *BMC biotechnology,* **12,** 95-95.

Sehgal, R. N., Jones, H. I. & Smith, T. B. (2001). Host specificity and incidence of *Trypanosoma* in some African rainforest birds: a molecular approach. *Molecular ecology,* **10,** 2319-27.

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial cell factories,* **14,** 41.

Singh, M., Yadav, A., Ma, X. & Amoah, E. (2010). Plasmid DNA transformation in *Escherichia coli*: effect of heat shock temperature, duration, and cold incubation of CaCl2 treated cells. *International journal of biotechnology and biochemistry,* **6,** 561-568.

Siqueira-Neto, J. L., Debnath, A., McCall, L.-I., Bernatchez, J. A., Ndao, M., Reed, S. L. & Rosenthal, P. J. (2018). Cysteine proteases in protozoan parasites. *PLoS neglected tropical diseases,* **12,** e0006512.

Smith, D. H., Pepin, J. & Stich, A. H. R. (1998). Human African trypanosomiasis: an emerging public health crisis. *British medical bulletin,* **54,** 341-355.

Smith, P. e., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. (1985). Measurement of protein using bicinchoninic acid. *Analytical biochemistry,* **150,** 76-85.

Steverding, D. (2008). The history of African trypanosomiasis. *Parasites & vectors,* **1,** 3.

Steverding, D. (2013). Proteases of *Trypanosoma brucei*. *Trypanosomatid diseases(eds P.M. Selzer, T. Jäger, O. Koch and L. Flohé). Wiley-VCH Verlag GmbH & Co. KGaA, Wenheim, Germany* 365-382.

Steverding, D. (2015). Evaluation of trypanocidal activity of combinations of anti-sleeping sickness drugs with cysteine protease inhibitors. *Experimental parasitology,* **151-152,** 28-33.

Steverding, D. (2017). Sleeping sickness and nagana disease caused by *Trypanosoma brucei*.

Stijlemans, B., De Baetselier, P., Magez, S., Van Ginderachter, J. A. & De Trez, C. (2018). African trypanosomiasis-associated anemia: the contribution of the interplay between parasites and the mononuclear phagocyte system. *Frontiers in immunology,* **9,** 218-218.

Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology,* **189,** 113-30.

Sutherland, C. S., Stone, C. M., Steinmann, P., Tanner, M. & Tediosi, F. (2017). Seeing beyond 2020: an economic evaluation of contemporary and emerging strategies for elimination of *Trypanosoma brucei gambiense*. *Lancet glob health,* **5,** e69-e79.

Szallies, A., Kubata, B. K. & Duszenko, M. (2002). A metacaspase of *Trypanosoma brucei* causes loss of respiration competence and clonal death in the yeast *Saccharomyces cerevisiae*. *FEBS letters,* **517,** 144-150.

Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the national academy of sciences of the United States of America,* **76,** 4350-4354.

Troeberg, L., Pike, R. N., Morty, R. E., Berry, R. K., Coetzer, T. H. & Lonsdale-Eccles, J. D. (1996). Proteases from *Trypanosoma brucei brucei*: purification, characterisation and interactions with host regulatory molecules. *European journal of biochemistry,* **238,** 728-736.

Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zavialov, A., Lam, E. & Bozhkov, P. V. (2011). Metacaspases. *Cell death and differentiation,* **18,** 1279.

Van Den Abbeele, J., Claes, Y., van Bockstaele, D., Le Ray, D. & Coosemans, M. (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology,* **118 (Pt 5),** 469-78.

Verma, S., Dixit, R. & Pandey, K. C. (2016). Cysteine proteases: Modes of mctivation and future prospects as pharmacological targets. *Frontiers in pharmacology,* **7,** 107-107.

Welburn, S. C., Molyneux, D. H. & Maudlin, I. (2016). Beyond tsetse – implications for research and control of human African trypanosomiasis epidemics. *Trends in parasitology,* **32,** 230-241.
Wiegertjes, G. F., Forlenza, M., Joerink, M. & Scharsack, J. P. (2005). Parasite infections revisited. *Developmental & comparative immunology,* **29,** 749-758.
Wright, A. D., Li, S., Feng, S., Martin, D. S. & Lynn, D. H. (1999). Phylogenetic position of the kinetoplastids, *Cryptobia bullocki, Cryptobia catostomi,* and *Cryptobia salmositica* and monophyly of the genus *Trypanosoma* inferred from small subunit ribosomal RNA sequences. *Molecular and biochemical parasitology,* **99,** 69-76.
Yamaguchi, H. & Miyazaki, M. (2014). Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules,* **4,** 235-251.
Yamamoto, Y., Kurata, M., Watabe, S., Murakami, R. & Takahashi, S. Y. (2002). Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. *Current protein & peptide science,* **3,** 231-8.
Zwickl, D. (2006). Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.

