Investigation into the Major Surface Proteases of African Trypanosomes

by

Alexandré Marie Chaplin Delport

BSc. (*Hons*) Biochemistry (*cum laude*)

Submitted in fulfilment of the academic requirements for the degree of Master of Science in Biochemistry,

School of Life Sciences,
University of KwaZulu-Natal,
Pietermaritzburg.

2016

Preface

The experimental work described in this dissertation was carried out in the School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, from January 2014 to October 2016, under the supervision of Professor THT Coetzer. 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.

____/____/2017 _____

Miss AMC Delport

As the candidate's supervisor I agree to the submission of this dissertation.

____/____/2017 _____

Prof. THT Coetzer

Declaration - Plagiarism

I, Alexandré Marie Chaplin Delport, 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 references sections.

____/____/2017 _____

Miss AMC Delport

Abstract

The unicellular parasite of the genus *Trypanosoma* infects a number of mammalian species including livestock and humans. In sub-Saharan Africa three main parasitic species cause disease: *Trypanosoma brucei, T. congolense and T. vivax.* The lack of sensitive diagnosis and increased drug resistance leaves an avenue in trypanosome research for exploring novel virulence factors as diagnostic and chemotherapeutic agents. The work reported in this dissertation involved investigation of the identified virulence factor, the Major Surface Protease (MSP), of African trypanosomes. The MSPs comprise a group of metalloproteases which have been found in over 13 Leishmanian species, *T. cruzi, T. brucei* and *T. congolense* as well as other *Kinetoplastids*.

In this study, putative M8 metalloprotease sequences were also identified in the *T. vivax* genome. These putative sequences were grouped into four classes of protein, *Tv*MSP-A, -C, -D and -E by phylogenetic comparison with other MSPs. Three-dimensional modelling showed high structural identity with leishmanolysin from *Leishmania major*. The *T. vivax* MSP sequences were used, in conjunction with *T. brucei and T. congolense* sequences, to select immunogenic peptide regions to produce anti-peptide antibodies. Three peptides were selected with the intention to 1) detect both *Tb*MSP-B and *Tco*MSP-B (peptide *Tb/Tco*MSP:303-314, cross-species), 2) detect only *Tb*MSP-C (peptide *Tb*MSP:400-412) and 3) detect only *Tv*MSP-C (peptide *Tv*MSP:686-697). They were used to generate two types of detection molecules: complete IgY anti-peptide antibodies and single chained fragments (scFvs), with the capability to detect the peptide in an ELISA format. Single scFv expressing *E. coli* colonies were successfully selected and shown to detect two (*Tb/Tco*MSP:303-314 and *Tb*MSP:400-412) of the three peptides. The anti-peptide antibodies, produced in chickens, were used to successfully detect native MSP within *T. brucei* and *T. congolense* parasite lysates; however, cross-reactivity between species was seen.

The *T. brucei* MSP-C class of protease was successfully cloned, expressed and purified, although, numerous truncated proteins and gene mutations occurred [truncated (t)*Tb*MSP-C]. The expression constructs r*Tb*MSP-C and r*Tco*MSP-C were hence synthesised. These two enzymes were successfully expressed and purified and were shown to form high molecular weight multimers. Furthermore, the enzymes were able to cleave the peptide substrate H-Suc-Leu-Tyr-AMC with acidic pH optima and activity was inhibited with metalloprotease inhibitors, EDTA and 1,10 phenanthroline.

The successful detection of r*Tco*MSP-C by *T. congolense* infected cattle sera was also observed. t*Tb*MSP-C, r*Tb*MSP-C and r*Tco*MSP-C were detected with the chicken anti-peptide antibodies and it was again found that these antibodies cross-reacted with different species MSPs. The high identity shared between MSPs from all *Trypanosoma* species made selecting species-specific antibodies difficult.

Further work to detect native MSP-C protease within infected sera or blood would give a definitive conclusion of its use in diagnostics. Moreover, antibodies that detect just *T. brucei* and *T. congolense* still need to be produced. Preliminary activity assays were performed on r*Tb*MSP-C and r*Tco*MSP-C but, additional research on the kinetics of these proteases is still needed.

In summary, it was shown that MSPs have the potential to be novel diagnostic markers especially for cross-trypanosomal species detection. Furthermore, activity of r*Tb*MSP-C and r*Tco*MSP-C has substrate cleavage specificity and pH optima that are comparable with leishmanolysin.

Acknowledgements

I would like to extend my gratitude and thanks to the following people and institutes:

Firstly, a very special thank you to my supervisor, Prof. Coetzer for all her help, assistance, support and guidance in completing my MSc (especially reading and correcting this dissertation). We finally got there!!

I extend this thank you to the other academics within our department, especially Prof Goldring, Prof. Niesler and Dr Hewer.

To my colleagues in Lab 44 for all the support and laughs over the three years. A special thanks to Lauren Eyssen whom without, the laboratory would not run as it does. To the other laboratories within out department of biochemistry, thank you for your constant friendliness, help and advice.

I would also like to thank everyone who forms the driving force of our department and university: the technicians, financial division and cleaning staff. All of you make it pleasant and easy to work at UKZN.

I would like to thank the National Research Foundation (NRF) for funding my three years of study, without which this degree would have not been possible.

I would like to thank the University of KwaZulu-Natal for their financial support in my first year of masters.

I dedicate this dissertation to my dad who taught me to always be curious and that, without mistakes you can never learn. To my mom who taught me to never give up and always believe in myself.

Above all I dedicate this to my beautiful sister Jodi, who reminds me every day that things are never as bad as they seem, to always follow my heart and in her strength she gives me strength. You are forever my inspiration.

Contents

Preface ... **ii**

Declaration - Plagiarism ... **iii**

Abstract ... **iv**

Acknowledgements ... **vi**

Contents ... **vii**

List of Figures ... **xi**

List of Tables ... **xiv**

Abbreviations ... **xv**

Chapter 1 Literature Review: The Trypanosome Parasites and their Major Surface Proteases ... **1**

1.1 History of Trypanosomiasis ... 1

1.2 Classification of Trypanosomes ... 2

1.3 Biology of the Trypanosome ... 5

1.3.1 The Genome of African Trypanosomes ... 5

1.3.2 Morphology of Trypanosomes ... 5

1.3.3 Antigenic Variation ... 7

1.4 Life Cycle of the African Trypanosome ... 7

1.5 African Trypanosomiasis ... 9

1.5.1 Pathogenesis ... 10

1.5.2 Diagnosis ... 11

1.5.3 Treatment and Control of Trypanosomiasis ... 13

1.5.4 Trypanotolerance ... 16

1.6 Leishmania, a *Trypanosomatidae* Parasite ... 17

1.7 Major Surface Proteases (MSP) in African Trypanosomes ... 18

1.7.1 Genomic Organisation of MSP Genes ... 18

1.7.2 Classification of Proteases ... 21

1.7.3 Structure of MSPs ... 24

1.7.4 Mechanism of Substrate Cleavage ... 31

1.7.5 Function of Major Surface Proteases 32
1.8 Objectives of the Study 35
Chapter 2: Major Surface Proteases in *Trypanosoma vivax* 37
2.1 Introduction 37
2.2 Materials and Methods 39
2.2.1 Selection of *T. vivax* MSP Sequences 39
2.2.2 Primary Structural Analyses 40
2.2.3 Secondary Structural Analyses 40
2.2.4 Tertiary Structural Analyses 40
2.3 Results 42
2.3.1 Selection of *T. vivax* Protein Sequences and Genomic Organisation 42
2.3.2 Primary Structure Analyses 43
2.3.3 Secondary Structure Analyses 44
2.3.4 Tertiary Structural Analyses 45
2.4 Discussion 52
Chapter 3: Anti-peptide Antibody Production against *T. brucei, T. congolense* and *T. vivax* MSP-B and -C 55
3.1 Introduction 55
3.2 Materials and Methods 60
3.2.1 Materials 60
3.2.2 Peptide Design 61
3.2.3 Coupling Peptide to Carrier Protein and Chicken Immunisation 62
3.2.4 Purification of Anti-Peptide Antibodies 63
3.2.5 Enzyme-Linked Immunosorbent Assay (ELISA) 64
3.2.6 Phage Display and Anti-Peptide scFvs 65
3.2.7 Parasite Purification and the Detection of MSP in Parasite lysate 69
3.3 Results 70
3.3.1 Peptide Design 70
3.3.2 Chicken Anti-Peptide Antibody Production 74

3.3.3 scFv Anti-Peptide Antibody Production 81
3.3.4 Detection of MSP in Parasite Lysate 84
3.4 Discussion 85
Chapter 4: Cloning of *T. brucei* MSP-C and Expression and Characterisation of *T. brucei and T. congolense* MSP-C 88
4.1 Introduction 88
4.2 Materials and Methods 90
4.2.1 Materials 90
4.2.2 Parasite Purification 91
4.2.3 Genomic DNA Isolation 91
4.2.4 Primer Design and Polymerase Chain Reaction (PCR) of *TbMSP-C* 91
4.2.5 Cloning of *TbMSP-C* into pGEM®-T and pGEX 4T-1 Vectors 93
4.2.6 Transformation and Restriction Digestion of Recombinant (r)*Tb*MSP-C and r*Tco*MSP-C Constructs 94
4.2.7 Recombinant Expression of *Tb*MSP-C, r*Tb*MSP-C and r*Tco*MSP-C. 95
4.2.8 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western Blot Analyses. 96
4.2.9 Purification of Truncated (t)*Tb*MSP-C, r*Tb*MSP-C and *rTcoMSP*-C 96
4.2.10 Protein Concentration Determination 98
4.2.11 Gelatin Zymography of t*Tb*MSP-C 99
4.2.12 Molecular Exclusion Chromatography of r*Tb*MSP-C and r*Tco*MSP-C 99
4.2.13 Hydrolysis of H-Suc-Leu-Tyr-AMC Peptide Substrate by r*Tb*MSP-C and r*Tco*MSP-C 99
4.2.14 Inhibition of r*Tb*MSP-C and r*Tco*MSP-C 99
4.2.15 pH Profile of r*Tb*MSP and r*Tco*MSP-C 100
4.2.16 Detection of r*Tco*MSP-C with Infected Cattle Sera 100
4.3 Results 100
4.3.1 Genomic DNA Isolation 100
4.3.2 PCR of *TbMSP-C* and Cloning into pGEM®-T and pGEX 4T-1 Vectors 101
4.3.3 Expression of *Tb*MSP-C (t*Tb*MSP-C*)* in *E. coli* 103

4.3.4 Gene Sequencing Analyses 104
4.3.5 Anti-Peptide Antibody Specificity against t*Tb*MSP-C 104
4.3.6 Purification and Activity of t*Tb*MSP-C 105
4.3.7 Transformation of Synthesised Constructs: r*Tb*MSP-C and r*Tco*MSP-C 107
4.3.8 Expression of r*Tb*MSP-C and r*Tco*MSP-C 108
4.3.9 Purification of r*Tb*MSP-C and r*Tco*MSP-C 109
4.3.10 Analysis of the Multimerisation of r*Tb*MSP-C and r*Tco*MSP-C. 111
4.3.11 Hydrolysis of H-Suc-Leu-Tyr-AMC Peptide Substrate 112
4.3.12 Anti-Peptide Antibody Specificity against r*Tco*MSP-C 115
4.3.13 Detection of Antibodies against r*Tco*MSP-C in Infected Cattle Sera 116
4.4 Discussion 116
Chapter 5: General Discussion 120
Appendix 1 129
Appendix 2 133
Appendix 3 134
Appendix 4 135
Appendix 5 136
Appendix 6 144
Appendix 7 147
References 149

List of Figures

Figure 1.1: Classification of trypanosomes 3

Figure 1.2: Diagrammatical representation of the morphology of a trypanosome. 6

Figure 1.3: Schematic representation of the life cycle of *T. brucei*. 8

Figure 1.4: HAT prevalence (A) and the overlap of cattle and tsetse fly distribution (B) in Africa. 11

Figure 1.5: Genomic organisation of Major Surface Proteases within various Trypanosomatidae species 20

Figure 1.6: Schematic representation of conserved regions within MSPs. 26

Figure 1.7: Crystal structure of leishmanolysin. 30

Figure 1.8: General mechanism of catalytic cleavage by thermolysin zinc metalloprotease 31

Figure 2.1: Schematic diagram of I-TASSER hierarchical method of three-dimensional protein modelling 38

Figure 2.2: Flow diagram outlining the methods used in the work reported in this chapter. 41

Figure 2.3: Phylogenetic representation and genomic organisation of *T. vivax* MSP protein sequences in relation to MSP classes from *T. brucei, T. congolense*, *T. cruzi* and leishmanolysin. 43

Figure 2.4: *Tv*MSP-A modelled by I-TASSER modelling program 46

Figure 2.5: *Tv*MSP-C modelled by I-TASSER modelling program. 48

Figure 2.6: *Tv*MSP-D modelled by I-TASSER modelling program. 49

Figure 2.7: *Tv*MSP-E modelled by I-TASSER modelling program 51

Figure 3.1: Schematic representation of a complete antibody and various antibody fragment types. 57

Figure 3.2: Schematic representation of the assembly of wild-type M13 phage coat proteins 57

Figure 3.3: Schematic representation of the use of phage display technology to select for scFv specific fragments. 59

Figure 3.4: Predict 7 plots of selected MSP peptides and their predicted positions in a homologous protease. 72

Figure 3.5: Sequence alignment of each peptide with full length MSP-B and –C from *T. brucei (*Tb), *T. congolense* (Tco) and *T. vivax* (Tv)..... 73

Figure 3.6: Analysis of antibody production in chickens against *Tb/Tco*MSP:303-314, *Tb*MSP:400-412 and *Tv*MSP:686-697 peptides by ELISA. 75

Figure 3.7: Elution profiles of anti-peptide antibodies from peptide-SulfoLink® affinity columns..... 76

Figure 3.8: Recognition of *Tb/Tco*MSP:303-314 peptide by corresponding affinity purified chicken anti-peptide antibody in an ELISA. 78

Figure 3.9: Recognition of *Tb*MSP:400-412 peptide by corresponding affinity purified chicken anti-peptide antibody in an ELISA. 79

Figure 3.10: Recognition of *Tv*MSP:686-697 peptide by corresponding affinity purified chicken anti-peptide antibody in an ELISA. 80

Figure 3.11: Polyclonal phagemid ELISA of panning rounds 1 to 4. 81

Figure 3.12: Monocolony scFv ELISA on 96 single colonies from round 4 of panning against *Tb/Tco*MSP:303-314. 82

Figure 3.13: Monocolony scFv ELISA on 96 single colonies from round 4 of panning against *Tb*MSP:400-412. 83

Figure 3.14: Monocolony scFv ELISA on 96 single colonies from round 4 of panning against *Tv*MSP:686-697. 83

Figure 3.15: Detection of MSPs in parasite lysate, using anti-peptide antibodies. 85

Figure 4.1: PCR cycling conditions used to amplify *TbMSP-C*. 92

Figure 4.2 Bradford standard curve used for protein quantification. 98

Figure 4.3: Genomic DNA extraction from procyclic (Lister 427) and blood stage (Litat 1.3) form *T. brucei* strains, evaluated on 1% (w/v) agarose gels. 101

Figure 4.4: Polymerase chain-reaction of *TbMSP-C* on both procyclic and blood stage form *T. brucei* genomic DNA, evaluated on 1% (w/v) agarose gels. 101

Figure 4.5: Undigested and digested pGEM®-T-*Tb*MSP-C and expression vector pGEX 4T-1 using EcoRI and Not 1, analysed on 1% (w/v) agarose gel. 102

Figure 4.6: Determining whether the expression vector is recombinant, as analysed on 1% (w/v) agarose gel. 103

Figure 4.7: Expression of (t)*Tb*MSP-C (recombinant pGEX 4T-1) in comparison with non-recombinant pGEX 4T-1. 104

Figure 4.8: Detection of t*Tb*MSP-C with anti-peptide antibodies. 105

Figure 4.9: Solubilisation of t*Tb*MSP-C as analysed on a 10% SDS-PAGE gel. 106

Figure 4.10: Purification of recombinantly expressed and solubilised t*Tb*MSP-C.....106

Figure 4.11: Protease activity of purified t*Tb*MSP-C as analysed on a 0.1% (w/v) gelatin zymogram stained with Coomassie blue. 107

Figure 4.12: Restriction digestion of r*Tb*MSP-C and r*Tco*MSP-C recombinant *E. coli* cells analysed on 1% (w/v) agarose gel. 108

Figure 4.13: Expression and solubility of r*Tb*MSP-C and r*Tco*MSP-C. 109

Figure 4.14: Purification of r*Tb*MSP-C and r*Tco*MSP-C. 110

Figure 4.15: Analysis of r*Tb*MSP-C and r*Tco*MSP-C under non-denaturing conditions. 111

Figure 4.16: Molecular exclusion chromatography of r*Tb*MSP-C and r*Tco*MSP-C.....112

Figure 4.17: Hydrolysis of peptide substrate H-Suc-Leu-Tyr-AMC by r*Tb*MSP-C and r*Tco*MSP-C. 113

Figure 4.18: Inhibition of r*Tb*MSP-C and r*Tco*MSP-C by metalloprotease inhibitor, EDTA and 1,10 phenathroline 114

Figure 4.19: pH profile of r*Tb*MSP-C and r*Tco*MSP-C.....114

Figure 4.20: Detection of r*Tco*MSP-C with anti-peptide antibodies 115

Figure 4.21: Detection of r*Tco*MSP-C with infected cattle sera. 116

Figure A1: Amino Acid sequence alignment of all Major Surface Proteases from *T. brucei, T. congolense, T. vivax, T. cruzi* and leishmanolysin from *L. major.* 132
Figure A2: Comparison of sizes (kDa) of various MSP protein, between different species (A) or between different families (B). 133
Figure A3: Secondary structure prediction of *TbMSP-C* using Quick 2D software. 134
Figure A6: Cloned TbMSP-C gene sequencing compared to actual sequence including translated protein sequence. 146
Figure A7.1: pGEM®-T vector map. 147
Figure A7.2: pGEX 4T-1 vector map. 147
Figure A7.3: pET100/D-TOPO vector map. 148

List of Tables

Table 4.1: Primers designed for amplification of *TbMSP-C*.....92
Table A1: Sequence identities of all Major Surface Proteases from *Trypanosoma brucei, T. congolense, T. vivax* and *T. cruzi* including leishmanolysin. 129
Table A2: Total number of amino acid residues and size of various Major Surface Protease from *T. brucei, T. congolense, T. vivax* and *T. cruzi* including leishmanolysin. 133
Table A4: MSPs predicted to be GPI-anchored or contain a transmembrane domain. 135
Table A5.1: Secondary structural analyses on identified TvMSP classes: comparison between four bioinformatical databases. 136
Table A5.2: Secondary Structure Analysis: PredGPI and DAS-TMfilter programs. 143
Table A5.3: Disulfide bond prediction by Scratch Protein Predictor. 143

Abbreviations

2xYT	2 x yeast tryptone
AAT	animal African trypanosomoasis
ABTS	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)
AMC	7-amino-4 methylcoumarin
amp	ampicillin
AP	alkaline phosphatase
APMA	4-Aminophenylmercuric acetate
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
CATT	card agglutination test
CNS	central nervous system
CSF	cerebrospinal fluid
C-terminal	carboxyl terminal
dH_2O	distilled water
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
Ellman's reagent	5,5'-dithio-bis-(2-nitrobenzoic acid)
EST	expressed sequence tag
Fab	fragment antigen binding
Fc	fragment crystallisable
FIA	fluorescent immunoassay
Fv	fragment variable
GPI	glycosyl-phosphatidylinositol
GPI-PLC	GPI-specific phospholipase C
GSS	genome survey sequence
GST	glutathione S transferase
h	hour/s
HAT	human African trypanosomiasis

HRPO	horse radish peroxidase
IFA	immunofluorescence assay
Ig	immunoglobulin
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISG	Invariable surface antigen
I-TASSER	iterative threading assembly refinement
kan	kanamycin
kb	kilo base
kDa	kilodalton/s
Lm	*Leishmania major*
MBS	m-Maleimidobenzoyl-N-hydroxysuccinimide ester
MCS	multiple cloning site
MHC	myosin heavy chain
min	minute/s
Mr	relative molecular mass
mRNA	messenger ribonucleic acid
MSP	Major Surface Protease
NCBI	National Centre of Biotechnology Information
N-terminal	amino terminal
OD	optical density
ODC	ornithine decarboxylase
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline, pH 7.2
PCR	polymerase chain reaction
PEG	polyethelene glycol
pI	isoelectric point
POP	prolyl oligopeptidase B
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RSA	rabbit serum albumin
RT	room temperature of 25^{o}C
r*Tb*MSP-C	synthesised constructed *Tb*MSP-C
r*Tco*MSP-C	synthesised constructed *Tco*MSP-C
s	second/s
scFvs	single chain variable fragments
SDS	sodium dodecyl sulfate

ss/dsDNA	single-stranded/double stranded DNA
Suc	succinyl
TAE	Tris-acetate-EDTA
TBS	Tris Buffered Saline
TM	trade mark
TNF	tumour necrosis factor
tris	2-amino-2-(hydroxymethyl)-1.3-propandiol
tRNA	transfer ribonucleic acid
t*Tb*MSP-C	truncated *Tb*MSP-C that was cloned here
TYE A/G	tryptone yeast extract ampicillin/glucose
UTRs	untranslated regions
VSG	variable surface antigen
WHO	World Health Organisation
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1 Literature Review

The Trypanosome Parasites and their Major Surface Proteases.

1.1 History of Trypanosomiasis

In the early 1900s two Congolese patients were shipped to London for observation and treatment of an unknown disease to which they succumbed. Mott conducted autopsies of their brains and identified many abnormalities including large abnormal plasma cells later known as the "morula cells of Mott" (Mott, 1908). This marks the first study done on humans suffering from human African trypanosomiasis (HAT-sleeping sickness); however, it took another four years to connect sleeping sickness with trypanosomes.

In the mid 1800s trypanosome parasites were identified in frogs (Kudo, 1922) and trout (Valentin, 1841) and in 1843, Gruby denoted the genus *Trypanosoma* (Gruby, 1843; Hoare, 1964). It was not until 1901 that live parasites were identified in a man suffering from recurring fever in Gambia (Baker *et al.*, 1902). Dutton identified the parasites as trypanosomes and in his publication in 1902 named them *Trypanosoma gambiense* (Dutton, 1902).

Only after Castellani's 1903 studies on the parasites morphology was it determined that *T. gambiense* (from trypanosome fever patients) and the parasites (from sleeping sickness) are of the same genus (Manson *et al.*, 1903; Steverding, 2008). These parasites were named *T. ugandensis* and Castellani hypothesised that sleeping sickness and trypanosome fever are the same disease.

After large numbers of cattle deaths in Zululand in the late 19th century, Sir David Bruce was asked to investigate the cause of the disease. In 1894, he showed in a series of experiments that tsetse flies (*Glossina spp.*) were able to transfer the disease from an infected dog to a healthy dog and that all infected animals had small parasites in their blood (Bruce, 1895). He gave definitive proof that the disease was transferred by the insect vector. This parasite was later named *Trypanosoma brucei brucei* (Steverding, 2008).

Bruce also did extensive comparisons of parasites from Uganda and Gambia and determined that they were identical proving finally that sleeping sickness and trypanosome fever are the same disease caused by trypanosome parasites.

1.2 Classification of Trypanosomes

Trypanosomes are protozoan parasites that cause diseases in both humans and animals (Uilenberg, 1998). Generally, their life cycle involves stages in both insect vector and mammalian host. Trypanosome parasites belong to the order Kinetoplastida since they have collections of DNA that form kinetoplasts, in addition to the nucleus (Stevens *et al.*, 2004). They fall under the family Trypanosomatidae which includes other parasites like those of the genus *Leishmania* and *Crithidia* (Figure 1.1). *Leishmania* is an intracellular parasite that causes leishmaniasis in Central and South America (Farrell, 2002), while *Crithidia* resides within a vector only (Smyth, 1994). The genus *Trypanosoma* is divided into two groups, depending on how the parasites are transmitted: Salivaria and Stercoraria (Lloyd and Johnson, 1924). Salivarian trypanosomes are transferred to the host via the bite of the infected vector whereas, Stercorarian trypanosomes develop in the hindgut of the vector and are posteriorly transmitted i.e. via contact with excreted faeces (Enyaru *et al.*, 2010). Salivarian trypanosomes also typically exhibit antigenic variation owing to the presence of variant surface glycoproteins (VSGs) on their surface (Stevens *et al.*, 2004). Trypanosomal parasites are highly diverse and contain a number of subgenera, species and subspecies (Figure 1.1). In this review, some significant trypanosome species will be discussed in more detail.