APPENDIX

Appendix 1: Multiple sequence alignment of *Tb*CATL and cysteine protease mRNA from *T. brucei*

```
apTeasyTb      ------------------ACACGGTMRGTGATTGTATACGACTCACTATAGGGCGAA---  39
X16465.1       GAAGGTCCATCCAAACATAAACAGCCGAAAGATGCC-TCGAACAGAAATGGTGCGTTTTG   59
                           * ** *        **    ***      ** * ***

pTeasyTb       TTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTTTGGAT   99
X16465.1       TACGTCTCCCCGTTGTCTTGCTGGCTATGGCAGCGTGCCTTGCGTCTGTCGCACTCGGGT   119
               *  * * *  *** *  *          * **  * * *    *    ***   * ** *

pTeasyTb       CCCTCCACGTGGAGGAGTCATTGGAGATGCGTTTTGCTGCGTTCAAGAAGAAGTACGGCA   159
X16465.1       CGCTCCACGTGGAGGAGTCATTGGAGATGCGTTTTGCTGCGTTCAAGAAGAAGTACGGCA   179
               * ************************************************************

pTeasyTb       AGGTGTACAAGGATGCTAAGGAGGAAGCATTCCGCTTCCGTGCCTTTGAGGAGAATATGG   219
X16465.1       AGGTGTACAAGGATGCTAAGGAGGAAGCATTCCGCTTCCGTGCCTTTGAGGAAAATATGG   239
               ************************************************** *******

pTeasyTb       AGCAGGCGAAGATTCAAGCTGCGGCGAACCCATACGCAACGTTTGGTGTGACACCCTTCT   279
X16465.1       AGCAGGCGAAGATTCAAGCTGCGGCGAACCCATACGCAACGTTTGGTGTGACACCCTTCT   299
               ************************************************************

pTeasyTb       CGGATATGACACGTGAAGAGTTCAGGGCACGCTACCGTAACGGCGCGTCCTACTTTGCAG   339
X16465.1       CGGATATGACACGTGAAGAGTTCAGGGCACGCTACCGTAACGGCGCGTCCTACTTTGCAG   359
               ************************************************************

pTeasyTb       CTGCGCAGAAGCGGCTACGCAAGACGGTGAACGTAACCACTGGCCGTGCTCCTGCAGCTG   399
X16465.1       CTGCGCAGAAGCGGCTACGCAAGACGGTGAACGTAACCACTGGCCGTGCTCCTGCAGCTG   419
               ************************************************************

pTeasyTb       TGGATTGGCGTGAGAAGGGAGCAGTGACCCCAGTGAAGGATCAGGGTCAGTGCGGCTCGT   459
X16465.1       TGGATTGGCGTGAGAAGGGAGCAGTGACCCCAGTGAAGGTTCAGGGTCAGTGCGGCTCGT   479
               ************************************ ********************

pTeasyTb       GCTGGGCCTTTTCAACTATCGGCAACATCGAAGGGCAGTGGCAGGTGGCAGGAAATCCTC   519
X16465.1       GCTGGGCCTTTTCAACTATCGGCAACATCGAAGGGCAGTGGCAGGTGGCAGGAAATCCTC   539
               ************************************************************

pTeasyTb       TCGTATCCCTCTCGGAGCAGATGCTAGTGTCATGTGATACCATTGATTTCGGCTGTGGTG   579
X16465.1       TCGTATCCCTCTCGGAGCAGATGCTAGTGTCATGTGATACCATTGATTCAGGTTGTAATG   599
               ************************************************  ** ***  **

pTeasyTb       GTGGGCTGATGGACAATGCCTTCAACTGGATAGTAAATTCAAACGGTGGAAACGTATTCA   639
X16465.1       GTGGGCTGATGGACAATGCCTTCAACTGGATAGTAAATTCAAACGGTGGAAACGTATTCA   659
               ************************************************************

pTeasyTb       CGGAGGCGAGCTATCCCTATGTTTCTGGGAATGGTGAGCAGCCACAGTGCCAGATGAATG   699
X16465.1       CGGAGGCGAGCTATCCCTATGTTTCTGGGAATGGTGAGCAGCCACAGTGCCAGATGAATG   719
               ************************************************************

pTeasyTb       GTCACGAGATCGGTGCTGCGATAACAGACCATGTTGACTTACCGCAGGATGAGGACGCTA   759
X16465.1       GTCACGAGATCGGTGCTGCGATAACAGACCATGTTGACTTACCGCAGGATGAGGACGCTA   779