An important subgenus of the group Stercoraria*, Schizotrypanum,* contains a significant species, *T. cruzi* that causes Chagas disease in humans in South America and Asia (Stevens *et al.*, 2004). A number of other species are included in this subgenus like *T. rangeli,* which is non-pathogenic to humans, domestic and sylvatic animals (Guhl and Vallejo, 2003) and *T. dionisii* and *T. myoti,* which are restricted to bats (Klimpel and Mehlhorn, 2014). Morphologically speaking, these species cannot be distinguished because they have a small bloodstream form size of 14-24 nm. Therefore, other methods such as isoenzyme analysis, randomly amplified polymorphic DNA (RAPD) and other genetic markers have been used to distinguish these species. In South America and Asia, *T. cruzi* is prevalent and is pathogenic to its hosts: humans, pigs, dogs as well as wild mammals. In humans, Chagas disease results in damage to the heart and central nervous system (Stevens *et al.*, 2004). It is non-pathogenic to its vector, the sandfly, and has also been found in leeches and ticks (Vallejo *et al.*, 2009; Telleria and Tibayrenc, 2010).

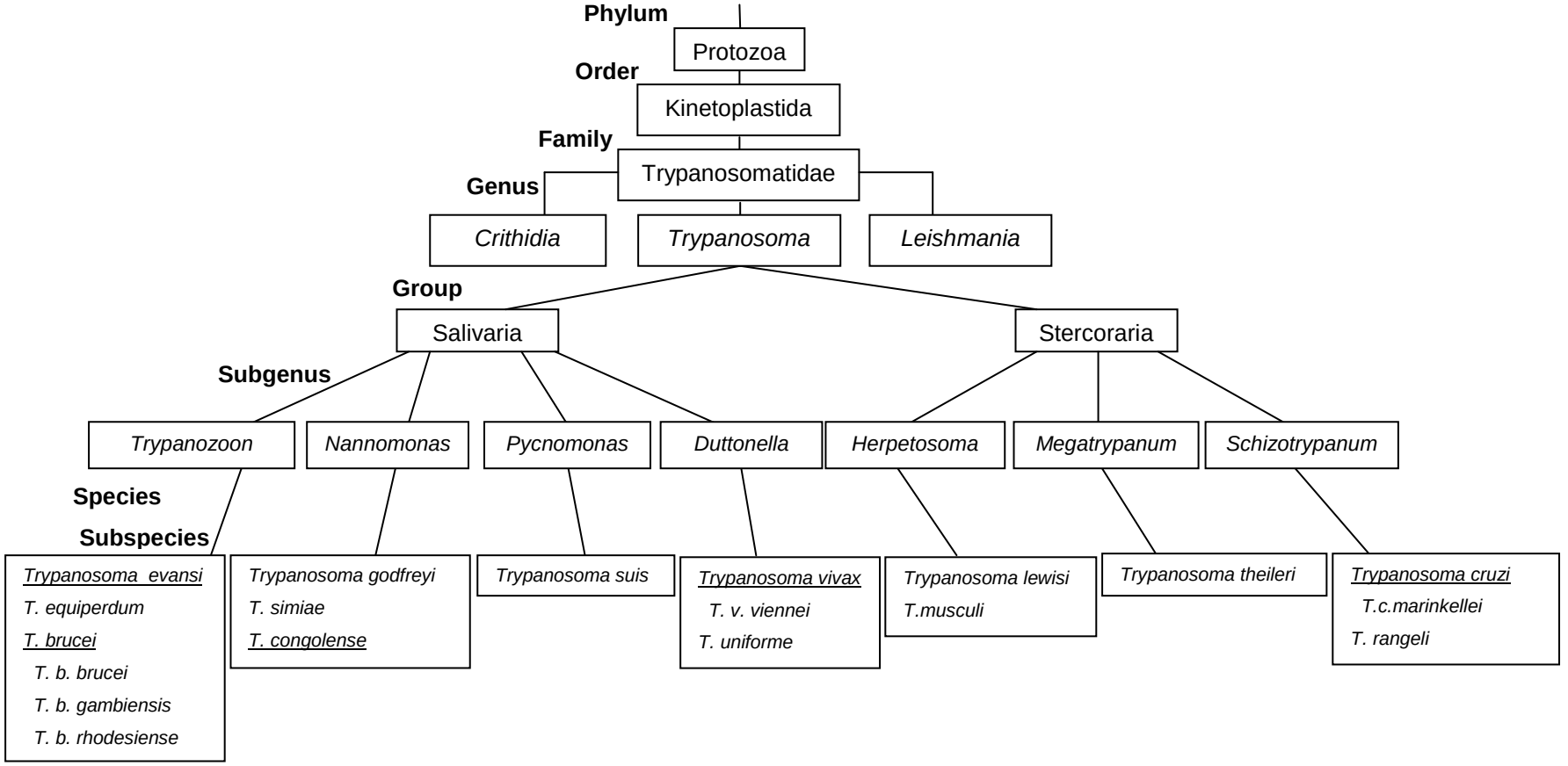


Figure 1.1: Classification of trypanosomes (Stevens and Gibson, 1999; Stevens *et al.*, 2004; Baral, 2010)
Species of importance in this review are underlined

Of the Salivaria, the subgenus *Duttonella* contains the trypanosome species, *T.* vivax, which shows rapid and distinctive movement within a wet blood film typical of this subgenus (Uilenberg, 1998). All these parasites develop within the vector's proboscis only (Foil, 1999). *T. vivax*, infects domestic and wild ungulates and is one of the key trypanosome species causing AAT (Ashford, 2001). It is unique in its length averaging around 21-25 µm. Its subspecies *T. v. viennei* is thought to have spread to South America and the Caribbean with the importation of zebu cattle from West Africa (Stevens *et al.*, 2004).

Subgenus *Nannomonas* contains three species of parasite: *T. congolense, T. simiae* and *T. godfreyi,* of which, *T. congolense* is of most interest. They are characteristically small parasites (8-24 µm) that lack a free flagellum (Hoare, 1964). *T. simiae* and *T. godfreyi* infection is limited to suids while *T. congolense* infects a range of mammals including but not limited to bovines, equines, sheep, goats and pigs (Simpson and Casas, 2009). *T. congolense* is probably the most commonly encountered pathogen of African livestock and results in a fatal disease called N'gana (Namangala and Odongo, 2014). For this reason, *T. congolense* is of significant importance in sub-Saharan Africa, since, the death of cattle causes economic loss (Namangala and Odongo, 2014).

Trypanozoon, another important subgenus of *Trypanosoma* includes three species *T. brucei, T. evansi* and *T. equiperdum.* They are morphologically indistinguishable but have unique pathological and genetic characteristics (Stevens *et al.*, 2004). *T. brucei*, a polymorphic parasite (Uilenberg, 1998), is one of the most significant species occurring on the African continent. The two subspecies, *T. b. rhodesiense* and *T. b. gambiense* cause HAT or sleeping sickness (Namangala and Odongo, 2014). A third subspecies, *T. b. brucei*, infects cattle and other domestic mammals but is non-pathogenic (Smyth, 1994). *T. evansi*, occurring in North African, the Middle East, Asia and South America, is transmitted by biting flies and causes disease in both domestic and wild animals (Desquesnes *et al.*, 2013; Namangala and Odongo, 2014). This species, unlike *T. brucei,* is monomorphic and only consists of long slender forms (Ashford, 2001). This parasite is able to infect a diverse number of mammalian hosts with varying degrees of severity. *T. equiperdum* is transmitted venereally between horses causing the disease dourine (Mohler and Schoening, 1920). The disease has largely been eradicated due to drug treatment, animal movement legislation and culling of infected individuals (Bowman, 2006).

The present study concentrates on the African trypanosome species, *T. brucei, T. congolense and T. vivax,* in the context of their effects on humans and cattle in sub-Saharan Africa.

1.3 Biology of the Trypanosome

Although, the genus *Trypanosoma* is diverse, these species share common genetic and morphological features. What follows is a general discussion on the genome and morphology of this genus of parasites, with specific reference to African trypanosome species.

1.3.1 The Genome of African Trypanosomes

Trypanosome parasites' genetic material is found in two organelles: the nucleus and the kinetoplast where DNA replication occurs simultaneously. The nuclear genome of *T. brucei* is around 35 million bp (Mb) in length with sizes varying between different isolates. This is expected to be the result of the presence of varying numbers of variant surface glycoprotein (VSGs) genes. The size of the nuclear genome of African trypanosomes is comparable with that of other parasites such as *Leishmania major* (32.8 Mb) and *Crithidia fasciculata* (32 Mb). The *Trypanosoma* genome contains 20 chromosomes (0.2 to 6 Mb) and >100 ~50 kb minichromosomes (Ersfeld and Gull, 1997). It is suspected that the genome is of low complexity because more than 50% of the genome contains single copy sequences (El-Sayed *et al.*, 2000). Uniquely, the VSG genes and expression sites are haploid, which is most likely due to the high degree of variety required in coding for the protein coat. The entire *T. brucei* genome has been completely sequenced (Berriman *et al.*, 2005) as well as a large portion of the *T. congolense and T. vivax* genome.

The kinetoplast, so named because it was originally thought to be involved in the movement of the parasite (*kineto* meaning movement), contains genomic material. Within the kinetoplast there are maxicircles (~50 copies/cell) and minicircles (~10 000 copies/cell) of DNA (Madison-Antenucci *et al.*, 2002). The maxicircles encode mitochondrial proteins, whereas the minicircles contain guide RNAs which act as templates for the editing of maxicircle DNA (Liu *et al.*, 2005).

1.3.2 Morphology of Trypanosomes

Trypanosomes are unicellular organisms and classification is based on the following morphological characteristics: presence of a single flagellum (be it free or not) and the presence of the kinetoplast (Hoare, 1964). These organisms contain general cellular components found in most eukaryotic cells: nucleus, cytoskeleton, endoplasmic reticulum, Golgi apparatus and mitochondria (Figure 1.2). Specifically, African trypanosomes show two morphological forms: the trypomastigote form (blood stream form within the mammalian host) and the epimastigote [in the vector; (Uilenberg, 1998)].

The polymorphic African trypanosomes are between 11-42 µm in length and 1-3 µm in width (Stevens *et al.*, 2004). The different forms depend on what part of the life cycle the parasites are in and whether they are infective or not (see section 1.4). These parasites are designed for motility within plasma, blood and other tissues within the host and vector. For this reason, they are generally streamlined in shape and tapered at both ends. They also contain a highly polarised microtubule cytoskeleton (Figure 1.2) with minus ends placed anteriorly while plus ends are positioned posteriorly (Robinson *et al.*, 1995). The single copy organelles are located specifically within this cytoskeleton. Parasites also contain a pellicle (thin layer of cells supporting the cell membrane) which has evolutionarily been designed to keep the shape of the parasite while still allowing enough flexibility for movement.

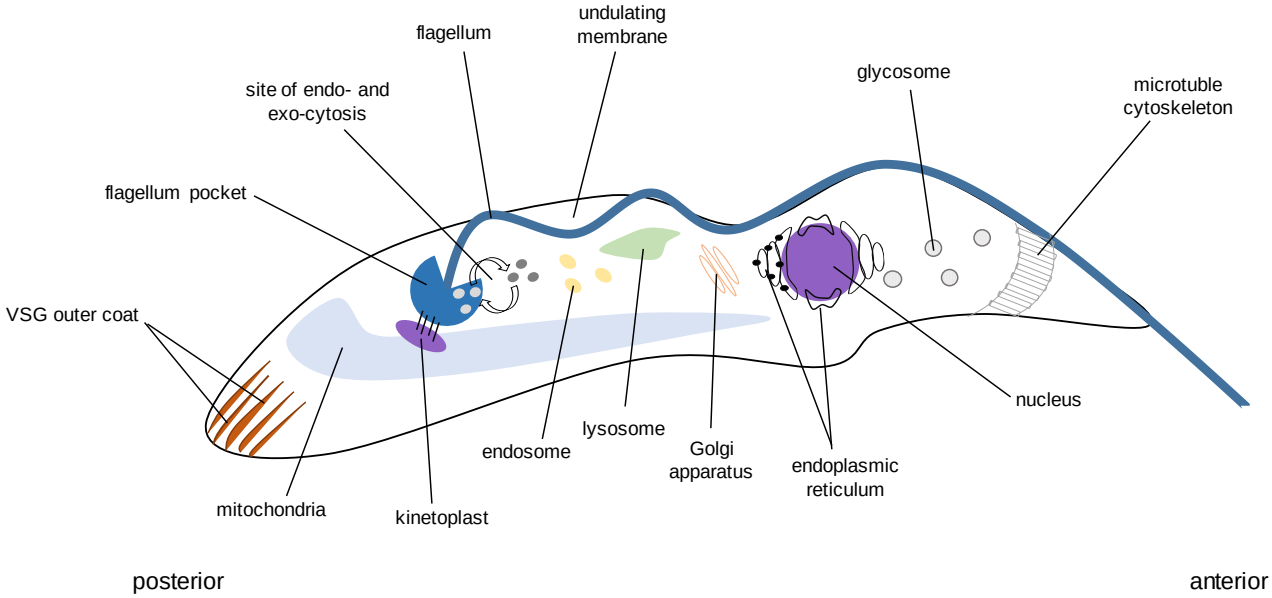


Figure 1.2: Diagrammatical representation of the morphology of a trypanosome.

A single flagellum, essential for parasite movement, is present which begins posteriorly from the flagellum pocket and travels along the length of the parasite (Bastin *et al.*, 1998). It can protrude anteriorly as a free flagellum. Whether a parasite has a free flagellum or not and the length of it again depends on the species and the life cycle stage it is in (Vickerman, 1985). The flagellum pocket is the only site of endo- and exo-cytosis for the parasite (Figure 1.2) and hence plays an important role in cell survival (Overath *et al.*, 1997). The pellicle and cytoplasm are pinched along the body of the parasite to form an undulating membrane. The kinetoplast is generally situated at the posterior end and is adjacent to the parabasal body [from where the flagellum arises; (Figure 1.2)]. This defining organelle can vary in size and position within different life cycle stages and species (Vickerman, 1985; Vickerman *et al.*,

1988). Characteristically, African trypanosomes (and all Salivaria) contain a variant surface glycoprotein (VSG) coat over their entire surface, except the flagellum pocket, during the trypamastigote forms (Stevens *et al.*, 2004). Movement from mammalian host to insect vector occurs in conjunction with 'coat shedding' i.e. rapid removal of glycoproteins from the surface (Bülow *et al.*, 1989). This allows for the development of the procyclin coat present during the insect vector life cycle stage.

1.3.3 Antigenic Variation

One of the most extraordinary processes African trypanosomes possess is their ability to change their coat proteins (VSGs) periodically during the trypomastigote forms (bloodstream form) to evade the host immune system. This is done via a process called antigenic variation, in which trypanosomes produce an enormous range of antigenically distinct VSGs (Taylor and Rudenko, 2006). These distinct VSGs have varying levels of expression. As the host immune response develops against a particular VSG coat, parasites containing a homologous coat will decline; however, those parasites that are heterologous will survive, divide and initiate another wave of parasitaemia within the infected individual (Barry and Carrington, 2004; Mugnier *et al.*, 2016). This variation of the protein coat is based on the switching of expression between various VSG genes and is comparable to immunoglobulin population development in higher eukaryotes (Robinson *et al.*, 1999). Due to this phenomenon, conventional vaccines against trypanosome infection have been unsuccessful and novel vaccine development methods are being researched (Black and Mansfield, 2016).

1.4 Life Cycle of the African Trypanosome

The trypanosome parasite life cycle involves the transmission from invertebrate vector to vertebrate host and back again and requires changes in morphology as part of a complex life cycle. This life cycle involves alternative proliferative and quiescent stages. The only stage capable of differentiation is the quiescent stage and, hence, is infective (Matthews, 1999). The infective quiescent metacyclic forms are found within the salivary glands of the tsetse fly [Figure 1.3; (Vickerman *et al.*, 1988)]. During the insect bite, saliva is injected into the host which contains anti-thrombotic, anti-platelet aggregating and anti-coagulation activities that prevent blood clotting and induce vasodilation during the blood meal (Telleria *et al.*, 2014). Infective trypanosome parasites are transferred and initially divide at the site of infection to form a chancre, after which, they are released directly into the bloodstream via lymph nodes and vessels (Vickerman, 1985).

Once in the bloodstream, *T. brucei* undergoes differentiation from a metacyclic form to a proliferative trypomastigote form (Vickerman, 1985). This form is generally long and slender and divides rapidly within the blood where it is able to spread to different organs within the host (Figure 1.3). Subsequently, these proliferative forms differentiate into a quiescent trypomastigote form which is infective and are ingested by tsetse flies feeding on an infected host. This non-dividing blood stream form of the parasite is generally short and stumpy [Figure 1.3; (Matthews, 2005)].

Parasites are now in the midgut of the fly where they differentiate into the procyclic form (Figure 1.3) via rapid VSG coat shedding performed by a GPI-anchor specific phospholipase C (GPI-PLC) and a major surface metalloprotease (LaCount *et al.*, 2003; Gruszynski *et al.*, 2006; Grandgenett *et al.*, 2007). Procyclic parasites develop a procyclin coat within the insect vector (Matthews *et al.*, 2004). At this point crucial adjustments of enzyme systems used by the parasite take place so that different nutrients can be used at the different temperatures within the vector and host. One of the most significant changes is the development of a functioning mitochondrion to provide energy via oxidiative phosophorylation. In comparison, the bloodstream form of the trypanosome uses glycolysis (Fenn and Matthews, 2007).

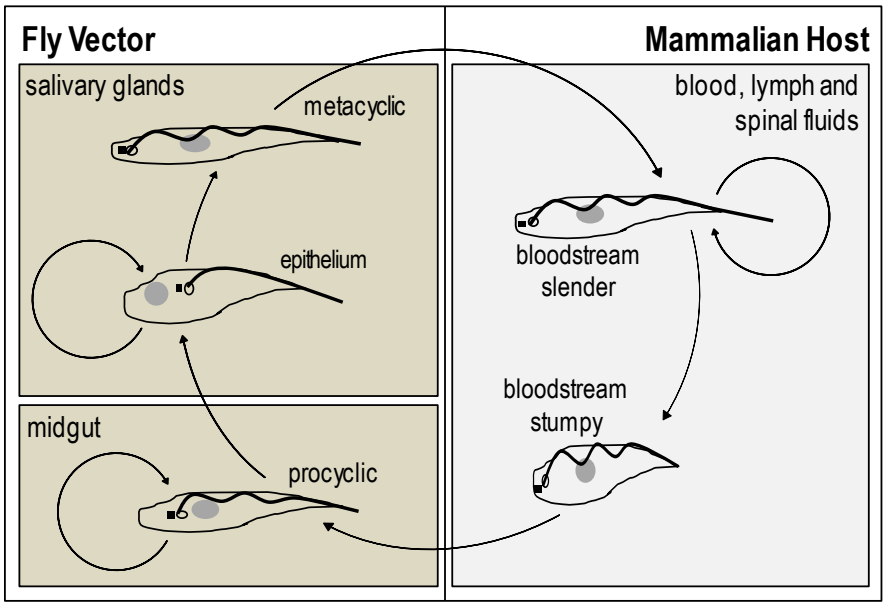


Figure 1.3: Schematic representation of the life cycle of *T. brucei*. Proliferative forms , Grey circle within the parasite represents the nucleus; note how its position changes with different life cycle stages [Adapted from Wheeler (2010)].

After ingestion by the tsetse fly, many parasites begin to die and this may lead to elimination of infection in some flies (Fenn and Matthews, 2007). Parasites that survive within the fly undergo differentiation and migration.

In the case of *T. brucei*, procyclic forms migrate to salivary glands and differentiate into epimastigote forms which become attached to the microvilli within the gland [Figure 1.3; (Vickerman *et al.*, 1988)]. Within the salivary glands, epimastigotes differentiate into the metacyclic form which is characterised by the redevelopment of the VSG coat (Vickerman *et al.*, 1988). Once the VSG coat has developed fully, parasites detach from the cell wall of the tsetse fly salivary glands and are able to infect the host; thus, perpetuating the life cycle (Figure 1.3).

By contrast, *T. congolense* develops into the procyclic form once it has migrated to the insect proboscis and after differentiation it enters the salivary glands while *T. vivax* develops entirely in the proboscis and no differentiation occurs within the midgut or the salivary gland (Matthews *et al.*, 2004).

1.5 African Trypanosomiasis

African trypanosomiasis (both HAT and AAT) is endemic in 36 African countries and a study published in 2011, which examined the risk factor over a ten year period, concluded that in central Africa about 3.5 million people are at risk of contracting HAT (Simarro *et al.*, 2011). The decline of HAT over the last decade has promoted the World Health Organisation (WHO) in conjunction with Pan African Tsetste and Trypanosomiasis Eradication Campaign (PATTEC) to implement eradication strategies and at the beginning of 2014 announced that they hope to completely eliminate *T. b. gambiense* by 2020 (this is less than 1 case per 10000 people in at least 90% of all endemic areas).

Within central and northern Africa, livestock farming especially cattle plays an important role in the lives of people as a source of food, income and stature (Herrero *et al.*, 2013). It has been known since the re-emergence of the disease in the 1980s, that the problem with animal trypanosomiasis needs to be resolved to ensure the economic development of Africa. The loss caused by this disease for the cattle industry within Africa is estimated as US$ 1.3 million annually (Maitima, 2007).

The presence of the tsetse fly and the disease generally overlaps geographically on the African continent (Figure 1.4). However, cattle farmers often migrate with their livestock into and out of tsetse fly areas where cattle become infected and carry the disease with them. It is here where the fact that biting flies (tabinids) can also carry the disease becomes important because, they aid in the spread of the disease even in tsetse fly free zones (Stevens *et al.*, 2004).

A study conducted to map the economic benefit of treatment and prevention of bovine trypanosomiasis determined that due to high costs and high chance of re-infection, only farmers that are geographically located on the fringe of infection zones will be likely to profit (Shaw *et al.*, 2014). Current assessment of sub-Saharan African communities, with prevalence of AAT, suggests that locally-adapted treatment and control programmes are required for the successful management of the disease (Holt *et al.*, 2016).

1.5.1 Pathogenesis

T. brucei rhodesiense and *T. b. gambiense* HAT, manifest differently within human patients as either an acute or chronic form. The biggest difference being that *T. b. rhodesiense* HAT may be fatal within a few months of infection (acute), due to heart damage, while *T. b. gambiense* HAT may be present for years after initial infection (chronic) with death resulting from other causes, such as, central nervous system damage and opportunistic infections (Pentreath and Kennedy, 2004).

The infections leading to HAT are divided into two stages: early and late. Early haemolymphatic stage trypanosomiasis begins with the formation of a chancre at the bite site. This is often considered to be a boil and disregarded by the infected patient (Kennedy, 2004). Five to twelve days after a tsetse fly bite, the parasite has invaded the blood which is associated with symptoms such as fever, headache and joint pain (Pentreath and Kennedy, 2004). At this point, a difference in the number of parasites in the blood for *T. b. rhodensiense* and *T. b. gambiense* infected patients is observed. High levels of *T. b. rhodensiense* parasites are found (10^5/ml); whereas, *T. b. gambiense* has low numbers and often goes undetected. Parasites also invade the lymphatic system and fever cycles are observed every seven to ten days.

The late encephalitic stage of HAT is when the parasite invades the central nervous system (CNS) and cerebrospinal fluid (Kennedy, 2004). The acute form of the disease sees parasites invading the CNS within a couple of weeks following the initial infection and patients may die before this invasion even occurs. The chronic form is slower and symptoms like personality changes, insomnia and irritability are indicative of CNS invasion. Death by the chronic form of HAT is usually the result of opportunistic infections.

A range of animal species are susceptible to AAT or N'gana caused by a variety of trypanosome species that could be described as 'haematic' where the parasite stays within the blood vessels of the host (*T. congolense*) or 'humoral' where the parasite invades other tissue [*T. vivax* and *T b. brucei*; (Uilenberg, 1998)]. Pathogenesis of trypanosomiasis within animals is similar to that in humans. Generally, cattle become chronically ill and infection is

accompanied by a loss of milk production, emaciation and eventually death (Taylor and Authié, 2004). It should be noted that most African trypanosome species are able to infect wild animals with no effect on their health, hence, they serve as a reservoir of parasite to replenish the life cycle and continue to infect host animals (Steverding, 2008). This is one of the many problems in the treatment and eradication of the disease.

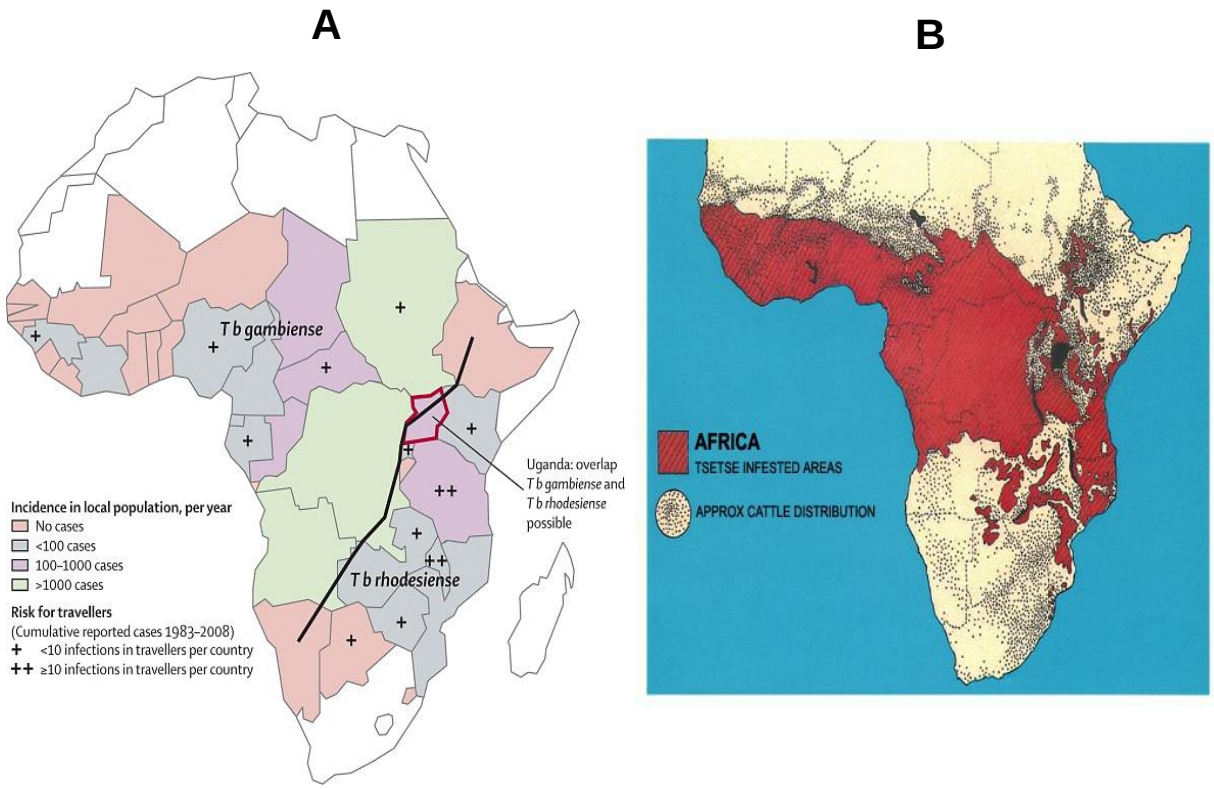


Figure 1.4: HAT prevalence (A) and the overlap of cattle and tsetse fly distribution (B) in Africa. Disease occurrence correlates with spread of the vector [Brun *et al.* (2010); http://www.irinnews.org/news/2009/05/12/tsetse-fly-costs-agriculture-billions-every-year].

1.5.2 Diagnosis

Diagnosis of trypanosomiasis, especially in humans, follows three general steps; screening, diagnostic confirmation and staging (Chappuis *et al.*, 2005). Diagnostic techniques can be either direct or indirect. Direct diagnostic methods involve either clinical or parasitological methods that detect the parasite in blood, lymph or cerebrospinal fluid (CSF) samples (Büscher and Lejon, 2004) while indirect diagnostic methods involve detection of factors produced either by the parasite or in response to it, such as antigens or antibodies.

There are specific approaches to developing diagnostic methods because of the different species and strains causing the disease, due to antigenic variation and the fact that the disease is most common in third world African countries with rudimentary infrastructure and depressed economy. The epidemiology of this disease has resulted in adapting WHO "ASSURED" criteria (**A**ffordable, **S**ensitive, **S**pecific, **U**ser-friendly, **R**apid and robust, **E**quipment-free and **D**eliverable to end-users) for the development of diagnostic methods (Peeling *et al.*, 2006).

Direct clinical diagnosis is an unreliable method because symptoms of trypanosomiasis resemble those of other tropical diseases such as babesiosis and theileriosis which can result in misdiagnosis. Even so, it remains one of the most widely used methods of diagnosis of AAT because often physical assessment is the only means available (Eisler *et al.*, 2004). In the case of HAT, this method has low sensitivity especially in identifying *T. b gambiense* which generally has low parasitaemia (Uilenberg, 1998). To increase sensitivity, various methods to concentrate parasite numbers have been developed such as removal of red blood cells using anion exchange chromatography (Nantulya, 1990), quantitative buffer coat technique (Bailey and Smith, 1994) and microhaematocrit centrifugation (Woo, 1970). These methods remain the most simple and inexpensive techniques.

Indirect methods of diagnosis include serological and molecular tests. Some serological tests are the Card Agglutination Test for Trypanosomiasis (CATT), indirect immunofluorescent antibody (IFA) test, and indirect and direct enzyme-linked immunosorbent assays (ELISAs). The CATT has been an important development in the diagnosis of *T. b gambiense*. This simple agglutination test detects *T. b. gambiense* specific antibodies from patients' blood or plasma (Magnus *et al.*, 1978). It contains lyophilised variable antigen type LiTat 1.3. antigens (Chappuis *et al.*, 2005). False negatives for this test can result if trypanosomes causing infection do not express the LiTat 1.3 gene; however, other tests like LATEX includes three different variant antigen types and has shown to have higher specificity (Penchenier *et al.*, 2003). However, it has lower sensitivity when compared to CATT. Both IFA and ELISA testing have shown promising diagnostic results because they have high sensitivity and specificity especially when antigens can be supplied at low cost and in high quantity (Uilenberg, 1998). However, these methods require sophisticated equipment which limits field use.

Lateral flow tests have also been developed that use a parasitic antigen for the detection of serum antibodies and include HAT SERO K-SeT (detects patient serum antibodies against *T. b gambiense* VSGs LiTat 1.3 and 1.5), HAT Sero-Strip (detects patient serum antibodies against *T. b. gambiense* VSGs LiTat 1.3 and 1.5), SD BIOLINE HAT (detects patient serum

antibodies against *T. b. gambiense* VSGs LiTat 1.3 and 1.5) and Surra SERO K-SeT [detects animal serum antibodies against *T. evansi* VSGs Rotat 1.2; (Büscher *et al.*, 2014; Birhanu *et al.*, 2015; Jamonneau *et al.*, 2015)]. A potential lateral flow test against the p310 invariant surface glycoprotein (ISG) antigen, for *T. vivax,* has given promising preliminary results (Fleming *et al.*, 2016). A disadvantage of antibody detecting serological tests is that cured or previously infected individuals may give positive results, if antibodies are still present (Garcia *et al.*, 2000; Solano *et al.*, 2002). Therefore, further research into antigen detecting serological tests is still required to distinguish between past and present infections.

Molecular tests involve various techniques to detect parasite DNA or RNA often with the use of polymerase chain reaction (PCR) or nucleic acid labelled probes. Advantages of these techniques are higher sensitivity and specificity even to subspecies level; however, sophisticated equipment is often required. Development of a PCR-oligochromatography dipstick has shown promising results since it is able to distinguish between *T. b. gambiense* and *T. b. rhodesiense* (Deborggraeve *et al.*, 2006; Büscher *et al.*, 2013). The 18S ribosomal RNA gene is a favourite in the detection of parasites using PCR diagnostics because of the high number of copies as well as the uniqueness of the genetic code (Desquesnes *et al.*, 2001). A two-year study published in 2011 by Deborggraeve *et al*. showed that PCR on blood and CSF may be a non-invasive technique to diagnose infected individuals with better or equal performance compared to other methods.

Diagnostic methods for the detection of African trypanosomiasis is continually being improved because of the regions where this disease occurs (Brun *et al.*, 2010). Also, current diagnostic techniques are often only effective in 95% of cases and as trypanosome infections drop below 5% of the population, more sensitive diagnostic methods are required (Chappuis *et al.*, 2005). It is clear that new diagnostic targets and techniques need to be developed to continue the fight against this disease that is a prominent burden in third world Africa.

1.5.3 Treatment and Control of Trypanosomiasis

1.5.3.1 Treatment

Treatment for HAT is stage specific; i.e., a different treatment is used for the early haematic and late CSF stages; although even today, there are only a few drugs available for HAT treatment.

Two main drugs used for early stage treatment are pentamidine and suramin (Brun *et al.*, 2010). Pentamidine is used in the treatment of *T. b. gambiense* HAT and is a general

anti-protozoal agent which is thought to disrupt nuclear metabolism by inhibiting DNA, RNA, lipid and protein synthesis (Nguewa *et al.*, 2005). Suramin, on the other hand, is used to treat first stage *T. b. rhodesiense* HAT and is a polyanionic compound that contains six electronegative sites and is thought to act by binding many important enzymes involved in respiration and glycolysis of the parasite (Hanau *et al.*, 1996; Barrett *et al.*, 2007).

Melarsoprol and eflornithine are used for the treatment of late stage AAT. Eflornithine is the only drug released in the last 25 years (released in 1990) for the treatment of late stage *T. b. gambiense* (Brun *et al.*, 2010). Nifurtimox, originally developed for the treatment of Chaga's disease, has been tested in combination with eflornithine for treatment of late stage *T. b*. *gambiense* as well (Priotto *et al.*, 2009). Melarsoprol can be used in the treatment of both chronic and acute HAT and is an arsenic-based anti-parasitic agent that binds thiol groups. Binding of thiol groups inhibits the action of many enzymes including those required for glycolysis (Hanau *et al.*, 1996) which are important in the production of adenosine triphosphate (ATP) and therefore energy within the parasite. It has also been shown to bind trypanothione, which provides a defence against damage by oxidants (Fairlamb *et al.*, 1989). Eflornithine specifically inhibits ornithine decarboxylase (ODC) within the parasite. This enzyme is important in the production of polyamines which play a significant role in growth, differentiation and cellular replication by actively participating in nucleic acid and protein synthesis (Van Bogaert and Haemers, 1989).

Diminazene aceturate, homidium (bromide and chloride), isometamidium chloride, quinapyramine (sulfate and dimethylsulfate), suramin and melarsomine dichlorohydrate are used either as chemoprophylaxis or chemotherapeutic agents for AAT (Holmes *et al.*, 2004) caused by *T. vivax, T brucei, T. evansi* and *T. simiae*. Prophylactic drugs are used as a preventative measure against infection in livestock. Diminazene has been shown to inhibit a trypanosome serine protease, oligopeptidase B (OPB), whereas isometamidium targets the kinetoplast DNA. Suramin was shown to inhibit OPB as well (Morty *et al.*, 1998). On the other hand, homidium interferes with glycosomal functions (Holmes *et al.*, 2004).

New drugs can be developed by looking at commercially available drug libraries that have chemical structures with the potential to be optimised against trypanosomiasis. Katiyar *et al.* (2013) identified kinases as new drug targets within trypanosomes and determined that current drugs such as 4-anilinoquinazoline and pyrrolopyrimidines have good scaffolds for further optimisation to develop new anti-trypanosomal drugs. However, drug development and testing is lengthy process that often ends in failure.

Since no new drugs have entered the market in the last 25 years, drug resistance has become a serious problem for the treatment of trypanosomiasis (Brun *et al.*, 2010).

Resistance against suramin for HAT is rare; however, it is seen for AAT. Resistance against melarsoprol and pentamidine has increased throughout Africa as well. Melarsoprol resistance occurred because the uptake of the drug is dependent on the P2 aminopurine transporter and resistant parasites have lost this transporter (Barrett *et al.*, 2007). A similar resistance mechanism is seen for pentamidine as well. Therefore, new pathogenic and virulence factors need to be identified within trypanosomes and a better understanding of the life cycle and differentiation of the parasite is required before new drug targets can be developed.

1.5.3.2 Control

Due to antigenic variation of trypanosomes and their immunosuppressive effect on the host immune system, conventional vaccination has been unsuccessful. Therefore, vector control is of utmost importance and numerous techniques and methods have been employed in sub-Saharan Africa to reduce risk and spread of the disease. Methods include aerial and ground spraying, artificial baits (Rayaisse *et al.*, 2010), insecticidal treatment of animals (Bouyer *et al.*, 2009) or insecticide-treated insect targets (e.g. Harrison trap) and the sterile insect technique (Vreysen, 2000). Tsetse flies are susceptible to most insecticides from the groups of organochlorines, organophosphate, carbamate, natural pyrethrum and synthetic pyrethroids. For over 40 years, the insecticide dichlorodiphenyltrichloroethane (DDT) was used throughout Africa from spraying against tsetse flies; however, it was later found that the long-lasting ability of this chemical was detrimental to the environment and was banned (Allsopp and Hursey, 2004). Aerial spraying is highly effective over large areas but may leave small areas untreated from where tsetse flies proliferate and spread. Ground spraying is labour-intensive and governments may not be able to afford the manpower for this method to be viable (Allsopp and Hursey, 2004).

Artificial baiting technologies, such as, odour or visual cue containing traps that contain insecticides have been shown to decrease tsetse fly populations. However, a study conducted by Rayaisse *et al.* (2010) showed that current traps employed in the Ivory Coast only kill ~50% of the tsetse flies attracted to the area. In an assessment of AAT vulnerability within sub-Saharan Africa, areas that show high vulnerability to the disease reported few traps or targets (Holt *et al.*, 2016). Since this method of control is much cheaper compared to the spraying of insecticide, it would be expected to have higher prevalence; however, traps require maintenance and are labour-intensive which could have had a negative impact on their availability. Improvements of these traps or targets are required, to have a significant effect on the tsetse fly population.

Animals can be treated with pour-on insecticide and/or repellent to avoid being bitten which has also been shown to be an effective vector control method and it minimises chemical spread to the environment (Bouyer *et al.*, 2009). However, in high density tsetse fly areas the efficacy of this method is diminishing. The need to supply farmers with dip that is inexpensive enough to be viable can also be a challenge. Synthetic and natural repellents (identified from wild animals that are generally not bitten by the tsetse fly) have been identified and are undergoing validation (Gikonyo *et al.*, 2003; Bett *et al.*, 2010).

Sterile insect technique when sterilised males are released into an area which slows population growth until it stops and starts to die out (Vreysen, 2000), has shown moderate success. Sterilised flies have to be continually released to counteract fertile males being born in the new generation. There has been successful use of this technique in eradicating the tsetse fly from the island Zanzibar (Vreysen *et al.*, 2000), as well as, for other insect species like the screwfly from North America and Libya (Lindquist *et al.*, 1992). Doubts have been expressed as to whether it will work over such large areas as in the case of the *Glossina* species habitat. It has potential in smaller pockets, such as, northern KwaZulu-Natal, where its use is in the planning stage.

1.5.4 Trypanotolerance

Animals that do not show symptoms of trypanosomiasis despite being infected are termed trypanotolerant. This is thought to be genetically predetermined due to the evolution of animal species with the parasite over thousands of years (Murray *et al.*, 2004). Wild animals such as zebra, buffalo, wildebeest and waterbuck are able to resist infection but are still carriers and act as reservoirs for the continuation of the life cycle of the parasite (Njiokou *et al.*, 2006). However, there are also species of cattle, like the N'Dama cattle, which are trypanotolerant (Murray *et al.*, 2004).

Trypanotolerant animals are able to reduce parasitaemia levels and self-cure and it has been shown that β_2-microglobulin plays an important role in this resistance. Hill *et al.* (2005) showed that during high parasitaemia increased expression of β_2-microglobulin, MHC-1 lymphocyte antigen and protein kinase C could all be important in the observed tolerance.

This resistant N'Dama cattle breed is, however, smaller than general cattle farming breed like Zebu cattle and therefore, the use of susceptible cattle is more common and widespread throughout Africa. Cross-breeding to select for resistant larger breeds should be considered by local farmers in conjunction with other control methods (d'Ieteren *et al.*, 1998).

Whether trypanotolerance occurs in humans is a question that still needs to be answered. Older literature reported sleeping sickness in self-curing patients but these studies used diagnostic tests that lack the advancements we have today and have been deemed unreliable (Jamonneau *et al.*, 2012). However, a 15 year follow-up study reported that ten out of eleven patients that refused treatment showed reduced parasitaemia as observed with microscopy and PCR, as well as, some showing reduced serological responses becoming negative for trypanosome variable antigens over time (Jamonneau *et al.*, 2012). Even more recently, it has been determined that pro-inflammatory cytokines are involved in this observed 'tolerance' with high levels of IL8 being associated with serologically negative patients while those with high levels of IL10 and low levels of TNF-α have higher risk of developing sleeping sickness (Ilboudo *et al.*, 2014). This may mean that the disease could follow an alternate progression of tolerance rather than being 100% fatal every time.

1.6 Leishmania, a Trypanosomatidae Parasite

There are many other parasitic species that fall under the family Trypanosomatidae. These species all have different mechanisms of infection, different vectors and different hosts. In the discussions that follow, *Leishmania* is included for comparitive purposes.

Leishmania causes the disease leishmaniasis which infects mainly vertebrate hosts like rodents and humans (Dedet, 2002). Its vector, the sandfly of the genus *Lutzomyia,* is found in subtropical and tropical countries of South America and South-East Asia. The parasite causes two types of disease within humans; cutaneous leishmaniasis, which causes sores and nodules on the skin, and visceral leishmaniasis, which infects different tissues of important organs such the liver or bone marrow (Dedet, 2002). Like Salivaria, parasites are injected by sandflies into the human host during a blood meal. Conversely, *Leishmania* are intracellular parasites that invade or are engulfed by macrophages within the host (Dedet, 2002). African trypanosomes are exclusively extracellular, while *Leishmania* is comparable to *T. cruzi* which is also intracellular. It must be noted here that the Major Surface Protease (the main focus of the present study) was first identified in this species of Trypanosomatidae and its function is to allow parasites to invade macrophages and gives them resistance against cell lysis within these host cells (McMaster *et al.*, 1994).

1.7 Major Surface Proteases (MSPs) in African Trypanosomes

1.7.1 Genomic Organisation of MSP Genes

The Major Surface Protease (MSP) or Glycoprotein 63 (gp63) was first discovered in the 1980s in *Leishmania mexicana amazonensis* (Chaudhuri and Chang, 1988) and identified as a pathogenic and virulence factor (Oliver *et al.*, 2012). Later on, homologs were found in 13 other *Leishmania* species, as well as, *Crithidia fasciculata* and *Herpetomonas samuel pessoai* (Etges, 1992). Genes identified as *LmMSP* homologs were reported in both African and American trypanosomes (El-Sayed and Donelson, 1997a; Cuevas *et al.*, 2003; Marcoux *et al.*, 2010), as well as *Cryptobia salmositica* (Jesudhasan *et al.*, 2007) of the order Kinetoplastida (Thomas and Woo, 1991). The presence of MSPs on the *T. brucei* cell surface proteosome was recently confirmed (Shimogawa *et al.*, 2015). Within many of the species shown to have *MSP* genes, each have a distinct number of classes which have been shown to differ primarily in the C-terminal coding regions and 3′ untranslated regions [UTRs; Ramamoorthy *et al.* (1992); Voth *et al.* (1998)]. In most cases, differential expression during the life cycle was observed (Ramamoorthy *et al.*, 1992; Voth *et al.*, 1998) and, for *T. brucei* MSP, this expression was regulated by the 3'-UTRs (Helm *et al.*, 2009).

The genomic organisation of *MSP* genes of *Leishmania major* will be compared to that of *Trypanosoma brucei*, *T. congolense* and *T. cruzi*. To date seven *MSP* genes have been found in the *L. major* genome and were originally called *gp63 1-7*. The first five *MSP* genes are highly conserved tandem repeats on chromosome ten (Ivens *et al.*, 2005). The 8-10 kb downstream genes six and seven are less conserved compared to the first five (Figure 1.5). Later, they were re-named *MSP-L1-5, MSP-C (gene six)* and *MSP-S [(gene seven);* Roberts *et al.* (1993); Voth *et al.* (1998)] and each have varying levels of expression at different life cycle stages of the parasite. For example, *MSP-L* is expressed only in promastigotes while *MSP-C* is expressed in both promastigotes and amastigotes (Ramamoorthy *et al.*, 1992). It has also been concluded that the different isoforms of MSP, within *Leishmania*, have unique *pIs* which are related to the biological function of the protease (Hsiao *et al.*, 2008).

By chance, in an experiment to compare genome survey sequencing (GSS) and expressed sequence tag (EST) shotgun sequencing techniques, El-Sayed and Donelson (1997b) discovered *Trypanosoma brucei* sequences that were similar to genes that code for *L. major MSP (gp63*). Subsequently, cDNA libraries were used to identify *T. brucei* homologs to *Leishmania* and *Crithidia MSPs* (El-Sayed and Donelson, 1997a).

As seen in Figure 1.5, three gene classes were discovered; namely, *TbMSP-A, -B* and *-C* (LaCount *et al.*, 2003). *TbMSP-A* gene class consists of five closely related copies (1-5) and is located on chromosome 11 (LaCount *et al.*, 2003; Berriman *et al.*, 2005; El-Sayed *et al.*, 2005). The *TbMSP-B* class consists of four tandem genes located on chromosome eight of the *T. brucei* genome (LaCount *et al.*, 2003; Berriman *et al.*, 2005; El-Sayed *et al.*, 2005) while the *TbMSP-C* class consists of a single gene found at the end of a high gene density collection on chromosome 10 (LaCount *et al.*, 2003). Interestingly, *MSP-B* mRNA is found in both the blood stage and procyclic form of the parasite while both *MSP-A* and *MSP-C* mRNA is only found in the blood stage form (LaCount *et al.*, 2003).

Several genes were identified in the *T. congolense* genome as homologous to both *T. brucei* and *L. major MSP* genes. Marcoux *et al.* (2010) discovered a total of six full-length homologues and after extensive sequence analysis of each they were identified as *MSP-A*, *-B1, -B2, -C, -D* and *-E. MSP-B1* and *B2* were grouped together, because, both genes showed high sequence identity to *TbMSP-B* with low genetic distance. Genes encoding *MSP-D* and *-E* were found to have low sequence identity to all *TbMSP* gene classes and with high genetic distance were grouped on their own as separate gene classes. *TcoMSP-D, -A* and-*E*, similarly to *TbMSP-A*, all occur on chromosome 11 and cluster between 270 kb (Figure 1.5). *TcoMSP-C*, like *TbMSP-C,* occurs on chromosome 10 (Figure 1.5) within the genome while *TcoMSP-B2* and *-B1* may occur on chromosome 8 like *TbMSP-B.* However, this remains to be confirmed. Putative Major Surface Protease genes within *T. vivax* were identified and are discussed in full in Chapter 2.

In total, eight major surface protease genes have been identified in the *T. cruzi* genome (Grandgenett *et al.*, 2000) which have been separated into two groups: I and II (Cuevas *et al.*, 2003). It was also shown that within group I two members were detected; namely, a and b (Cuevas *et al.*, 2003). Perusing the phylogenetic tree constructed by Marcoux *et al.* (2010) shows that these two groups can be *MSP-B* and *MSP-C* with the likelihood that members a and b correspond to *MSP-B1* and *B2* as seen within the *T. congolense* genome (Figure 1.5). The fact that these two groups show identity with *T. brucei* and *T. congolense MSP* could mean that they occur on corresponding chromosomes within the genome. However, this also remains to be determined. There is uncertainty which class contains tandem repeats and on which chromosomes these genes reside. More information will come to light when more of the *T. cruzi* genome has been sequenced.