               ************************************************************

pTeasyTb       TCGCCGCGTATTTGGCAGAA-ACGGTCCCCTTGCTATTGCCGTTGACGCCACAAGTTTTA   818
X16465.1       TCGCCGCGTATTTGGCAGAAAACGGTCCCCTTGCTATCGCTGTTGACGCCGAAAGTTTTA   839
               ******************* *************** ** *********  ********

pTeasyTb       TGGACTATAACGGTGGGATTCTGACTTCATGCACCTCCGAGCAATTGGATCATGGTGTGC   878
X16465.1       TGGACTATAACGGTGGGATTCTGACTTCATGCACCTCCAAACAACTGGATCATGGTGTGC   899
```

```
                ************************************ * *** ***************

pTeasyTb        TCCTCGTTGGTTACAATGATAATAGCAATCCACCCTACTGGATCATCAAAAACTCGTGGA  938
X16465.1        TCCTCGTTGGTTACAATGATAATAGCAATCCACCCTACTGGATCATCAAAAACTCGTGGA  959
                ************************************************************

pTeasyTb        GCAACATGTGGGGCGAGGACGGCTACATCCGCATCGAGAAGGGCACAAACCAATGTCTCA  998
X16465.1        GCAACATGTGGGGCGAGGACGGCTACATCCGCATCGAGAAGGGCACAAACCAATGTCTCA  1019
                ************************************************************

pTeasyTb        TGAATCAGCCGTATCCTCCGCATGACTCGAGAAAA-TCACTAGTGAATTCGCGGCGCCTG  1057
X16465.1        TGAATCAGGCCGTATCCTCCGCAGTTGTTGGAGGCCCCACTCCACCAC-CACCACCACCG  1078
                ******** *     *  *    *      **     ****     *  * *  *  * *

pTeasyTb        CAGTCGACCATATGGGAGAGCTCCCACGCGTGATGCATAGCTGAGTATT---CTATAGTG  1114
X16465.1        CCGCCGCCTTCAGCAACTTTTACACAGGACTTCTGCGAGGGCAAGGGTTGTACCAAAGGC  1138
                * * ** *   *          * ** *  *  ***   *   **  **   * * **

pTeasyTb        TCACTAAATAGCTGCGTATCATGGTCATAGCTGTT------------CTGKGKGAAATGT  1162
X16465.1        TGCTCACATGCCACC-TTCCCCACTGGCGAGTGCGTCCAGACTACCGGCGTCGGCTCAGT  1197
                *    * **  *  * *  *    *      **                 *   *    **

pTeasyTb        TATCCSCTCACATCCMACA-----------ACATACGAG--------------CGGAGCA  1197
X16465.1        GATCGCCACATGTGGCGCAAGCAACCTTACACAAATAATCTACCCACTAAGCAGGAGCTG  1257
                 ***  * **  *    **           *** *  *               *

pTeasyTb        TAACGGTAGCCTGGGTGCTATGAGTGARCTACTCACATAATGSATGTGCCTYMATGTTCC  1257
X16465.1        CAGCGGTCCCTCTGTGCCGATTACTGTGCCACTGGATAAGTGCATACCCATTTTGATTGG  1317
                 * **** *   *   * ** * **  * ***     * ** **   * *     **

pTeasyTb        CGCTTTCW---CAGATTYCG-----------------------------------------  1274
X16465.1        GTCCGTTGAGTATCATTGCTCCACCAACCCACCCACTAAGGCGGCCAGGCTGGTCCCACA  1377
                  *  *        *** *

pTeasyTb        ------------------------------------------------------------  1274
X16465.1        CCAGTGAGGTGGCGTGGTCTTGGGTTGCGCTTCACCTTGTGATTGGTGTTTCTCGTTCAT  1437

pTeasyTb        ------------------------------------------------------------  1274
X16465.1        TATTGCTTGGTTTTTGTTTTTATTTCCTTCCCTTTTTACTCCCCCTCAATTCTATCTTCT  1497

pTeasyTb        ------------------------------------------------------------  1274
X16465.1        TCGGCGAGCGGCTGTGTAGTGGAACTGTAATACACCTGGAGGGCATGTGCGTGTGGTATG  1557

pTeasyTb        ------------------------------------------------------------  1274
X16465.1        TGCATAGCAGAATGTGCTGCGCTTCGGTACAGTGAGTGCTGATGCCGTCTTAGCCACGCG  1617

pTeasyTb        ----       1274
X16465.1        TGCC       1621
```