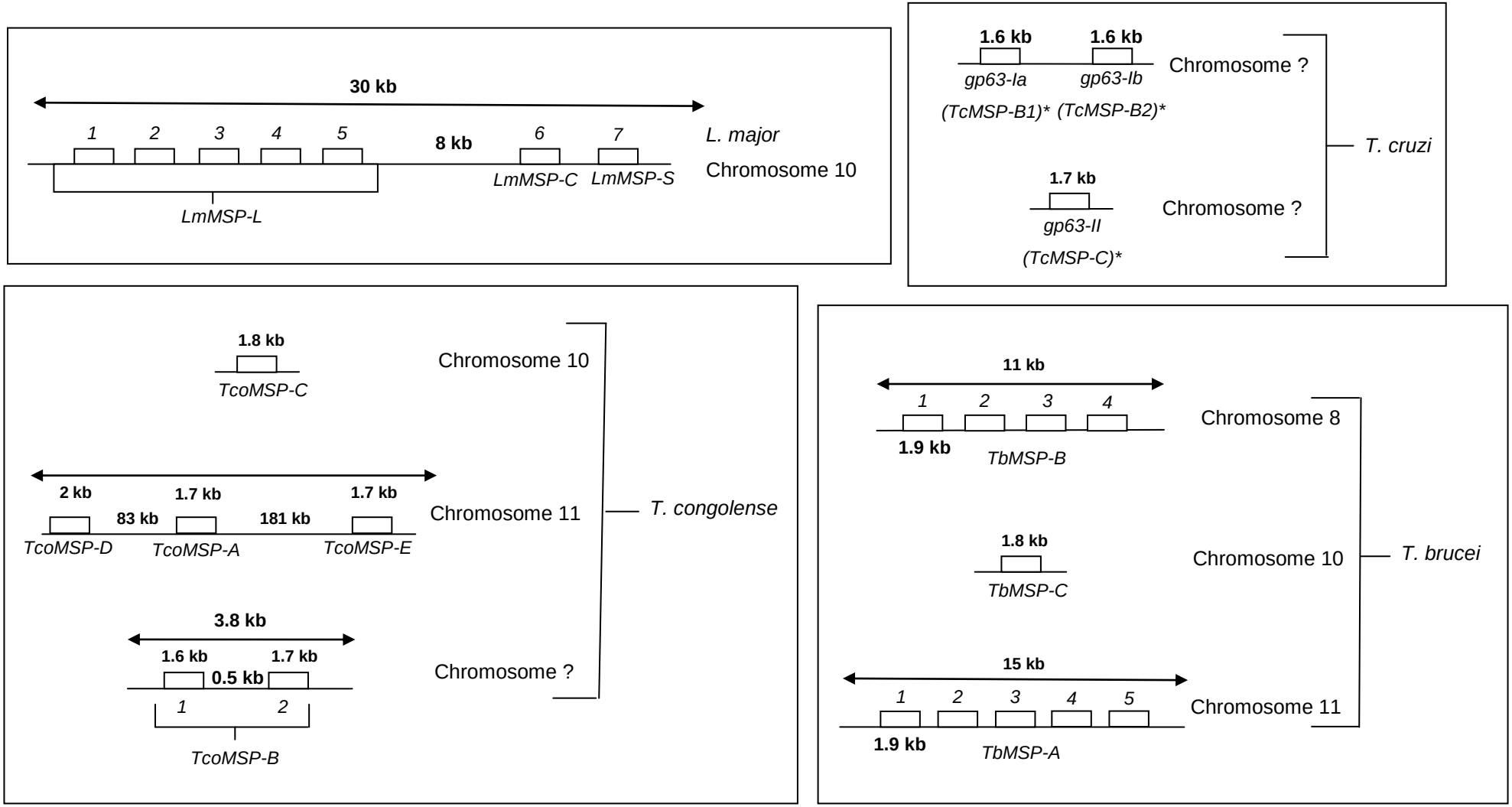


Figure 1.5: Genomic organisation of Major Surface Proteases within various Trypanosomatidae species. Arrangement of different *MSP* classes on corresponding chromosomes within the genomes of specific species is shown. ? denotes unknown information. *name according to phylogeny constructed by Marcoux *et al.* (2010).

1.7.2 Classification of Proteases

A protease is a unique enzyme because it recognises the carboxyl side of an amino acid residue regardless of the protein it belongs too (Garcia-Carreon and Del Toro, 1997). Mechanisms by which proteases cleave the peptide bonds are used to determine to which class they belong. There are seven classes of proteases identified to date: serine, cysteine, threonine, aspartic acid, metallo, glutamic acid and asparagine peptide lyases. Classification depends on residues within the active site of an enzyme and how they affect the cleavage of a protein. Some common proteases from each types are; chymotrypsin (serine), papain (cysteine), proteosome (threonine), pepsin (aspartic), thermolysin (metallo), scytalidopepsin (glutamic) and nodavirus coat protein [asparagine; (Rawlings *et al.*, 2014)]. Proteases are grouped into families because of their evolutionary relationships. As classification continues from family to subfamily followed by clan and subclan, evolutionary distance and difference between peptidases becomes less and less (Rawlings *et al.*, 2014). Classification into clans includes analysis of secondary and tertiary structures.

All serine proteases contain a catalytic serine residue in their active site with the hydroxyl group acting as the nucleophile involved in peptide cleavage. Cysteine and threonine catalytic types are the only other two groups that use an amino acid residue to act as a nucleophile rather than a water molecule. The most common catalytic mechanism of serine proteases involves a His, Asp and Ser triad which was previously known as clan SA (Rawlings and Barrett, 1993). However, after more analysis of the secondary and tertiary structure of these proteases it was found that there are a number of cysteine proteases with similar folds and cleavage mechanisms and so they were grouped in subclan PA (S) for serine and (C) for cysteine (Barrett *et al.*, 2004). There is an assortment of different catalytic sites which includes diads to tetrads of varying types of residues and these are grouped into separate clans.

Cysteine peptidases contain a catalytic cysteine residue in their active site and the sulfhydryl group acts as the nucleophile that attacks the scissile peptide bond (Erez *et al.*, 2009). The cleavage mechanism, in this case, is similar to that of serine proteases in that a nucleophile and a proton donor (general base) are required (Barrett *et al.*, 2004). Papain was the first clearly recognised cysteine protease with a catalytic tetrad: Gln, Cys, His and Asn and this forms the foundational requirements for clan CA (Rawlings and Barrett, 1993). Again, variation of different active site residues exists: diads, triads and tetrads.

The threonine protease family was first identified following the three-dimensional construction of the proteosome from yeast (Barrett *et al.*, 2004). The proteosome contains a number of different catalytic activities grouped together in one super-enzyme (DeMartino and Slaughter, 1999) and those are called N-terminal nucleophile peptidases. Three families of threonine peptidases exist (T1-3) and are grouped in subclan PB (T) with PB (C-cysteine) and (S-serine) (Barrett *et al.*, 2004). In this case, the N-terminal nucleophile is a threonine.

Aspartic proteases require a water molecule to act as the nucleophile that attacks the scissile peptide bond (Berg *et al.*, 2002). These proteases are called aspartic, not because of the amino acid residues involved in catalysis, but because Asp residues are ligands of the activated water molecule (Barrett *et al.*, 2004). Interestingly, all characterised enzymes from this group are endopeptidases. There are five clans of aspartic proteases with general ligation of the water molecule involving two Asp residues (Barrett *et al.*, 2004). Clans AC and AD have no proteases where the crystal structure has been derived and so classification for these may change as more understanding of their structure becomes known.

Glutamic proteases have been reclassified as a sixth catalytic type from the A4 group of aspartic proteases (Sims *et al.*, 2004) and consists of the families G1 and G2 (Rawlings *et al.*, 2014). Scytalidoglutamic peptidase from *Scytalidium lignicolum* is the founding member of the G1 family (Pillai *et al.*, 2007). These proteases have a unique Glu-Gln catalytic dyad in which the glutamic acid acts to activate the nucleophilic water molecule while the glutamine acts to stabilise the tetrahedral intermediate (Sims *et al.*, 2004; Pillai *et al.*, 2007; Jensen *et al.*, 2010). These proteases also contain a novel β-sandwich domain. Glutamic proteases have been identified in fungi, bacteria, viruses and animal species and over 350 putative protease sequences (Rawlings *et al.*, 2014).

In 2011, a seventh protease family was identified: the asparagine peptide lyases (Rawlings *et al.* 2011). In this family five clans have been identified: NA-E consisting of a total of eight subclans (N1-8) (Rawlings *et al.*, 2011; Rawlings *et al.*, 2014). Clan NA is the largest consisting of over 900 putative sequences within bacteria and viruses only. There are putative sequences within plants and animals for these lyases as well (Rawlings *et al.*, 2014). These proteases all contain an Asn within their catalytic site which acts as the nucleophile during catalytic processing (Rawlings *et al.*, 2011). This shows that not all proteolytic enzymes are peptidases.

Metalloproteases are one of the largest groups of proteases containing about 50 families (Rawlings *et al.*, 2014). Like aspartic proteases, they use a water molecule to act as a nucleophile for the hydrolysis of the peptide bond (Barrett *et al.*, 2004). They differ from all other proteases in that they require a divalent metal cation as a cofactor to activate the water

molecule and perform cleavage (Barrett *et al.*, 2004). The most common cofactor is zinc while other protease families use other cofactors such as copper, magnesium or nickel. The zinc dependant proteases have the ability to use other cations in zinc deficient environments (Holmquist and Vallee, 1974). The metal ions are held in place by amino acid ligands. A significant amount of research has been done to further class the large number of various zinc metalloproteases known using techniques that involve analysis of primary to tertiary structure, mechanism of peptide substrate cleavage and zinc binding site residues (Hooper, 1994; Stocker and Bode, 1995; Stocker *et al.*, 1995; Rawlings *et al.*, 2014). The general zinc binding site motif is HEXXH (where X is any amino acid) and all proteases that contain this site are grouped together in the class called the zincins (Bode *et al.*, 1993) or clan MA (Barrett *et al.*, 2004). Proteases that do not contain the HEXXH motif are grouped in clans MC-H and MJ-N. Examples of variations include HXXEH (inverzincins), HXXE and HXH.

The zincin class is then further split into two super-families: gluzincins or subclan MA(E) (Hooper, 1994; Barrett *et al.*, 2004) and metzincins (Bode *et al.*, 1993) or subclan MA(M) (Barrett *et al.*, 2004). Proteases are allocated a family according to the third zinc ligand consensus sequence which is conserved in each case. Gluzincins, named because of their fifth zinc ligand being a glutamic acid residue, consists of the thermolysin, endopeptidase 24-11, angiotensin converting enzyme and aminopeptidase families (Hooper, 1994). The other superfamily is the metzincins, so called due to the presence of a much longer zinc binding consensus sequence HEXXHXXGXXH and a conserved methionine-containing turn [Met-turn; Bode *et al.* (1993); Hooper (1994); Barrett *et al.* (2004)]. Four main families that belong to this superfamily are the astacin, serratia, reprolysin and matrixins, which are further divided into sub-families M6-8, M10-12, M35, M43, M46, M57, M64 and M72. In this case, proteases are separated into corresponding families by determining which residue follows the third histidine zinc binding ligand and the residues surrounding the methionine in the Met-turn. This Met-turn is present in all metzincins and so must be important for the functioning of the enzymes (Tallant *et al.*, 2010).

Astacin, a digestive enzyme found in crayfish (Hopsu-Havu *et al.*, 1997) is the exemplar enzyme for the astacin family. Enzymes that belong to this group come from diverse sources including matrix metalloendopeptidase and developmental regulation proteases (Bond and Beynon, 1995). They have a conserved glutamic acid following the third histidine (HEXXHXXGXXH**E**). Zinc ligation involves the three histidines within the zinc binding motif, the fourth being water and uniquely a fifth tyrosine residue within the Met-turn which results in a pentacoordinated trigonal-bipyramidal geometry of zinc binding (Stocker *et al.*, 1993).

The serralysin family consists of proteases that are secreted into their environment by various pathogenic bacteria of the genera *Serratia* (Nakahama *et al.*, 1986), *Pseudomonas* (Duong *et al.*, 1992) and *Erwinia* (Dahler *et al.*, 1990). A conserved proline follows the third histidine (HEXXHXXGXXH**P**) within the longer zinc binding consensus sequence (Mock and Yao, 1997) and as in the astacin family, a tyrosine within the Met-turn acts as the fifth zinc binding ligand (Bond and Beynon, 1995).

The reprolysin family, which consists of a number of snake venom toxins (Bjarnason and Fox, 1995), has a conserved aspartic acid residue following the third histidine within the zinc binding motif (HEXXHXXGXXH**D**). Unlike astacin and serralysin protease families, reprolysin proteases lack a fifth zinc binding ligand which results in a tetrahedral geometry of zinc binding (Cirilli *et al.*, 1997; Gomis-Rüth, 2003).

Lastly, the matrixin family which mainly consists of mammalian matrix metalloprotease (Visse and Nagase, 2003; Nagase *et al.*, 2006) has a conserved serine following the zinc binding motif (HEXXHXXGXXH**S**). Like reprolysin, no tyrosine residue is present in the Met-turn resulting in a similar zinc binding topology.

Of all the Major Surface Proteases discovered only leishmanolysin from *Leishmania major* has been classed into a metalloprotease clan and family. Since it contains the HEXXH motif it belongs in the zincin or MA clan under the M8 peptidase family (Barrett *et al.*, 2004). Three-dimensional analysis showed that it contains the conserved Met-turn and hence belongs to the subclan, metzincin [(MA(M)); (Schlagenhauf *et al.*, 1998)], but determination as to which family it belongs or if it needs to be grouped into a family of its own is still uncertain. Further discussion on this aspect will be found in Section 1.7.3.3.

1.7.3 Structure of MSPs

None of the trypanosomal Major Surface Proteases have been crystallised but molecular modelling, using various bioinformatics programs, has been done. The structure of one of the *Leishmania* Major Surface Proteases, leishmanolysin, has been successfully determined. However, primary and secondary structure for the various proteases will be analysed first.

1.7.3.1 Primary Structure of MSPs

The amino acid sequences for all *T. brucei, T. congolense,* and *T. cruzi* MSPs was aligned with the amino acid sequence for leishmanolysin, using BioEdit, for the purpose of this discussion (Appendix 1, Figure A1).

A basic trend observed in sequence identity is that each class of protease shares highest identity; of example, *Tb*MSP-B shares approximately 55% identity with *Tco*MSP-B and 44% with *Tc*MSP-B (Appendix 1, Table A1). Also, in general, higher sequence identity is observed between proteins from *T. brucei* and *T. congolense* compared to those from *T. cruzi* which is expected since *T. cruzi* is a South American *Trypanosoma* species. Leishmanolysin gave low sequence identities ranging between 20-30%. Highest identity for leishmanolysin was seen between the MSP-B protein sequences of all species. Interestingly, 30% identity was seen between leishmanolysin and *Tco*MSP-B2 (Appendix 1, Table A1).The sequence identity also reinforces the deduction that *Tc*gp63-Ia, -Ib and II indeed belong to the family MSP-B1, B2 and C, respectively (Appendix 1, Table A1).

From the full alignment (Appendix 1, Figure A1), conserved regions were schematically mapped as seen in Figure 1.6. In total, 18 conserved cysteine residues and 12 semi-conserved proline residues (11 or more of the sequences contain a proline at that position) are present which means that the secondary and tertiary structure of these proteins is likely to be conserved. The conservation of important amino acid residues throughout different species shows that the basic function of the protein has not changed. This conservation again suggests these proteins are of importance within the various parasites (LaCount *et al.*, 2003). It is observed that all sequences contain a truncated zinc binding site HEXXHXXGF instead of the full HEXXHXXGXXH motif found in other metzincin metalloproteases (Figure 1.6). The truncated form is due to the approximately 40-60 amino acid insert present between the glycine and the 'third' histidine (Figure 1.6). The C-terminal domain contains few conserved regions (Figure 1.6) which corresponds to literature that shows that different classes of Major Surface Protease are either GPI-anchored or not (Cuevas *et al.*, 2003; LaCount *et al.*, 2003; Grandgenett *et al.*, 2007; Marcoux *et al.*, 2010).

Evaluation of the number of amino acid residues in each sequence showed that on average the MSP protein is ~595 amino acid residues (65 kDa). If the different classes of MSPs of *Trypanosoma* are compared, the sizes MSP-A, -C, -D and -E were similar which correspond with sequence identities and may be related to where they are found within parasites (Appendix 1 and 2). Greatest variation in size is observed between members of the MSP-B family (Appendix 2, Figure A2.1).

The analysed sequences included the propeptide of the protein (Figure 1.6). It has been shown that in leishmanolysin a cysteine residue bound to the zinc cofactor is expected to block the active site resulting in protease inhibition (Macdonald *et al.*, 1995).

Interestingly, catalytically active leishmanolysin is unable to cleave the propeptide of inactive enzyme (McGwire and Chang, 1996). Organomercurial compounds can be used to break

this cysteine zinc complex which then allows the enzyme to autocatalytically remove its propeptide. This has been shown to occur, within leishmanolysin, between two valine residues downstream from the cysteine (Button and McMaster, 1988). There is a conserved cysteine in all other MSPs studied (not present in *Tco*MSP-B1) which may indicate that the propeptide inhibits protein functioning in a similar manner to leishmanolysin (Figure 1.6). All of *T. cruzi's* MSPs, *Tb*MSP-A and *Tco*MSP-E contain a valine dyad downstream from the conserved cysteine (Figure 1.6). These five MSPs homologues may remove their peptide in a similar manner to leishmanolysin but since a large number do not contain this dyad, different amino acid residue cleavages may exist too.

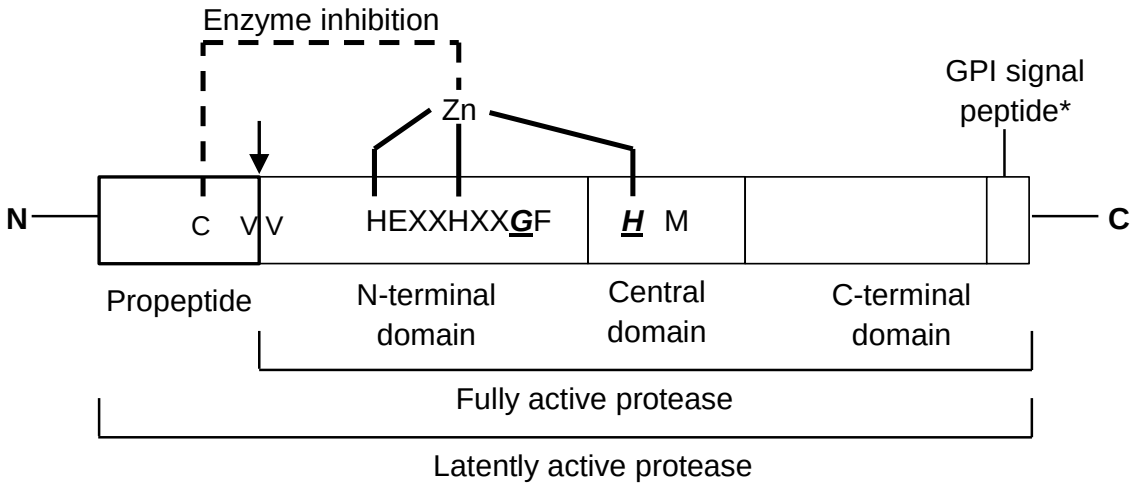


Figure 1.6: Schematic representation of conserved regions within MSPs. C, cysteine residue suspected to be involved in peptide inhibition; VV - valine-valine dyad where propeptide is removed (arrow); HEXXHXXGF, conserved active site motif; *H*, third histidine ligand; M, Met-turn methionine. From *G* (Gly) to *H* (His) a 40-60 amino acid insert is present. The solid line shows the three His (H) ligands bound to the zinc (Zn) cofactor while the dotted line shows interaction of Cys with Zn for protease inhibition. N- and C-termini are indicated. GPI-anchor signal peptide* is only present in MSP-A, -B and -E (as predicted).

1.7.3.2 Secondary Structure of MSPs

Various prediction programmes are available to predict the secondary structure of proteins. Since the structures of all *Trypanosoma* MSPs are unknown, we used the Quick2D bioinformatics tool (http://toolkit.tuebingen.mpg.de/quick2_d) to predict the secondary structures discussed here.

For simplicity, only *Tco*MSP-B2 and *Tc*MSP-B2 (gp63-Ib) were used in secondary structure analysis. When all MSP outputs were examined, it was observed that all these proteases have two main helical structures of which the second contains the HEXXHXXGF motif

following, 50-90 residues later, the 1,4 Met-turn and a third relatively prominent helix (examples in Appendix 3). These helices may correspond to helix A, B and C, respectively, which are seen in a number of gluzincins and metzincins like thermolysin (Matthews *et al.*, 1972), astacin (Gomis-Rüth *et al.*, 1993) and leishmanolysin (Schlagenhauf *et al.*, 1998). The C-terminal domain is dominated by β-sheets with few helices which is also seen in leishmanolysin (Schlagenhauf *et al.*, 1998). Interestingly, *Tco*MSP-D does not contain the 1.4 Met-turn methionine that is conserved in all the other MSPs and no methionine is in that vicinity either (Appendix 1).

The Quick2D analysis also showed that all the MSPs contain a signal peptide and signal peptide cleavage site which is indicative of being exported from the cell (Kapp *et al.*, 2009). An additional prediction program, PredGPI (http://gpcr2.biocomp.unibo.it/gpipe/pred.htm), was used to determine the likelihood that the MSP classes have a glycosyl-phosphatidylinositol (GPI) anchor or not (Appendix 4). It was found that all MSP-As and MSP-Bs are highly likely to have a GPI-anchor. All MSP-Cs and MSP-Ds are highly unlikely to have a GPI-anchor. On the other hand, the MSP-E class has a probable chance of having a GPI-anchor like classes MSP-A and -B. Taking into account both the presence of a signal peptide and a GPI-anchor and if prediction corresponds to biological function, it would seem that classes MSP-A, MSP-B and MSP-E are all located via a GPI-anchor to the surface of the parasite while MSP-C and MSP-D families are either secreted by the parasite or anchored to the surface in an alternative manner. This conclusion corresponds to a study done by Grandgenett *et al.* (2007) who showed that VSG cleavage and coat shedding in *T. brucei* is performed by a zinc metalloprotease on the surface of the parasite. Furthermore, the recent analysis of the *T. brucei* cell surface proteome places both MSP-A and MSP-B on the surface of the parasite, confirming these predictions (Shimogawa *et al.*, 2015). It has been shown that GPI-anchored proteins can act as activator antigens within the immune system which may suggest that these proteases cause immune responses in their host and may be a good diagnostic target (Brown and Waneck, 1992).

The sequences were also subjected to predictions on whether transmembrane domains were present using DAS-TMfilter (http://mendel.imp.ac.at/sat/DAS/DAS.html). Generally, GPI-anchored proteins do not contain transmembrane domains but are rather synthesised with a transmembrane tail which is removed after attachment to the GPI-anchor (Brown and Waneck, 1992; Ferguson and Hart, 2009).

It was found that the MSP-D family and *Tb*MSP-C are unlikely to have a transmembrane domain (Cserzo *et al.*, 2002, 2004) and unlikely to be GPI-anchored (Appendix 4) and so could be secreted by the parasite. *Tco*MSP-C and *Tc*MSP-C are both unlikely to have GPI-anchors but likely to have transmembrane domains (Appendix 4).

By comparison, leishmanolysin has a predominant β-sheet secondary structure (Figure 1.7) and the C-terminus contains the GPI-anchor used for surface attachment via an asparagine residue (Schlagenhauf *et al.*, 1998). In total; there are nine disulfide bridges which help in the formation and stabilisation of the tertiary structure. There are three major helices named A, B, and C as well as the tight 1,4 Met-turn, represented in all metzincin metalloproteases (Gomis-Rüth, 2009). It is also observed that leishmanolysin has a number of extra amino acids between helices and disulfide bridges that result in a number of flaps which are not seen in other zincin proteases (Schlagenhauf *et al.*, 1998).

Overall, the MSP proteins seem to contain a number of secondary structure similarities to leishmanolysin as well as other metzincins so the tertiary structure and mechanism of cleavage employed by these metalloproteases are likely to be similar to those of leishmanolysin.

1.7.3.3 Tertiary Structure of MSPs

Since high similarity was observed between the MSPs of *Trypanosoma* species and that of *Leishmania major*, an analysis of the tertiary structure of leishmanolysin, solved by Schlagenhauf *et al.* (1998), was done highlighting both similarities and differences it has with other zincin metalloproteases.

Other than the short HEXXH zinc binding motif, leishmanolysin shows little sequence identity with other metalloproteases, other than the homologs occurring in *Trypanosoma* (El-Sayed and Donelson, 1997a; Grandgenett *et al.*, 2000; Marcoux *et al.*, 2010), *Crithidia* and *Herpetomonas* (Etges, 1992). After crystallisation of leishmanolysin by Schlagenhauf *et al.* (1998), insight into the mechanism of cleavage and similarities to other zinc metalloendoproteinases was possible. The overall structure consists of three domains: N-terminal, central and C-terminal (Figure 1.7). The N-terminal domain consists of helices A and B (Figure 1.7) found on one side of a five-stranded β-sheet complex, which is also present in both gluzincin and metzincin families (Matthews *et al.*, 1972; Gomis-Rüth *et al.*, 1993; Gomis-Rüth *et al.*, 1994). The HEXXH motif is located within helix B of the N-terminal domain where the two histidines bind the zinc cofactor (Figure 1.7). The predicted secondary structure of other MSPs showed that the second prominent helix contained the zinc ligation

motif (Appendix 3). Leishmanolysin also has two 35 and 40 residue flaps which cover the opposite side of the β-sheet.

The central domain consists of compact anti-parallel helices and an anti-parallel β-sheet. In other metzincins, the glycine within the full HEXXHXX**G**XXH motif causes a tight turn at the end of the active site and results in the third histidine being in position for zinc ligand binding (Stocker *et al.*, 1995). However, as shown in Figure 1.6, a truncated form of the consensus sequence is present in the MSPs which results in the third histidine ligand in leishmanolysin being about 60 residues further away compared to other metzincins (Figure 1.7). The truncated form of the metzincin zinc binding motif results in an active site geometry which is different from other metzincins known to date.

Schlagenhauf *et al.* (1998) results show that the conserved histidine is still the third zinc ligand and that leishmanolysin has a pentacoordination of zinc binding similar to the astacin family. However, instead of the water molecule and tyrosine within the Met-turn being the fourth and fifth ligand respectively, a different setup is observed. After electron density studies on leishmanolysin and comparison to astacin and matrix metalloproteases, it was found that a glycine's amino nitrogen is in a similar position to the catalytic water of these enzymes while the glycine's carboxyl oxygen is orientated like the phenolic oxygen of tyrosine in astacin. So overall, three histidines represent the first three zinc ligands while a glycine molecule's amino nitrogen and carboxyl oxygen form the fourth and fifth ligands, respectively. A similar conformation of N-O zinc ligation has been found in stromelysin-1 matrix metalloprotease (Gomis-Ruth, 1997). Moreover, all the MSPs have a conserved glycine residue C-terminal from the active site motif (Appendix 1, Figure A1)

The Met-turn and helix C is present in the central domain of leishmanolysin (Figure 1.7). The function of the Met-turn within leishmanolysin (since it is unlikely to be used for zinc binding) and whether these proteases actually need it is uncertain. Since MSP-D of trypanosome species do not contain the Met-turn, suggests that it is not essential.

The C-terminal domain consists mostly of β-sheets and minimal helices are present which is comparable to the predicted secondary structure of other MSPs (Appendix 3). It also contains six of the nine disulfide bridges present within the entire molecule (Figure 1.7). From its N-terminus to its C-terminus the structures are as follows: two-stranded β-sheets, two short helices and a six stranded anti-parallel β-sheet sandwich (Figure 1.7) followed by two short helices which are attached to a GPI-anchor on the surface of the parasite.

Since the truncated form of the metzincin motif is present in all trypanosoma MSPs it is probable that similar residues and pentacoordinate geometry of zinc ligation may be used in the cleavage of peptide substrates.

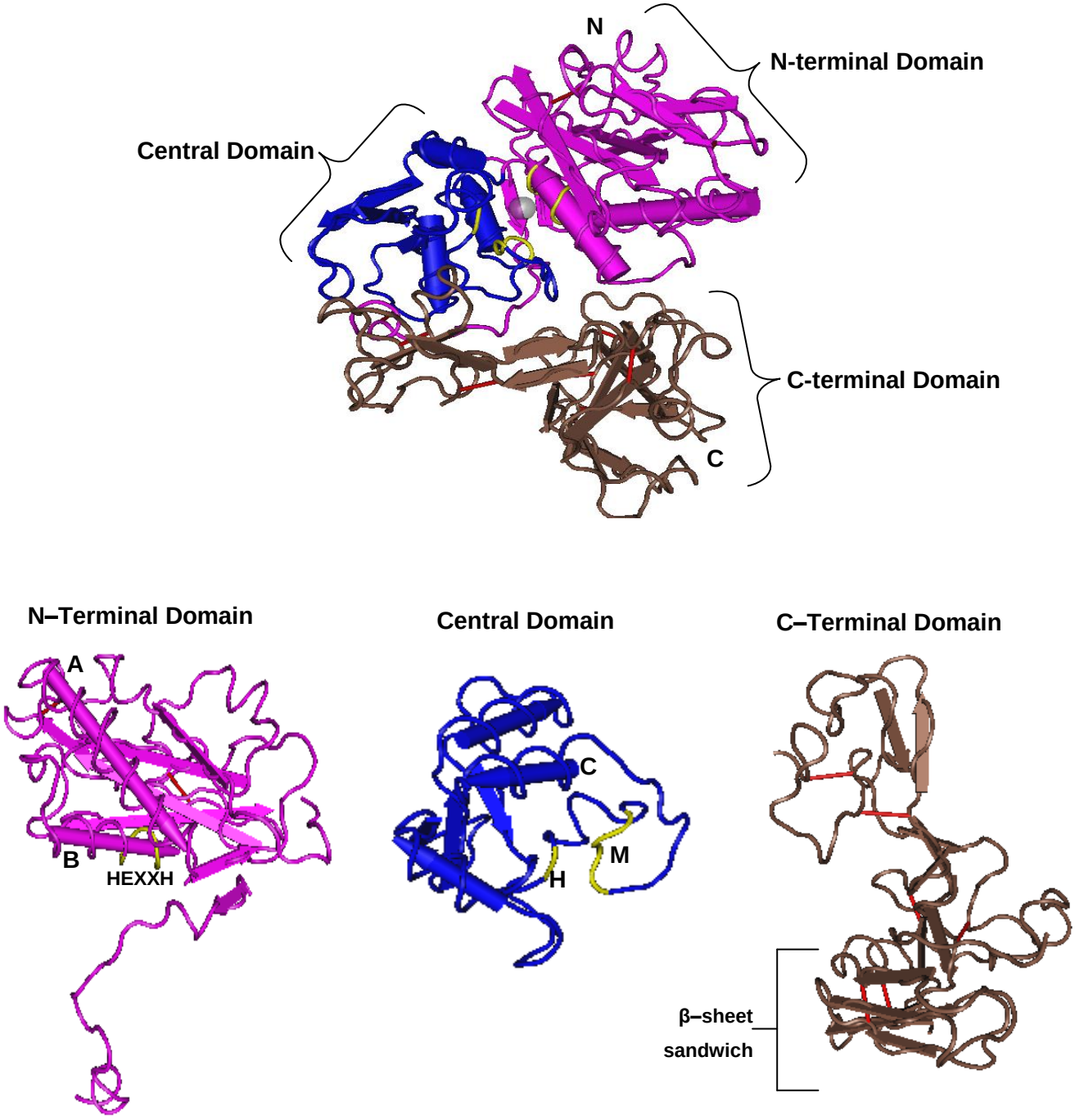


Figure 1.7: Crystal structure of leishmanolysin. The published structure was edited using Cn3D. The N-terminal domain is shown in purple, central domain in blue and C-terminal domain in brown. The grey sphere represents zinc binding and red lines represent the disulfide bonds. N- and C-termini are labelled. Yellow represents the HEXXH motif, third histidine ligand (H) and the Met-turn (M). Helix A, B and C are labelled. The six-stranded anti-parallel β-sheet sandwich is also labelled. Adapted from Schlagenhauf *et al.* (1998).

1.7.4 Mechanism of Substrate Cleavage

With knowledge of how the zinc cofactor is ligated and the different use of a nucleophile within leishmanolysin, the cleavage of the peptide bond by MSPs can be considered. Thermolysin is a common zinc metalloprotease and its mechanism of cleavage is illustrated here (Figure 1.8).

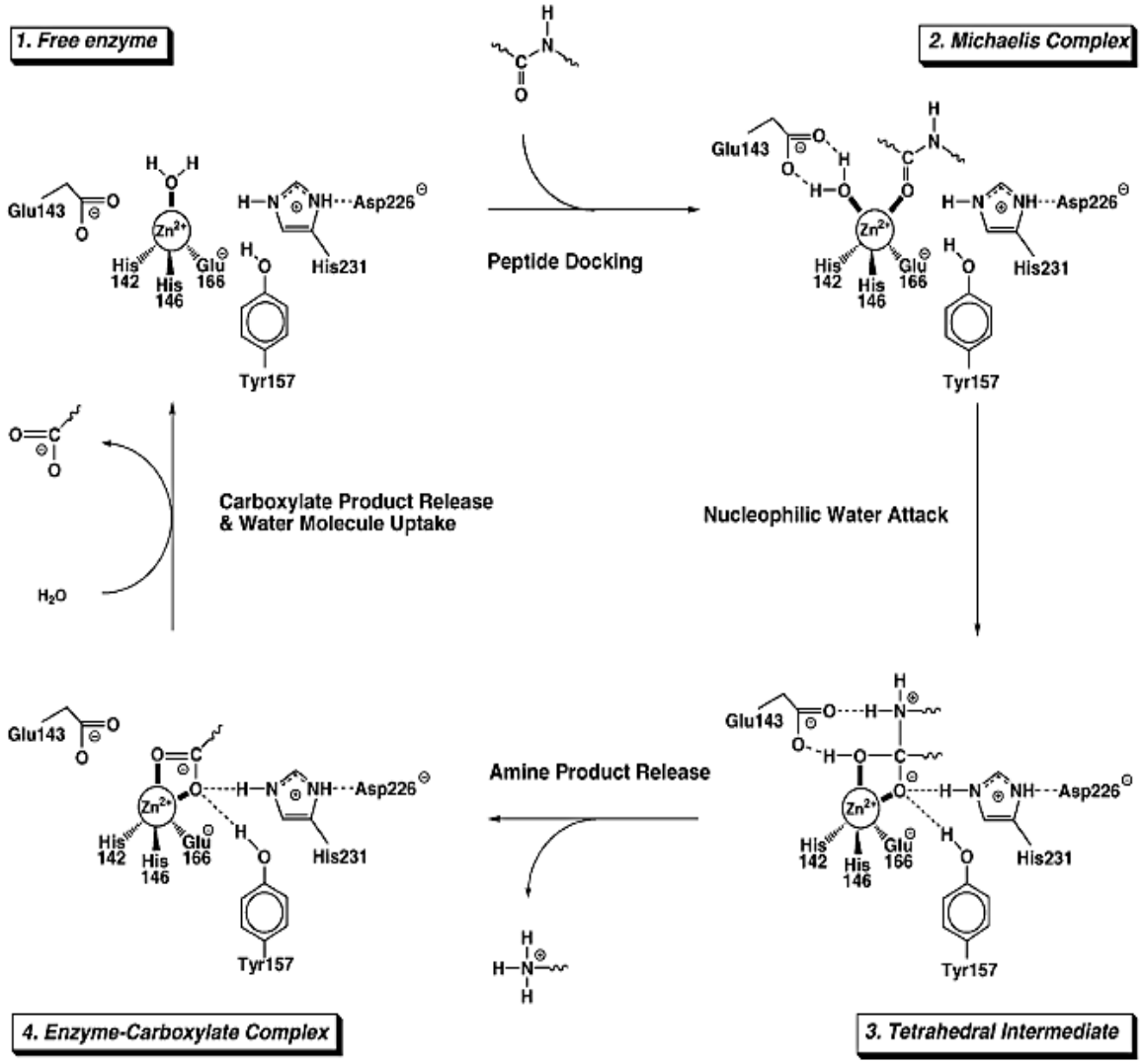


Figure 1.8 General mechanism of catalytic cleavage by thermolysin zinc metalloprotease (Pelmenschikov *et al.*, 2002).

The active site of the free thermolysin enzyme contains the zinc cofactor bound by three amino acid ligands. A water molecule is also associated with the zinc cofactor and a glutamate residue is the near vicinity (1. Free enzyme; Figure 1.8). On peptide entry, the carbonyl oxygen polarises with the zinc atom causing the water molecule to associate with the glutamate (2. Michaelis complex; Figure 1.8).

This complex then undergoes nucleophilic attack by a water molecule which results in the amino nitrogen of the peptide bond associating with glutamate (3. Tetrahedral complex; Figure 1.8). The tetrahedral complex is stabilised by a tyrosine and a histidine held in position by an asparagine. The water molecule donates its hydrogen to the amino nitrogen causing the peptide bond to break and the release of the amine product. The glutamate is also freed. The carboxylate region of the peptide remains attached to the zinc atom (4. Enzyme carboxylate complex; Figure 1.8). Addition of a water molecule causes the release of the carboxylate product and regeneration of the active site.

This is the general mechanism of cleavage and there are a number of variations. As described before, leishmanolysin is unlikely to use a water and glutamate but rather the amino nitrogen and carbonyl oxygen from a glycine in its cleavage (Schlagenhauf *et al.*, 1998). This is also likely for other MSPs.

The general substrate specificity for leishmanolysin is -Ala-Tyr↓Leu-Lys-Lys- (Bouvier *et al.*, 1990) from P2-P3′ (Schechter and Berger, 1967). Studies have shown that P2 can contain a Leu as well (Schomburg and Stephan, 2012). Since tyrosine contains a benzene ring as well as a hydroxyl group in the R-chain, the S1 site of leishmanolysin must be large to accommodate this. The S2’ and S3′ sites have been shown to favour the lysine dyad and, therefore, must be able to accommodate their long side chains (Bouvier *et al.*, 1993).

1.7.5 Function of Major Surface Proteases

Proteases within parasitic protozoa have been identified to play an important role in these unicellular organisms. Identifying different proteases and understanding the way they function within the life cycle of parasites have helped to understand how the parasite functions, as well as giving insight to vaccine development and diagnostics especially, in cases where these organisms infect and cause disease (Klemba and Goldberg, 2002).

Isoforms of leishmanolysin are expressed in both the promastigote and amastigote stages of the life cycle of the parasite and is likely to have different functions within different stages (Frommel *et al.*, 1990). One of the most important functions of leishmanolysin is that it is able to bind to complement component C3 which the parasite encounters on injection by the sand fly vector into the mammalian host (Russell, 1987). Leishmanian parasites expressing wild type leishmanolysin on their surface act as primary acceptors of C3b and are involved in its conversion into a molecule that contains iC3b. It then uses this to bind Mac-1 receptors for cell invasion enabling the parasites to use complement factors without being affected by their lytic properties (Brittingham *et al.*, 1995). This aids the parasites’ engulfment by

macrophages where they reside within the mammalian host (Russell, 1987; Wilson and Hardin, 1988). More recently, it has been shown that leishmanolysin is able to inhibit the recruitment of LC3 to phagosomes via the down-modulation of VAMP8 (endocytic SNARE protein associated with phagolysosome biogenesis and function), protecting the parasite from the antimicrobial activity of the host (Matte *et al.*, 2016). Furthermore, leishmanolysin has been identified as a virulence factor for the parasite where it acts, in conjunction with cysteine peptidase B, to regulate the trafficking of VAMP3 and VAMP8, two endocytic SNARE proteins (Casgrain *et al.,* 2016). Cysteine peptidase B mutant parasites, lack expression of leishmanolysin and the ability to regulate the two SNARE proteins (Casgrain *et al.*, 2016).

Other roles that have been suggested for this protease are that they are able to protect liposomal contents of the parasite from degradation (Chaudhuri and Chang, 1988) and may have an important function in avoiding antimicrobial peptides within the host and vector (Kulkarni *et al.*, 2006).

The discovery of MSP genes within African trypanosomes was a surprise since these parasites already have an efficient and effective way of evading the immune system, i.e. antigenic variation of their VSG type, which enables different generations of parasite to express different surface proteins and therefore evade the immune system (Pays *et al.*, 2004). Also, African trypanosomes do not contain an intracellular stage of their life cycle, like *Leishmania* does, and so the function of MSPs to aid parasite entry into macrophages or other host cells does not apply either (Vickerman *et al.*, 1988). It was concluded that either these proteases have novel functions in trypanosomes which do not apply in leishmanian species or they have functions that have not yet been discovered in *Leishmania*.

Of the small amount of research done on MSPs in *T. brucei,* it has been shown that *Tb*MSP-A and -C are only expressed in blood stage parasites while *Tb*MSP-B is expressed in both blood stage and procyclic forms (El-Sayed and Donelson, 1997a; LaCount *et al.*, 2003; Grandgenett *et al.*, 2007). Expression of MSP-B in the procyclic form is higher compared to in the blood stage form. For this reason, LaCount *et al.* (2003) decided to do an in depth study on *Tb*MSP-B. Their study showed that knockout of the *Tb*MSP-B coding gene had no effect on growth of parasites at either the procyclic or blood stream stage of development and is, therefore, not essential for parasite survival (LaCount *et al.*, 2003). In 1997, Bang *et al*. showed that VSGs are exported to the parasite surface with a phosphatidylinositol-specific phospholipase C-resistant GPI-anchor and that these VSGs are released into the culture medium during parasite differentiation into the procyclic form.

This release of VSG is dependent on a zinc metalloprotease since addition of 1,10-phenanthroline, a common metalloprotease inhibitor, stops this process of protein release (Bangs *et al.*, 1997). It was found later that *Tb*MSP-B functions in this coat shedding by parasites since RNAi knockout of the gene inhibits release of VSG during parasite differentiation (LaCount *et al.*, 2003). *Tb*MSP-B may also be used to ensure that VSG and other unwanted procyclin proteins do not accumulate on the surface of the parasites. This ability of *T. brucei* to rapidly shed its VSG coat has now been shown to be due to both MSP-B and GPI-PLC activity (Grandgenett *et al.*, 2007). Since *Tb*MSP-A and -C are only expressed in the blood stage form of the parasite, these proteases may have life cycle stage specific functions.

After the discovery of homologues for MSP in *T. congolense*, a study was conducted to determine the functioning of *Tco*MSP-D. *Tco*MSP-D was studied since conservation with *T. cruzi* and *Leishmania* MSP was considerable and it is likely to be secreted (Marcoux *et al.*, 2010). After recombinant expression of the amino terminus of the protein, mice were immunised and then subjected to parasitic infection. It was found that mice immunised against *Tco*MSP-D became more susceptible to parasitic infection and, therefore, *Tco*MSP-D is in fact a virulence factor (Marcoux *et al.*, 2010).

A very interesting study was done by Bangs *et al.* (2001) using existing metzincin metalloprotease inhibitors, known to inhibit *Leishmania* MSP, on *in vitro* growing *T. brucei* parasites. Excitingly, four inhibitors were found to reduce proliferation and differentiation of blood stage form of the parasite and it was concluded that metalloprotease have an important function in the life cycle of trypanosomes. Therefore, this may open a new avenue for drug development and therapy (Bangs *et al.*, 2001).

T. cruzi MSPs are expressed at different life-cycle stages, specifically most abundantly in amastigote parasites (Cuevas *et al.*, 2003; Kulkarni *et al.*, 2009). Since *T. cruzi* has an intracellular life cycle stage like *Leishmania*, it is suspected that its MSP proteases function in host cell invasion (Cuevas *et al.*, 2003; Kulkarni *et al.*, 2009). Also, like *T. brucei, T. cruzi* is required to shed its VSG coat during cell differentiation and so its MSP proteases may function here as well.

It has also been shown that a recombinant form of *T. carassii* gp63 is able to significantly reduce expression of inducible nitric oxide synthase, TNFα-1 and -2 within macrophages and internalisation of recombinant *Tca*gp63 was observed by goldfish macrophages (Oladiran and Belosevic, 2012). This MSP is grouped together with *T. brucei* and *T. cruzi* MSPs in phylogenetic analysis.

Considering that it has been shown that leishmanolysin is able to protect the parasites from antimicrobial peptides found within its mammalian host (Kulkarni *et al.*, 2006), it is possible that its trypanosome homologues could have the same function. Leishmanolysin does this by rapidly cleaving these peptides so they become ineffective (Kulkarni *et al.*, 2006). Antimicrobial peptides have been designed to avoid MSP cleavage which has been successfully shown to have anti-leishmanian properties (Kulkarni *et al.*, 2014). This could then be applied to *Trypanosoma* species as an anti-parasitic treatment since African trypanosomes have already been shown to be susceptible to antimicrobial peptides *in vitro* (McGwire *et al.*, 2003b).

Also, a zinc metalloprotease was identified in crude lysate from *T. brucei* blood stream form parasites (de Sousa *et al.*, 2010). Although, it is uncertain whether this protease is part of the MSP family or not, it was suggested that it may play a role in crossing the blood brain barrier and invading the central nervous system since it was able to degrade collagen, fibronectin and laminin, all of which are important vascular matrix components (de Sousa *et al.*, 2010). Could it then be possible that MSP plays a similar role?

Clearly there are still a number of questions involving the function of these proteases that need to be answered. Also, they have been identified as virulence factors and, hence, important but whether they open up new avenues for diagnostics and treatment of trypanosomiasis is still uncertain. Knowledge of their biochemical properties could further this investigation.

1.8 Objectives of the Study

Major Surface Proteases (MSPs) have been identified in the African trypanosomes, *T. brucei* and *T. congolense* but not in *T. vivax*. The first aim of the study was to analyse protein databases to identify putative *T. vivax* MSPs and attempt to classify the proteins found. Additional bioinfomatic analyses were also done and are presented in Chapter 2.

Once MSPs were identified in all African trypanosomes, each protease was analysed to determine conserved and nonconserved regions. The second aim was therefore to use these regions to produce anti-peptide antibodies with differing detection abilities. Three peptides were selected with the intention to 1) detect *Tb*MSP-B and *Tco*MSP-B (*Tb/Tco*MSP:303-314), 2) detect *Tb*MSP-B and *Tb*MSP-C (*Tb*MSP:400-412) and 3) detect *T. vivax (Tv)* MSP-C (*Tv*MSP:686-697). They were used to generate two-types of detection molecules: full IgY antibodies and single chain variable fragments (scFvs). This would give

insight into the possibly of these proteases as diagnostic markers and these results are reported in Chapter 3.

Of all the MSP classes, MSP-C was of most interest because it is found in all African trypanosome species, is expressed in the blood stage form of the parasite and is highly unlikely to be GPI-anchored to the surface of the parasite. It, therefore, may be a secreted protease. Since these trypanosomes are extracellular, this protease would be secreted directly into the host blood stream. This suggests that they could be detected in a diagnostic test or may have adverse effect on the host. The last aim was to clone, express and characterise the protease under non-denaturing condition. The ability of MSP-C, from *T. brucei* and *T. congolense,* to hydrolyse the fluorescent substrate H-Suc-Leu-Tyr-AMC, the effect of this hydrolysis in the presences of metalloprotease inhibitors and the pH optima of the enzymes were also assessed. Furthermore, the detection of recombinant protease by infected cattle sera was also determined and these results are reported in Chapter 4.

Chapter 2

Major Surface Proteases in *Trypanosoma vivax*: A Bioinformatics Analysis

2.1 Introduction

Numerous Major Surface Proteases (MSPs) have been identified in a number of trypanosomal and leishmanial species including the African trypanosomes, *T. brucei* and *T. congolense,* whereas little has been reported on the *T. vivax* MSPs*.* Putative MSP gene and protein sequences for *T. vivax* exist on a number of databases; however, these sequences have not been organised into the MSP classes A to E. Therefore, before any study can be undertaken on the *T. vivax* proteases, putative gene and protein sequences need to be analysed in comparison with other MSP sequences for their classification as MSP-A, B, C, D or E. The National Centre for Biotechnology Information (NCBI-www.ncbi.nlm.nih.gov/) has the largest collection of sequence information in a single site and it was used to search for putative *T. vivax* MSPs.

When the metabolic requirement of a gene product is in high demand, gene duplication, which can result in multiple copies of a gene, can fulfil this requirement (Ohno, 2013). It is known that *TbMSP-B* has four tandemly repeated genes while there are two classes of *TcoMSP-B* (LaCount *et al.*, 2003; Marcoux *et al.*, 2010). *TbMSP-A* has five tandemly repeated genes (LaCount *et al.*, 2003) which suggests that these proteases are required in significantly large amounts by these parasites. However, posttranslational regulation also affects protein concentration within a particular cell and it is not solely based on mRNA levels [reviewed in Brockmann *et al.* (2007)].

Determining the conserved regions within an enzyme's primary structure can suggest similarities in function and substrate specificity. Amino acid residues such as cysteine and proline indicate conserved secondary and tertiary structure as mentioned in Chapter 1. There are 18 cysteine and 12 proline residues conserved in all MSP sequences (Appendix 1, Figure A1) which led to the deduction that these proteases have conserved structural characteristics.

Open source structure prediction programmes were used to determine the possible secondary and tertiary structures of putative *T. vivax* MSPs*.* The I-TASSER (**I**terative **T**hreading **ASSE**mbly **R**efinement; Figure 1.2) program, used to predict the three-dimensional structures of each *T. vivax* MSPs, was selected based on its rating, its accessibility and its ease of use.

Rating of protein prediction software is conducted by the community-wide Critical Assessment of Structural Prediction (CASP) experiments for which I-TASSER has been rated top for a number of years running (2007-2011). Tertiary protein prediction programs, such as MODELLER, which is rated as one of the best software, requires program download and an understanding of programming language, while I-TASSER is completely web-based and results are sent to the user after modelling is complete. It was for these reasons that this software was selected for protein modelling.

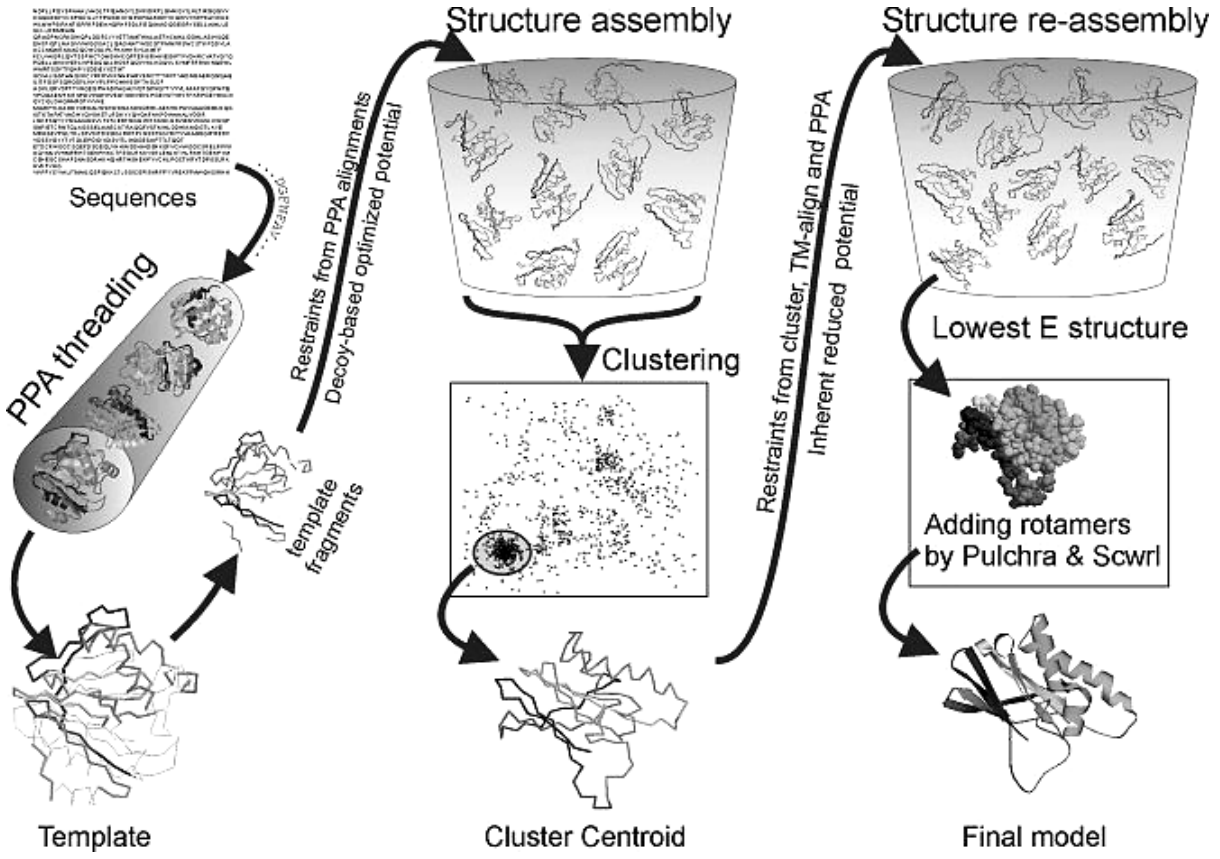


Figure 2.1: Schematic diagram of I-TASSER hierarchical method of three-dimensional protein modelling. The protocol is based on the use of protein structure categorisation using secondary structure profiling. Sequences are threaded through profile-profile alignments (PPA) until all regions are aligned with known structures which allows for the development of template structural fragments. All possible fragments are grouped together in such a way to form a number of full length structural models which are compared to known structures. The models that cluster together with known structures are then used to undergo profile-profile alignment again as well as using the structural alignment program called TM-align. This then allows for structural reassembly where those models with the lowest E-values are selected for the addition of rotamers (or N, C, O backbones) to produce a final model. The newer versions of I-TASSER also include function prediction but it was not included in this schematic (Zhang, 2007).

The I-TASSER method of modelling is based on hierarchical protein structure categorisation using secondary structure profiling (Figure 2.1). Sequences are threaded through a number of secondary structure profiles until all regions are aligned with the primary and secondary structures of proteins with known tertiary structure (Zhang, 2008). Regions that show gaps or a low level of alignment are designated as separate domains and multiple domains are then determined. This allows for modelling of both full length protein and separate domains giving a more complete final model (Zhang, 2008). Finally, domains are docked together to form a final full length protein which is superimposed with the most closely related protein that has a known three-dimensional structure. Five models are produced and users then decide which best represents the protein's structure. The full outline of I-TASSER method of protein modelling is explained by Roy *et al.* (2010).

The results described in this chapter include the identification of putative gene and protein sequences from NCBI for a number of *T. vivax* MSP classes. These sequences were subjected to secondary and tertiary structural bioinfomatic investigations in an attempt to show conservation of key characteristics of this family of proteases. This analysis was done in parallel to the Marcoux *et al.* (2010) modelling study on MSPs in *T. congolense*.

2.2 Materials and Methods

2.2.1 Selection of *T. vivax* MSP Sequences

The methods used within this Chapter are outlined in Figure 2.2 below. The NCBI database (www.ncbi.nlm.nih.gov/) and Basic Local Alignment Search Tool for protein (BLASTp-www.ncbi.nlm.nih.gov/BLAST/) were used to find putative MSP sequences within the *T. vivax* genome. The NCBI protein database was searched using these specific phrases: '*Trypanosoma vivax* gp63′, '*Trypanosoma vivax* major surface protease', '*Trypanosoma vivax* putative M8 peptidase'. Six sequences comprising 400-700 amino acid residues were selected as putative MSP sequences. BLASTp was then used to localise sequences from *T. vivax* with all *T. brucei* and *T. congolense* MSP sequences. Two additional putative MSPs comprising 400-700 amino acid residues were found. Eight putative *T. vivax* MSP protein sequences were identified (Accession numbers: CCC52269.1, CCC53317.1, CCC53338.1, CCD18177.1, CCD18122.1, CCD19977.1, CCD21328.1, CCD20103.1). ClustalW (www.ebi.ac.uk/Tools/msa/clustalw2) was used in an alignment of all MSPs from *T. brucei, T. congolense, T. cruzi* and leishmanolysin. This alignment was used to construct an unrooted phylogenetic tree. The sequences that shared the same clade with each class (A, B, C, D or E) of MSP were selected as putative *Tv*MSPs.

The genomic organisation of the identified proteases, within the *T. vivax* genome, was done by finding the corresponding gene for each selected protease as well as the size and chromosome on which each gene resides, using the NCBI nucleotide database.

2.2.2 Primary Structural Analyses

Primary structural analysis was done by using ClustalW within the BioEdit platform to align the selected sequences with other MSPs. The alignment was then used to identify conserved amino acid residues, such as histidine and cysteine, which are known to have important functions. Sequence identity and theoretical sizes of *Tv*MSPs were used to assess the similarities between each MSP class and within each trypanosome species, in comparison with the analyses outlined in Chapter 1.

2.2.3 Secondary Structural Analyses

T. vivax MSPs were analysed in four secondary prediction programmes: Quick2D (http://toolkit.tuebingen.mpg.de/quick2_d), PredictProtein (https://www.predictprotein.org/), PSIPred (http://bioinf.cs.ucl.ac.uk/psipred/) and Scratch Protein Predictor (http://scratch.proteomics.ics.uci.edu/). The resulting information was tabulated for comparison and identification of important structural characteristics. The sequences were also subjected to GPI predictions software: PredGPI (http://gpcr.biocomp.unibo.it/predgpi/).

2.2.4 Tertiary Structural Analyses

Each *T. vivax* MSP was submitted for three-dimensional structure modelling using the software I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). The method with which this software designs three-dimensional models is explained in Section 2.1. The program outputs five of the best tertiary models in protein data bank (.pdb) format. Those with the highest confidence score were then converted to Cn3D format (.cn3) using VAST search (http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html) for their manipulation in the Cn3D software (downloadable from the NCBI structure database). Cn3D-converted models produce structures highlighting individual domains. Model one of *Tv*MSP-A and -E with the highest confidence score from the I-TASSER server, gave three domains as seen for leishmanolysin and these models were therefore used. Model one of *Tv*MSP-C and -D with the highest confidence scores gave domain outlines that were incomparable to leishmanolysin and so model two from the I-TASSER server was used. The models were compared to the known leishmanolysin structure and the addition of a zinc atom was done to help with the orientation of each model. All editing of the models was done within the Cn3D software and saved as PNG images (.png).

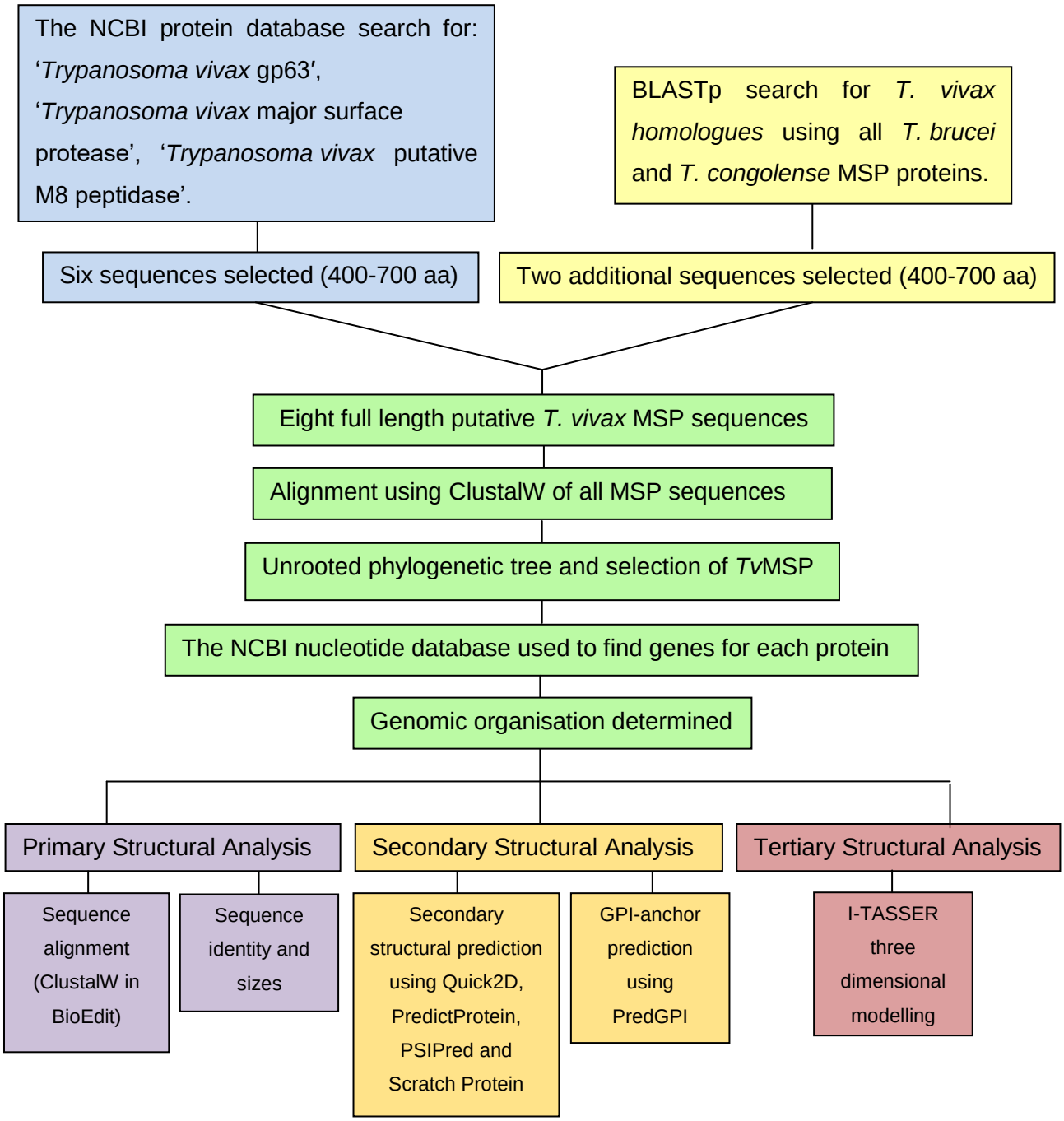


Figure 2.2: Flow diagram outlining the methods used in the work reported in this chapter. NCBI, National Centre for Biotechnology Information; BLAST, Basic Local Alignment Search Tool; aa, amino acid residues; MSP, Major Surface Protease; GPI, Glycosylphosphatidylinositol; and I-TASSER, Iterative Threading ASSEmbly Refinement.

2.3 Results

2.3.1 Selection of *T. vivax* Protein Sequences and Genomic Organisation

The NCBI protein and BLASTp databases were searched to find a total of eight 'full length' MSP protein sequences (400-700 aa) for *T. vivax* (Jackson *et al.*, 2012). A multiple sequence alignment was done using ClustalW (Larkin *et al.*, 2007) with these MSP sequences and used to construct an unrooted phylogenetic tree (Figure 2.3). An unrooted tree was used because the ancestry of these proteases is uncertain and since evolutionary distance is not considered in this study, a rooted tree is not required. The tree was used to simply give visual representation of which sequences group together. In Figure 2.3 (panel A), CCC52269.1 is grouped in the same clade as *Tb*MSP-A and *Tco*MSP-A and is named *Tv*MSP-A in further analyses. Sequence CCC53317.1 is grouped in the same clade *Tco*MSP-D and is named as *Tv*MSP-D. Sequence CCC53338.1 is grouped in the same clade as *Tco*MSP-E and is named *TvMSP-E*. Sequence CCD18177.1 is grouped in a subclade of *Tb*MSP-C and *Tco*MSP-C and is named *Tv*MSP-C. None of the sequences found were grouped in the same clade as *Tb*MSP-*B, Tco*MSP-B or *Tc*MSP-B and therefore no protein sequence for MSP-B were identified for *T. vivax*. All the other sequences are a branch of CCD18177.1 (*TvMSP-C*). As shown in Section 1.7, MSP-C and -D, and MSP-B and -E show higher sequence identity with each other while the MSP-As are grouped on their own (Figure 2.3, panel A). Leishmanolysin is grouped in the same clade as *Tco*MSP-D which was unexpected as its sequence identity is highest with *Tco*MSP-B (Appendix 1).

Corresponding nucleotide sequences, for each protein sequence, were found and the genomic organisation is shown in Figure 2.3 (panel B). The *TvMSP-A* nucleotide sequence occurs on chromosome 10, which is unlike other *MSP-A* genes from other *Trypanosoma* species where it resides on chromosome 11. *MSP-C* from *T. congolense* and *T. brucei* reside on this chromosome instead. However, like *TcoMSP-D* and *-E*, *TvMSP-D* and *-E* also resides on chromosome 11 approximately 38 kb apart. The size of *TvMSP-D* and *TcoMSP-D* are the same (2 kb), while the size of *TvMSP-E* (1.64 kb) is similar to *TcoMSP-E* (1.7 kb). *TvMSP-A* (1.8 kb) is similar to *TbMSP-A* (1.9 kb) and *TcoMSP-A* (1.7 kb), as well. The chromosome on which the *TvMSP-C* gene resides could not be determined and its size (2.1 kb) is larger than any other *MSP-C* genes.

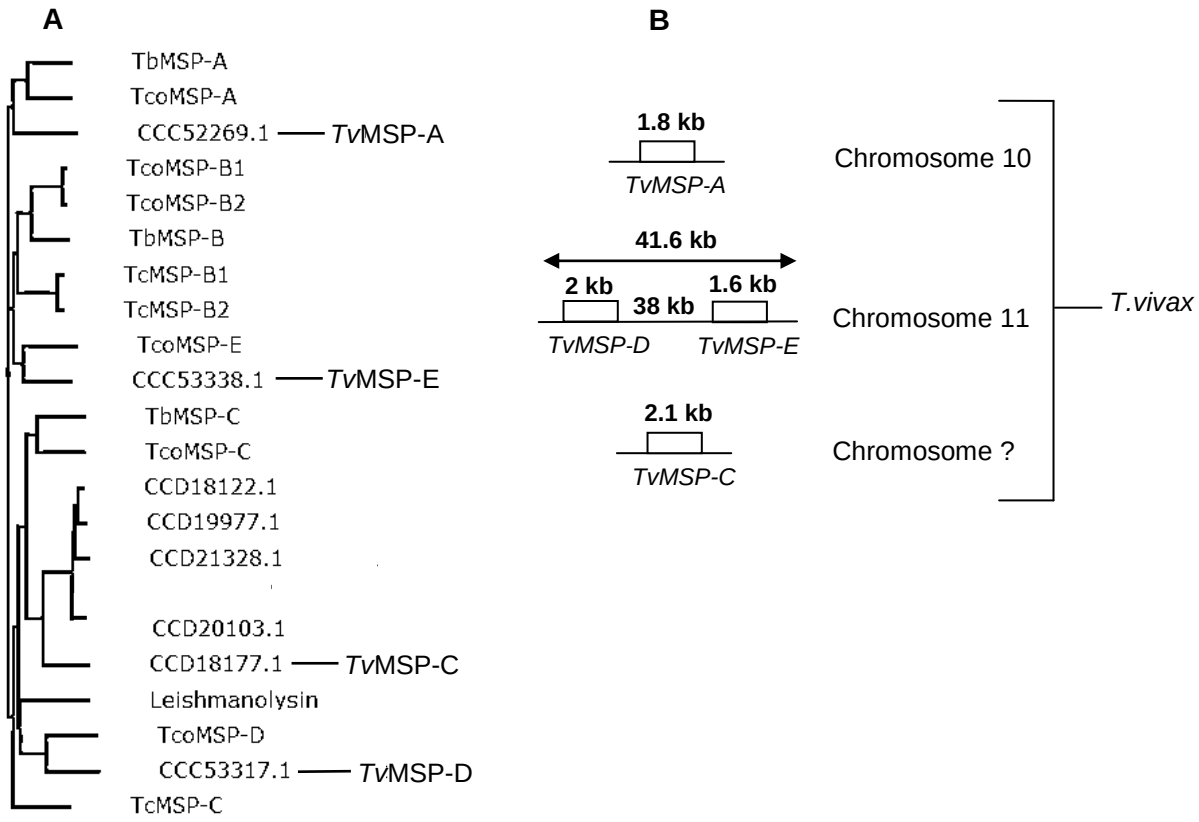


Figure 2.3: Phylogenetic representation and genomic organisation of *T. vivax* MSP protein sequences in relation to MSP classes from *T. brucei, T. congolense*, *T. cruzi* and leishmanolysin. (A) Phylogenetic tree showing selected sequences for each class of *Tv*MSP. Note that no sequence was found to be in the same clade as MSP-B. (B) Arrangement of different *TvMSP* genes on corresponding chromosomes within the *T. vivax* genome. ? denotes unknown information. (Saitou and Nei, 1987; Larkin *et al*., 2007).

2.3.2 Primary Structure Analyses

For primary structural analysis the sequence alignment shown in Appendix 1 and the schematically mapped conserved regions were shown (see Figure 1.6). The primary structure of *T. vivax* MSPs contains all conserved cysteine residues and all semi-conserved proline residues (Appendix 1) which is indicative of conserved secondary and tertiary structure (LaCount *et al.*, 2003). All *Tv*MSP sequences included the HEXXHXXGF truncated motif, and ~60-70 amino acid residues C-terminal to this motif the 'third' histidine ligand (conserved in all sequences).

Interestingly, *Tv*MSP-A shares the highest identity with *TbMSP*-A but *Tv*MSP-C, -D and -E have highest identity with *Tco*MSPs of the corresponding classes (Appendix 1). Overall, *T. vivax* MSPs at ~70 kDa are generally larger in size compared to all other MSPs with *Tv*MSP-C being the largest at 77 kDa (Appendix 2).

T. vivax MSPs show highest variability in size between each class of protein (Appendix 2). All *T. vivax* sequences also contain the conserved cysteine residue within their propeptide (Figure 1.6), like leishmanolysin, which suggests a similar mechanism of enzyme inhibition. *Tv*MSP-E is the only *T. vivax* sequence that contains a valine dyad where leishmanolysin cleaves its pro-region to become fully activated (Button and McMaster, 1988).

2.3.3 Secondary Structure Analyses

Secondary structure analysis was done by comparing four prediction programs: Quick2D, PredictProtein, PSIPred, and Scratch Protein Predictor and is presented in Appendix 5. All *Tv*MSP sequences for all predictions show three prominent helixes which are called helix A, B and C as seen in leishmanolysin and other metzincin and gluzincin proteases (Matthews *et al.*, 1972; Gomis-Rüth *et al.*, 1993; Gomis-Rüth *et al.*, 1994). There are highly variable regions N-terminal from helix A, in all *Tv*MSPs, which is most likely due to the presence of a propeptide. *Tv*MSP-C has the largest propeptide with helix A starting at the 200th residue compared to the ~120th residue for the other *Tv*MSPs.

All contain the HEXXH motif within helix B as seen for all zincin proteases (Bode *et al.*, 1993) and the amino acid insert before the 'third' histidine seen only in MSPs. C-terminal to the 'third' histidine ligand is the 1, 4 Met-turn (which gives metzincins their name) present in all *Tv*MSPs, except the D class, and almost all MSPs (Appendix 1). Both *Tv*MSP-D and *Tco*MSP-D lack this methionine residue. Comparable to all metzincin proteases, helix C follows shortly after the Met-turn for all *Tv*MSPs (Barrett *et al.*, 2004). Within the C- terminal regions the predicted number of β-sheets exceeds the number of α-helices by greater than 10 for all classes of the *Tv*MSP proteases.

Leishmanolysin has three main domains: N-terminal, central and C-terminal. Helices A and B, including the HEXXH motif, are present in the N-terminal domain, the third histidine ligand, Met-turn and helix C are in the central domain with a β-sheet sandwich within the C-terminal domain. The domain predictions for *Tv*MSPs show no correlation to the leishmanolysin domains. None correspond to the pattern seen with helices A and B in domain one; third ligand, Met-turn and helix C in domain two followed by domain three. However, C-terminal regions of all the proteases are dominated by β-sheets, which are comparable to that of leishmanolysin.

All *Tv*MSP are unlikely to be transmembrane proteins as expected (Appendix 5). These proteases are suspected to be attached by a GPI-anchor to the surface of the protein as seen for *Tb*MSP-B (LaCount *et al.*, 2003; Grandgenett *et al.*, 2007).

Further analysis using the pred-GPI program showed that *Tv*MSP-A and -E are highly likely to be GPI-anchored to the surface of the parasite while *Tv*MSP-D is highly unlikely to be GPI-anchored, like all other MSPs. In contrast to other MSP-C classes, *Tv*MSP-C is highly likely to contain a GPI-anchor comparable to *Tv*MSP-A and -E (Appendix 5). GPI-anchor prediction was also done on the unused sequences (CCD18122.1, CCD19977.1, CCD21328.1, CCD20103.1) from the *T. vivax* genome and it was found that two were weakly probable to have an anchor while two were highly probable to have an anchor comparable to *Tv*MSP-C (CCD18177.1).

2.3.4 Tertiary Structural Analyses

I-TASSER was used to predict the three-dimensional structure of the four putative *Tv*MSPs (Figures 2.4-2.7). Each model is shown with a zinc atom used for their orientation in comparison with leishmanolysin shown previously (Figure 1.7).

Models for *Tv*MSP-A and –E have similar structures to each other and to leishmanolysin while models constructed for *Tv*MSP-C and -D are similar to each other but contain a number of extra loops and flaps particularly N- and C-terminally. The core structure, however, is highly comparable to that of leishmanolysin as seen in Figure 1.7 with each containing an observed helix A, helix B (with the truncated HEXXH active motif), followed by a third histidine ligand, Met-turn, helix C and a β-sheet sandwich. The domain prediction follows the expected pattern in most cases. All modelled structures contain a predicted propeptide which is not seen in the active leishmanolysin three-dimensional structure (Figure 1.7).

The predicted propeptide *Tv*MSP-As (Figure 2.4), shown in yellow, is tightly coiled and remains atop the structure. The propeptide contains the conserved cysteine residue and forms a ‘lid’ with possible blocking of the putative active site which may prevent entry of larger substrate entry (Figure 2.4). The leishmanolysin propeptide is suspected to act using the cysteine switch mechanism in which a highly conserved cysteine residue forms a complex with the zinc cofactor in the active site (Macdonald *et al.*, 1995). In this case, the conserved cysteine residue would be in the correct orientation to form a complex with the zinc atom.

The N-terminal domain of the protein, shown in purple, has a compacted C-terminal ‘tail’ and two extra helices (C-terminal to helix B) not seen in the N-terminal domain of leishmanolysin (compare Figure 2.4 to Figure 1.7). This domain, however, contains conserved structures such as helices A and B (with the HEXXH motif). Conversely, helix A is separated into two smaller helical structures rather than a single large one.

The central domain, shown in blue, containing the third histidine ligand, Met-turn and helix C (Figure 2.4), is very similar to the leishmanolysin central domain. It can also be seen that the HEXXH motif and the third histidine ligand are in a similar position to that of leishmanolysin which is optimal for zinc binding (shown in Figure 2.4 with a grey sphere). Shown in brown is the C-terminal domain which has the same orientation as in leishmanolysin with two anti-parallel sheets leading to the β-sheet sandwich.

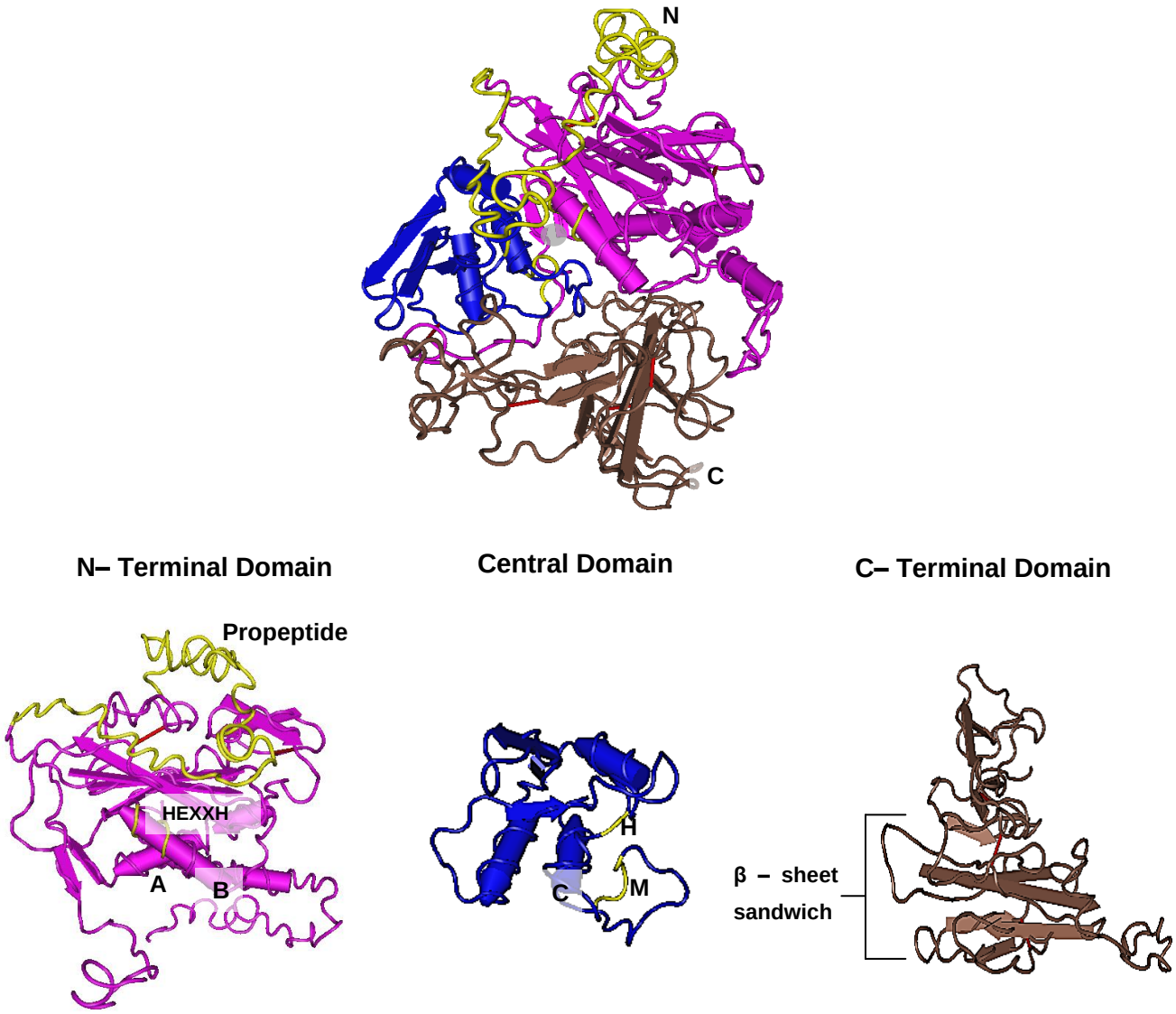


Figure 2.4: *Tv*MSP-A modelled by I-TASSER modelling program. Purple indicates predicted N-terminal domain, blue-central domain and brown-C-terminal domain. N- and C- termini are labelled. Yellow indicates predicted pro-region of the enzyme while the HEXXH motif, third histidine ligand (H), Met-turn (M) and helices A, B and C are indicated on the structure. Red lines indicate suspected disulfide bonds. The grey sphere represents the possible positioning of the zinc atom as in leishmanolysin (added after modelling and used for orientation).

The comparison of modelled *Tv*MSP-A to modelled *Tco*MSP-A (Marcoux *et al.*, 2010) shows similarity in zinc atom orientation with a helical structure (assumed to be helix B) and a second helical structure in close proximity (assumed to be helix A). However, differences seen between the two include a reduced number of helices and β-sheets and the lack of a β-sheet sandwich.

The propeptide of modelled of *Tv*MSP-C (indicated in yellow) is longer in length and its orientation does not block the putative active site in any way that is visible (Figure 2.5). The N-terminal domain, shown in purple, contains helices A and B with the HEXXH motif as expected (Figure 2.5). Helix A is a single long helical region as seen in leishmanolysin (Figure 1.7) but unlike *Tv*MSP-A (two separate helices; Figure 2.4). Overall, the conformation of this domain resembles that of leishmanolysin including the compacted β-sheets; loop regions and the C-terminal tail (compare Figure 1.7 to Figure 2.5). It also does not contain any extra helical structures like those in *Tv*MSP-A.

The central domain, shown in blue, has an arrangement of the third conserved histidine ligand, Met-turn and helix C that resembles those of leishmanolysin and seem to be conserved throughout all classes of the protease. Again, helix B and the third histidine ligand are in optimal positioning for zinc binding as compared with leishmanolysin (Figure 2.5, grey sphere).

The C-terminal domain was predicted to have two separate domains (brown and grey) but when grouped together and compared to leishmanolysin's C-terminal domain (compare Figure 1.7 to Figure 2.5), the grey (extra domain) resembles the N-terminus of the domain with the same two anti-parallel β-sheets. The domain shown in brown contains the C-terminal β-sheet sandwich. Unlike leishmanolysin, this protein is predicted to have an extra C-terminal loosely folded tail. This region is not found in any other family of MSPs and may take into account the larger size of *Tv*MSP-C compared to its homologues (Appendix 2).

Modelling of *Tco*MSP-C by Marcoux *et al.* (2010) shows the zinc atom orientation to be similar to that in *Tv*MSP-C again with a prominent helix as well as a loop region (assumed to be helix B and the third His ligand). No extra tail region was observed.

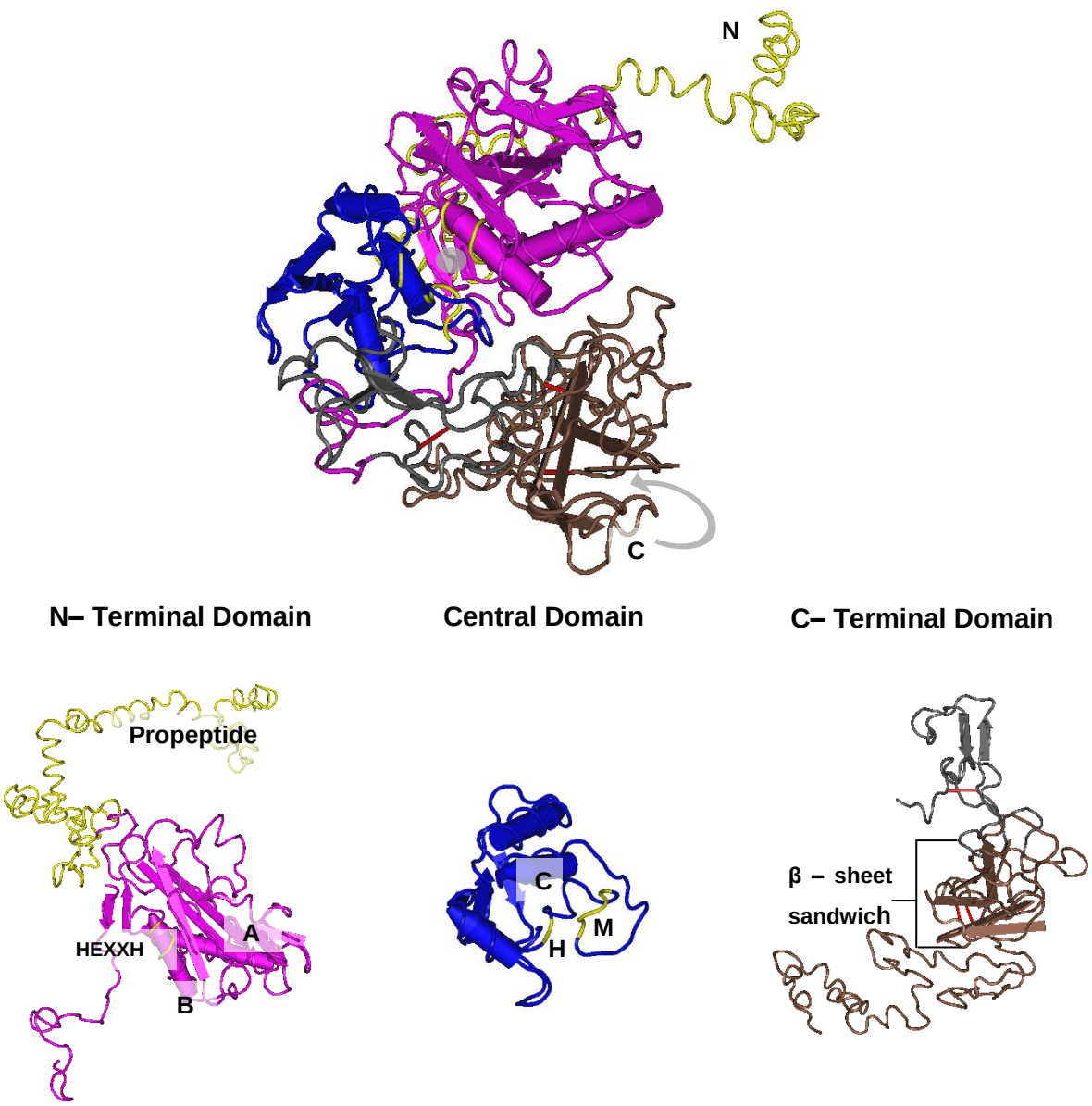


Figure 2.5: *Tv*MSP-C modelled by I-TASSER modelling program. Purple indicates predicted N-terminal domain, blue-central domain and brown-C-terminal domain. N- and C- termini are labelled. Yellow indicates predicted pro-region of the enzyme, while the HEXXH motif, third histidine ligand (H), Met-turn (M) and helices A, B and C are indicated on the structure. Red lines indicate suspected disulfide bonds. The grey sphere represents the possible positioning of the zinc atom as in leishmanolysin (added after modelling and used for orientation).

The MSP-D class of proteins is only present in *T. congolense* and *T. vivax.* In the model shown in Figure 2.6, the propeptide is coloured yellow and contains an extra helical structure which corresponds with predicted secondary structure (Appendix 5), but is not seen in other *Tv*MSPs. Orientation of the propeptide, which again contains the conserved cysteine residue expected to be involved in protease inhibition, is behind the putative active site.

The N-terminal domain is shown in purple (Figure 2.6) and contains helix A and B as well as the HEXXH motif. Helix A is also shown as a single long helical region rather than two separate smaller ones, comparable with *Tv*MSP-C and leishmanolysin. This domain does, however, contain two extra N-terminal helical regions which are not present in leishmanolysin (N-terminal domain dominated with β-sheets and helix A and B are the only two present: compare Figure 1.7 to Figure 2.6). This is comparable to *Tv*MSP-A; however, the two extra helical structures of *Tv*MSP-D are more C-terminal (compare Figure 2.4 to Figure 2.6). This domain lacks the C-terminal tail seen in leishmanolysin, *Tv*MSP-A and -C.

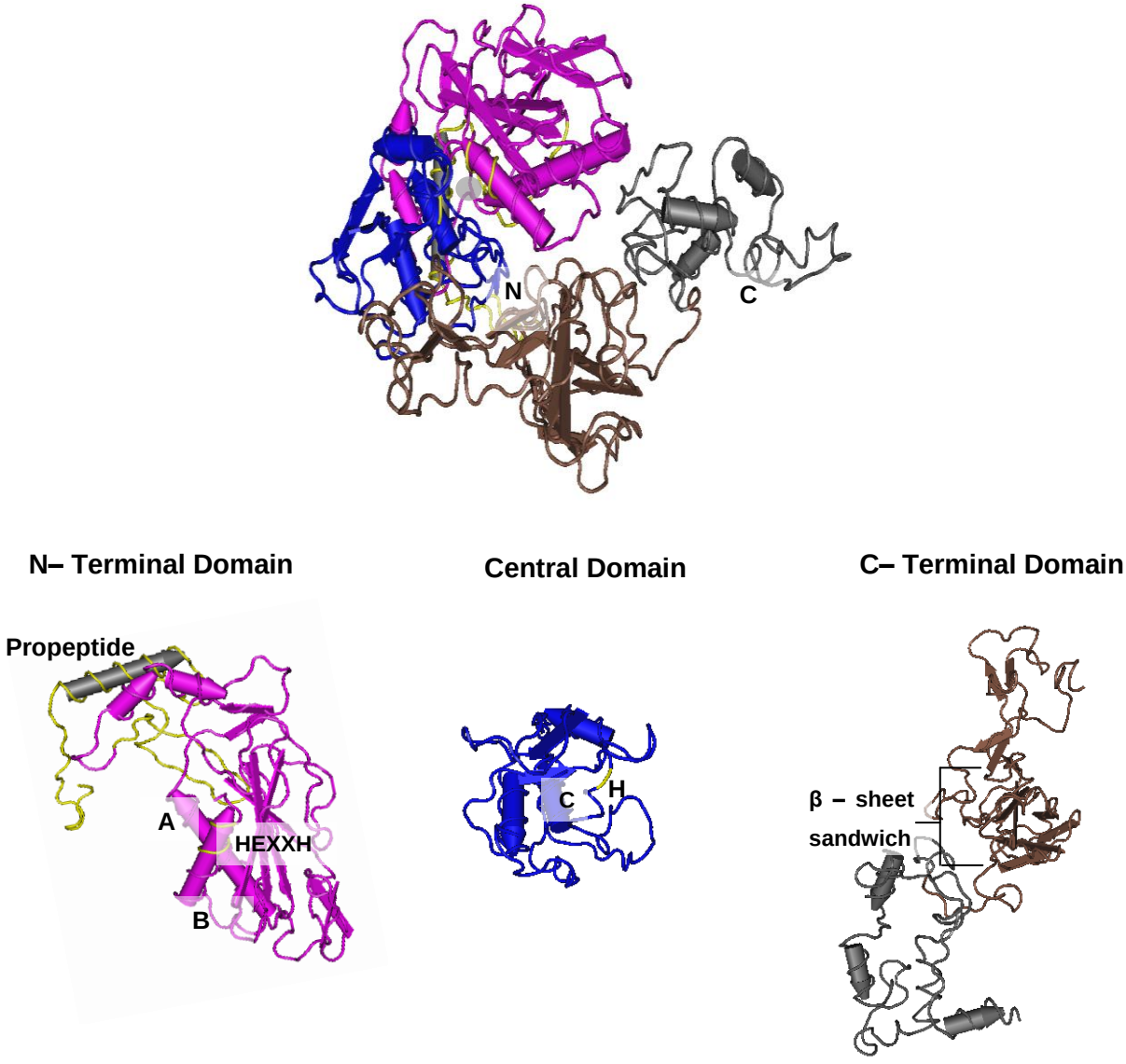


Figure 2.6: *Tv*MSP-D modelled by I-TASSER modelling program. Purple indicates predicted N-terminal domain, blue-central domain and brown-C-terminal domain. N- and C- termini are labelled. Yellow indicates predicted pro-region of the enzyme, while the HEXXH motif, the third histidine ligand (H) and helices A, B and C are indicated on the structure. No Met-turn is present within D classes of the protein. The predicted structure shows no disulfide bonds. The grey sphere represents the expected positioning of the zinc atom as in leishmanolysin (added after modelling and used for orientation).

The MSP-D class lacks the conserved methionine (Figure 1.6) which is characteristic of metzincin metalloproteases and is present in all other MSPs studied thus far. Thus, unlike leishmanolysin and other *Tv*MSPs, the central domain, shown in blue, lacks this feature. However, it does contain the third histidine ligand and helix C in the same conformation as other *Tv*MSPs and leishmanolysin which results in a similar folding compared to other central domains. The histidine ligands are in the correct orientation for zinc binding which is necessary for enzyme activity.

Like *Tv*MSP-C, *Tv*MSP-D's C-terminal domain consists of two separate domains shown in brown and grey (Figure 2.6). However, the first brown domain resembles the entire leishmanolysin C-terminal domain and the 'fourth' grey domain represents an extra C-terminal region of the protein. The predicted structure does contain the two anti-parallel β-sheets leading to the compacted β-sheet sandwich, like all other *Tv*MSPs and leishmanolysin. The extra domain is incomparable to the extra region seen in *Tv*MSP-C which is more a tail rather than an individual domain. The sizes of *Tv*MSP-D (75 kDa) and *Tv*MSP-C (77 kDa) are relatively similar which may indicate that *Tv*MSP-C might have extra loops and flaps within the actual protein structure rather than excess N- and C-terminal regions only. Whether this extra fourth domain in *Tv*MSP-D has a functional role in this protein class is unknown.

The *Tco*MSP-D model (Marcoux *et al.*, 2010) shows the same zinc orientation as seen for all other MSPs modelled to date including the expected helix A and helix B. However, it is lacking a number of β-sheets in the N-terminal domain that *Tv*MSP-D and leishmanolysin contain, as well as the full β-sheet sandwich seen in all *Tv*MSPs. It does not have the extra C-terminal domain seen in *Tv*MSP-D either.

The MSP-E class of proteins, like MSP-Ds, are also found in *T. congolense* and *T. vivax* only, but show most sequence identity with MSP-A. The propeptide of *Tv*MSP-E, shown in yellow, is predicted to lie atop the main protein structure (Figure 2.7) and like *Tv*MSP-A forms a 'lid' over the active site. It also contains the conserved cysteine residue which, in this orientation, would be able to form a complex with the zinc cofactor.

The N-terminal domain, shown in purple, contains the same helix A and B with the HEXXH motif (Figure 2.7). It resembles *Tv*MSP-D since it lacks the C-terminal tail observed in the N-terminal domain of *Tv*MSP-A, -C and leishmanolysin (compare Figure 2.7 to Figure 1.7); however, its helix A is separated into two smaller helical regions like *Tv*MSP-A. It does not contain the extra two helices as is the case for *TvMSP*-A and -D which is comparable with leishmanolysin and *Tv*MSP-C.

The central domain, shown in blue, resembles all other classes containing the third histidine ligand, Met-turn and helix C (Figure 2.7). Helix B and the third His ligand are also in optimal positioning for zinc binding (Figure 2.7, grey sphere). The C-terminal domain, shown in brown, contains the two anti-parallel β-sheet regions leading to the β-sheet sandwich. This domain closely resembles *Tv*MSP-A.

The zinc atom orientation of the *Tco*MSP-E model's (Marcoux *et al.*, 2010) is incomparable to that of *Tv*MSP-E with no helix B being present for binding. There is also no β-sheets and only three prominent helices (one forming helix A) throughout the entire structure.

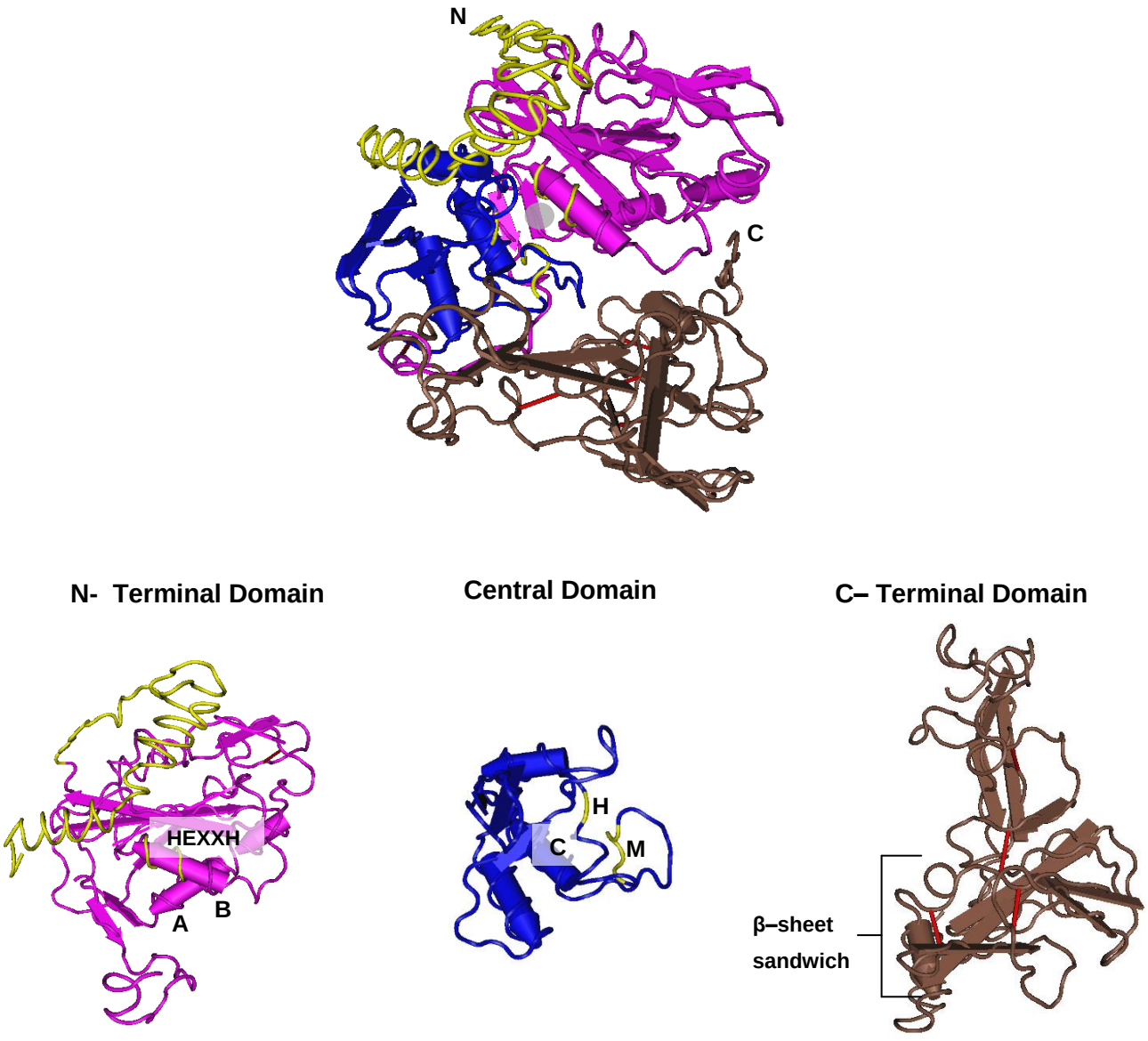


Figure 2.7: *Tv*MSP-E modelled by I-TASSER modelling program. Purple indicates predicted N-terminal domain, blue-central domain and brown-C-terminal domain. N and C termini are labelled. Yellow indicates predicted pro-region of the enzyme, while the HEXXH motif, third histidine ligand (H), Met-turn (M) and helices A, B and C are indicated on the structure. Red lines indicate suspected disulfide bonds. The grey sphere represents zinc atom expected positioning as in leishmanolysin (added after modelling and used for orientation).

The modelling software (Figures 2.4-2.7) and the Scratch Protein Predictor (Appendix 5) showed various numbers of disulfide bonds in each *Tv*MSP, however, since all *Tv*MSPs contain the 18 conserved cysteine residues found within leishmanolysin and other trypanosoma MSPs (Figure 1.6, Figure 1.7 and Appendix 1), it is suspected that all contain nine disulfide bonds. Similar to leishmanolysin though, all disulfide bonds were predicted to be in the C-terminal domain for all four *Tv*MSPs (Figure 2.4-2.7).

All protein sequences selected to represent the *T. vivax* MSP classes have key characteristics which shows that they belong to this group of metalloproteases. However, each individual protein also has unique features which shows why they are in different classes of MSPs.

2.4 Discussion

The NCBI database was searched for putative MSP sequences within the African trypanosome, *T. vivax* for which eight sequences were found ranging in size from 400-700 amino acid residues. After phylogenetic comparison of all eight sequences with all other MSP proteases from *T. brucei, T. congolense, T. cruzi* and leishmanolysin, four were selected to represent classes MSP-A, -C, -D and -E. These four *T. vivax* MSP sequences then were subjected to a number of secondary and tertiary structure bioinfomatic analyses. They showed key characteristics within their predicted structures that are conserved and are comparable with leishmanolysin and metalloprotease of the metzincin family. Each class also contained unique features explaining the presence of individual classes of the protease. The evidence provided here again shows how these proteases are conserved across species, genera and families which could suggest that their function has some importance in these parasites (Chaudhuri and Chang, 1988; Etges, 1992; Ramamoorthy *et al.*, 1992; El-Sayed and Donelson, 1997a; Cuevas *et al.*, 2003; Jesudhasan *et al.*, 2007; Marcoux *et al.*, 2010; Oladiran and Belosevic, 2012).

No protein sequence was found to be phylogenetically grouped with the MSP-B class and for the remainder of this study, this protein class was omitted. *Tv*MSP-C, when subjected to phylogenetic construction with the omission of all other subsequences listed in Figure 2.3, appears as a parent sequence to the other MSP-Cs (data not shown). This sequence was also the only MSP-C protein that was predicted to be GPI-anchored to the surface of the parasite, unlike other MSP-Cs which are unlikely to be GPI-anchored [Marcoux *et al.* (2010) and the points described in Chapter 1]. GPI-anchor prediction was also done on the unused sequences (CCD18122.1, CCD19977.1, CCD21328.1, CCD20103.1) from the *T. vivax* genome and it was found that two were weakly probable to have an anchor while two were

highly probable to have an anchor (data not shown) comparable to *Tv*MSP-C (CCD18177.1). These characteristics of *Tv*MSP-C (and sequences CCD20103.1 and CCD21328.1) resemble MSP-Bs from other species and so it is hypothesised that '*Tv*MSP-B' may have higher identity with possible *Tv*MSP-Cs. Sequences CCD18122.1 and CCD19977.1 would then represent the true MSP-C class from *T. vivax* which is unlikely to be GPI-anchored as expected. This is, however, speculative and it is uncertain whether the *T. vivax* genome contains an MSP-B gene at this point.

All the sequences selected were predicted to include the propeptide of the protein. Leishmanolysin has been shown to have latent activity which is activated by the treatment with organomercurials (Macdonald *et al.*, 1995). The activated enzyme is unable to cleave latent enzyme into its fully activated form which means that leishmanolysin is not autocatalytic (McGwire and Chang, 1996). Treatment by organomercurials, as a means to activate the enzyme, is commonly observed with activation of matrix metalloproteases and the cysteine switch mechanism (Stricklin *et al.*, 1983). Matrix metalloproteases all contain the consensus sequence 'P**RC**GVPDV' which includes this cysteine residue (Park *et al.*, 1991). The theory is based on the binding of the sulfhydryl group of a conserved cysteine residue with the zinc cofactor of the enzyme thereby blocking the active site (Van Wart and Birkedal-Hansen, 1990). Organomercurials are able to break this bond which then allows removal of the propeptide. Mercury chloride has been shown to fully activate leishmanolysin (Macdonald *et al.*, 1995). A conserved cysteine residue is found in all MSP sequences and a number contain the 'RC' dyad seen in the matrix metalloprotease consensus sequence (Park *et al.*, 1991).

Orientation of the propeptide shown in the modelling of *Tv*MSP-A and -E forms a 'lid' over the active site and is in prime position for cysteine interaction with the zinc cofactor within the putative active site. This folding would explain the latent activity in leishmanolysin, before activation, since the active site is not completely blocked and may still allow low levels of substrate cleavage. *Tv*MSP-C and -D propeptide modelling showed the active site open and exposed.

Since I-TASSER bases its modelling on predicted domains and because *Tv*MSP-A and -E both have the 'RC' dyad within their propeptide, modelling of this region for these two sequences could have been based on known structures of modelled matrix metalloprotease propeptides such as procollagenase, transin (rat stromelysin) and human stromelysin (Stricklin *et al.*, 1983; Park *et al.*, 1991; Gomis-Ruth, 1997). *Tv*MSP-C and -D do not contain the 'RC' dyad and perhaps its propeptide modelling was based on a different known structure and would explain the differences seen here.

The *T. congolense* and *T. vivax* genomes contain the classes MSP-D and -E which are not present in *T. brucei*. These proteases are suspected to have similar roles to MSP-C and -B, respectively because they are either known (Grandgenett *et al.*, 2007; Shimogawa *et al.*, 2015) or predicted to be GPI-anchored (MSP-B, -A and -E) or predicted not to be GPI-anchored (MSP-C and -D). Since *T. brucei* has four tandemly repeated *MSP-B* genes while *T. congolense* has two classes of MSP-B (Figure 1.5), it may suggest that MSP-E in *T. congolense* acts like MSP-B when this protease is required in high amounts. This suggests different regulation of expression of these proteases within the two species. Furthermore, because *T. vivax* has both MSP-A and -E, that are suspected to have similar roles to MSP-B, it does not require this enzyme in such high quantities and the MSP-B class could have been lost from the *T. vivax* genome.

The MSP-D class is unique because it does not contain the conserved methionine residue (Marcoux *et al.*, 2010) found in all metzincin proteases (Bode *et al.*, 1993). In studying the crystal structure of leishmanolysin, it was determined that due to the extra loops between the second and 'third' histidine ligand, the methionine turn which helped in zinc binding, loses its function (Schlagenhauf *et al.*, 1998). Therefore, this class of protease is more than likely to have the same function as the other proteases. *Tv*MSP-D does have ~10 extra amino acid residues between the second and 'third' histidine ligands which may help in ligand orientation in the absence of the methionine.

Overall, four putative *T. vivax* MSP sequences were selected to represent classes A, C, D and E. These selected sequences underwent bioinformatic investigation to determine important key features of these proteases as well as unique features within each class. These identified sequences were used in conjunction with *T. brucei* and *T. congolense* sequences to identify immunogenic peptides for the production of anti-peptide antibodies (Chapter 3). This allowed for the exploration of these proteases as potential diagnostic markers. These proteases were also recombinantly expressed and purified, to give insight into their multimerisation and enzymatic function (Chapter 4).

Chapter 3

Anti-peptide Antibody Production against *T. brucei, T. congolense* and *T. vivax* MSP-B and -C: IgY antibodies vs single chain Fv fragments (scFvs)

3.1 Introduction

Human African trypanosomiasis (HAT) and animal African trypanosomiasis (AAT) are two of the many neglected tropical diseases causing health, agriculture and economical burdens to the continent. To combat any disease, accurate diagnosis is required to ensure correct treatment and since the symptoms of tropical diseases are similar (Tiberti *et al.*, 2013), direct clinical diagnosis is challenging and often unreliable. For HAT, disease transmission is suspected to be under control since the number of reported cases are on a decline, however, aspects of patient management still need improvement [reviewed in Tiberti *et al.* (2013)]. AAT control has been less effective with the disease being highly prevalent throughout countries of sub-Saharan Africa [reviewed in Odeyemi *et al.* (2015)]. AAT has also spread to previously trypanosomosis-free areas such as the Jos Plateau, Nigeria [occurred in the last two decades; Majekodunmi *et al.* (2013)]. A more recent publication showed that the Mayo Rey division of Cameroon's cattle still have an annual prevalence of bovine trypanosomiasis of 55.2% (Abdoulmoumini *et al.*, 2015). Furthermore, assessment of AAT susceptibility of cattle farming communities within sub-Saharan Africa showed that these communities can be clustered into groups of low to high vulnerability (Holt *et al.,* 2016). It was suggested that for effective control, specialised methods for each area of vulnerability is necessary (Holt *et al.*, 2016). Therefore, new diagnostic and disease staging tools are required to continue the fight against both HAT and AAT.

Synthetic peptides corresponding to immunogenic epitopes in proteins have been used to produce site-specific antibodies for their use in both diagnostic tests and vaccine development (Posnett *et al.*, 1988; Gómara and Haro, 2007). These peptides provide a powerful tool for the production of antibodies that can be polyclonal, yet mono-specific. This would allow for the detection of an array of antigens to just a single antigen (Lee *et al.*, 2010). The generation of these types of antibodies involves the determination of immunogenic regions within the antigen using epitope prediction software. Whether the peptide is conserved between proteins of different species or unique to a protein from a single species determines the detection specificity of the resulting antibody.

Recently developed diagnostic tests for HAT and AAT that use parasite antigen for the detection of serum antibodies include HAT SERO K-SeT (detects patient serum antibodies against *T. b. gambiense* VSGs LiTat 1.3 and 1.5), HAT Sero-Strip (detects patient serum antibodies against *T. b. gambiense* VSGs LiTat 1.3 and 1.5), SD BIOLINE HAT (detects patient serum antibodies against *T. b. gambiense* VSGs LiTat 1.3 and 1.5) and Surra SERO K-SeT [detects animal serum antibodies against *T. evansi* VSGs Rotat 1.2; (Büscher *et al.*, 2014; Birhanu *et al.*, 2015; Jamonneau *et al.*, 2015)]. The identification of antigen p310 (an ISG) with the ability to specifically detect *T. vivax* infected cattle sera in a diagnostic setup, is in early stage field testing (Fleming *et al.*, 2016). These detection systems are rapid, easy to use and field-applicable which is suitable for diagnosis in the sub-optimal, poor infrastructural environment found in sub-Saharan Africa where trypanosomiasis occurs. The tests are similar to more commonly used diagnostic tests, in both specificity and sensitivity, with most detecting trypanosomiasis with 95% accuracy (Büscher *et al.*, 2014; Birhanu *et al.*, 2015; Jamonneau *et al.*, 2015). As HAT infected cases decreases, diagnostic tests require increased sensitivity. Studies have shown that detecting more than one type of antigen within a single test may help solve this problem (Jamonneau *et al.*, 2015). It has also been suggested that to distinguish between post and current infection an antigen detecting diagnostic test is better.

The research reported in this chapter focused on the production of anti-peptide antibodies against Major Surface Proteases (MSPs) found in *T. brucei, T. congolense* and *T. vivax*. Two methods for anti-peptide antibody production were used; namely, antibodies produced in chickens and using scFv phage technology. Since MSP-A and MSP-B are known to be present on the surface of the trypanosome (LaCount *et al.*, 2003; Shimogawa *et al.*, 2015) and MSP-E is suspected to be additionally GPI-anchored in the trypanosome membrane, while, MSP-C and -D are suspected to be secreted into the mammalian host circulation, would infer that there are a high number of these genes within the trypanosome genome. This may result in a number of possible antigen targets for the development of a diagnostic test. Additionally, it has also been shown that GPI-anchored proteins are likely to induce an immune response (Brown and Waneck, 1992) which may result in serum antibodies available for diagnostic detection. Although, the expression level of these proteins by the parasites is unknown at present, they are highly conserved amongst *Trypanosoma* species. Therefore, a diagnostic assay based on MSPs could provide a pan-specific test for trypanosomes.

The general structure of an antibody consists of two heavy chains and two light chains. A single heavy chain consists of a single variable (VH) and either three constant (CH) domains for IgG, IgA and IgD or four CH domains for IgM, IgE and IgY.

A single light chain consists of a single variable (VL) and a single constant (CL) domain (Figure 3.1). These heavy and light chains can be linked together to form smaller antibody types (Figure 3.1) that have different detection abilities (Skottrup, P 2010). One of these antibody fragments involves the joining of just the variable regions, VH and VL, to form fragment variable (Fv) which together still contains the antigen binding site (Skottrup, 2010). When recombinantly expressed with a linker region $(Gly_4Ser)_3$ that joins them together, these variable regions are called scFvs (Holliger *et al.*, 1993). These recombinantly expressed scFvs have the ability to recognise and bind antigens and are produced using phage display technologies (Skottrup, 2010).

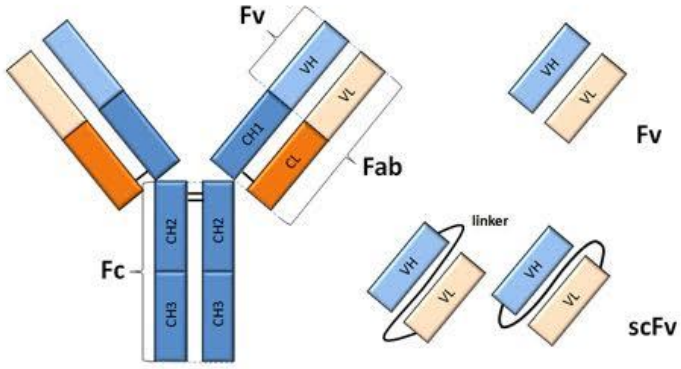


Figure 3.1: Schematic representation of a complete antibody and various antibody fragment types. CH1-3, constant heavy chain 1 - 3; VH, variable heavy chain; CL, constant light chain; VL, variable light chain; Fc, fragment crystallisable; Fab, fragment antigen binding; Fv, fragment variable; scFv, single chain fragment variable (Nelson, 2010).

Phage display technology is a molecular technique that uses filamentous phage particles to express protein libraries, such as scFvs, and display them on their surface [reviewed in Azzazy and Highsmith (2002)]. Filamentous phages are viruses that are able to infect Gram negative bacteria (like *Escherichia coli*) where they use the bacteria to propagate the next generation of virus. The virus structure consists of five major and minor coat proteins (pIII, pVIII, pVI, pIX and pVII) surrounding single-stranded (ss) DNA [Figure 3.2; (Wang *et al.*, 2006)].

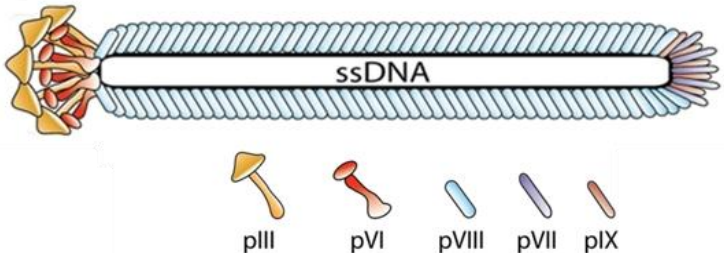


Figure 3.2: Schematic representation of the assembly of wild-type M13 phage coat proteins. (Løset *et al.*, 2011).

During bacterial infection, the ssDNA is injected and converted to double-stranded (ds) DNA that acts as a template for the production of the various viral coat proteins as well as more ssDNA for the next phage generation. These coat proteins are then expressed on the infected *E. coli* surface which allows for non-lytic blebbing of new viral particles or phage (Endemann and Model, 1995). Essentially, the virus that leaves the *E. coli* is a protein coat encapsulating the viral ssDNA (Figure 3.2). Coat proteins have functions in binding *E. coli*, injecting DNA into *E. coli* and general protection and encapsulation of its DNA. Phage display technology uses coat protein genes to attach foreign genes that code for the selected protein library (Smith, 1985).

In the present study, the Nkuku® phagemid library, which codes for Fv regions originally from the chicken genome, was used (Van Wyngaardt *et al.*, 2004). The phagemid is a vector of DNA that resembles a plasmid vector because it contains an origin of replication, a *c-myc* detection tag, ampicillin resistance gene, some restriction sites, an amber codon as well as a coat protein gene followed by a foreign gene (Hoogenboom *et al.*, 1991). In the case of the Nkuku® phagemid library, the coat protein pIII's gene was selected to attach the scFv gene which results in an expressed pIII-scFv coat protein. The genes inserted are an array of chicken immunoglobulin genes plus those that contain inserted mutations to give a library of possible scFvs which are transformed into *E. coli* TG1 cells (Figure 3.3). The phagemids' DNA alone is ineffective at forming new viral particles, since, the pIII contains the scFv portion and the viral particle is thus unable to reinfect *E. coli,* which means it remains within *E. coli* cells where it replicates in a similar manner to a plasmid. In culture, the ampicillin resistant *E. coli* cells grow with pIII-scFv coat proteins displayed on their outer membrane (Figure 3.3).

To allow for the reformation of phagemid viral particles, the full viral genome is required including wild-type pIII coat proteins. This is done by adding helper phage (wild type M13K07 phage) which are able to infect *E. coli* cells, injecting the full viral genome. Infected *E. coli* TG1 cells produce all viral coat proteins including wild-type pIII and pIII-scFv, resulting in a heterogenous population, containing both phagemid genome and helper phage genome (Figure 3.3). M13K07 ssDNA has a kanamycin resistance gene while the phagemid has an ampicillin resistance gene; therefore, the *E. coli* cells are grown in the presence of both antibiotics to select for cells that contain both genomes. This results in the formation of a mosaic of viral particles that contain either phagemid genome or wild-type genome with either pIII-scFv only, wild-type pIII only or both pIII-scFv and wild-type pIII on their surface (phages: single copy of DNA, *E. coli:* more than one copy of DNA; Figure 3.3).

This mosaic population of phages are then purified and allowed to react with the antigen (in this case a peptide) which allows for selection of phages that contains scFvs that recognise the antigen. Phages that do not recognise antigen, as well as those only expressing pIII without the scFv fragment are lost. The phages that contain specific pIII-scFv fusion protein bound to antigen are then allowed to reinfect *E. coli* cells (with a loss of phage that do not contain wildtype pIII) and grown in ampicillin containing medium (selects for phagemid genome and loss of helper phage genome; Figure 3.3). This results in the selection of the phagemid genome that expresses scFvs specific for the antigen. This entire selection process is called panning which is repeated a minimum of four times and results in a polyclonal scFv population (can detect a range of epitopes on the antigen).

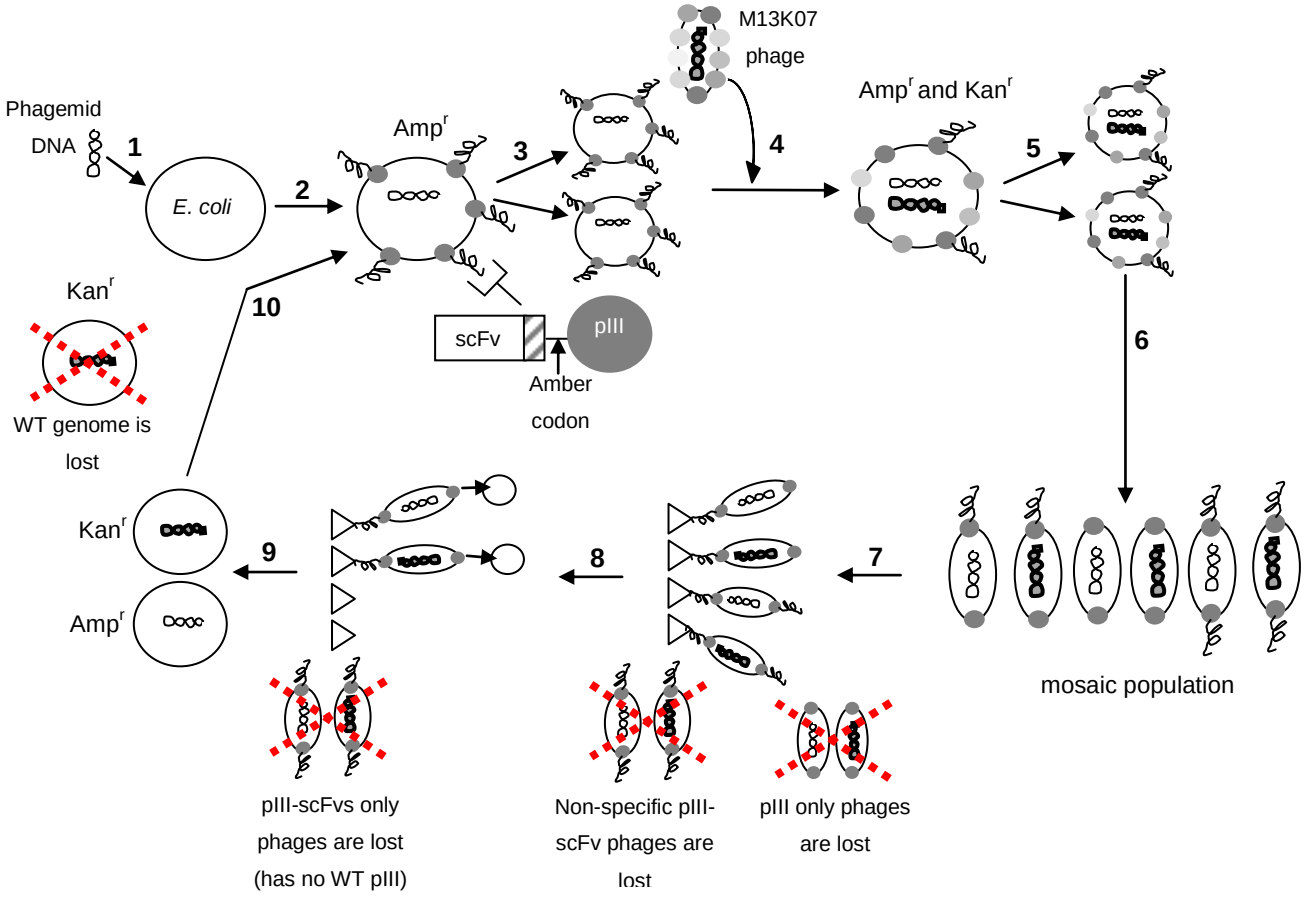


Figure 3.3: Schematic representation of the use of phage display technology to select for scFv specific fragments. 1 - phagemid DNA containing gene coding for scFvs is transformed into *E. coli* to form a library of possible scFv fragments. 2 - pIII-scFv conjugate expressed on *E. coli* surface. Striped block represents *c-myc* tag. 3 - *E. coli* cells with phagemid DNA are Amp^r and are able to propagate. Phagemid DNA is also copied. 4 - phagemid DNA is rescued by the addition of helper phage which also infects *E. coli* cells. 5 - *E. coli* with both phagemid (Amp^r) and helper phage (Kan^r) genome are propagated in the presences of both Amp and Kan. 6 - blebbing of new phage particles results in mosaic population. 7 - phagemids allowed to react with antigen (need pIII-scFv to bind antigen). 8 - bound phagemid allowed to infect *E. coli* (need WT pIII to infect *E. coli* cells). 9 - *E. coli* cells have either phagemid genome (Amp^r) or helper genome (Kan^r). 10 - growth in ampicillin medium selects for cells with phagemid genome only. Amp^r - ampicillin resistance. Kan^r - kanamycin resistance. WT-wild-type.

The polyclonal phagemid population can be used to select for monocolony scFv expressing clones (individual colonies that express a single scFv phenotype and detect a single epitope). The specific phagemid population derived finally is used to express scFvs in a soluble form via an amber codon selection process (Qi *et al.*, 2012). This codon is present between the pIII and scFv gene in the phagemid genome (Figure 3.1). It can either act as a codon coding a glutamic acid or as a stop codon, depending on the strain of *E. coli* containing the DNA. The *E. coli* TG1 strain is unable to recognise the stop codon and so expresses a pIII-scFv fusion protein. However, strains of *E. coli,* such as TOP 10, are able to read the stop codon which results in expression of pIII and scFv as separate proteins (Qi *et al.*, 2012). The scFv protein can then be affinity purified using anti-*c-myc* antibodies.

The objective of the work described in this chapter was to produce antibodies and scFvs against three MSP peptides. Peptides were selected from protease MSP-B and -C from *T. brucei, T. congolense* or *T. vivax* with the objective that the anti-peptide antibodies would detect full length MSP proteins. Peptides were selected from only these two protease classes with the purpose of showing native MSP-B and MSP-C in parasite lysate. Comparison of antibody binding capacity of the two different scaffolds of antibody produced (complete antibody and scFvs) may result in a novel diagnostic detection system. These antibodies were used for the further analyses of recombinant MSP-C described in Chapter 4.

3.2 Materials and Methods

3.2.1 Materials

All reagents used during the study were of analytical grade. Reagents Na_2HPO_4, NaH_2PO_4, NaN_3, Ethylenediaminetetraacetic acid (EDTA), sodium dodecyl sulfate (SDS), NaCl, KCl, KH_2PO_4, NaOH, HCl, PEG (Mr 6000), Tween 20, H_2O_2, citric acid, tryptone, yeast extract, bacteriological agar, glycine and glycerol were purchased from Merck (Gauteng, RSA). 5,5'-dithio-bis-[2-nitrobenzoic acid] (Ellman's reagent), dimethylsulfoxide (DMSO) and dimethylformamide (DMF) were purchased from Sigma-Aldrich (USA). Tris-HCl was purchased from Melford (UK) and dithiothreitol (DTT) was purchased from ThermoScientific (USA).

IgY antibody production and purification: Peptides were synthesised by GL Biochem (China) and SulfoLink® resin was purchased from ThermoScientific (USA). Freund's complete and incomplete adjuvants, rabbit serum albumin (RSA), m-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) and L-cysteine were purchased from Sigma-Aldrich (USA).

Phage Display Technology: M13K07 wild type phage, Nkuku® phagemid display library and *E. coli* TG1 cells were obtained from Prof. Dion du Plessis and Dr Jaenni Fehrsen, Immunology Division, Onderstepoort Veterinary Institute, South Africa. NUNC® Immuno™ tubes were purchased from ThermoFisher Scientific (USA). Ampicillin was from Amresco (USA) and kanamycin was purchased from Life Technologies (UK). Sterile filters (0.22 µm and 0.45 µm) were purchased from PALL Life Sciences (USA) and sterile 96 well microtitre plates from Nest BioTech co (China).

Enzyme-Linked Immunosorbent Assay (ELISA): Bovine serum albumin (BSA) and 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic) acid (ABTS) were purchased from Roche (Germany). NUNC® Maxisorb non-sterile 96 well microtitre plates were purchased from ThermoFisher Scientific (USA) and non-fat milk powder from Elite (RSA). Goat anti-mouse IgG HRPO antibody was purchased from Jackson Immunochemical (USA), mouse anti-M13 antibody from ThermoScientific (USA) and rabbit anti-chicken horse radish peroxidase (HRPO) antibody from Sigma-Aldrich (USA).

Parasite Purification: DE-52 pre-swollen diethylaminoethyl (DEAE) cellulose resin was purchased from Sigma-Aldrich (Germany).

3.2.2 Peptide Design

T. brucei, T. congolense and *T. vivax* MSP-B and -C amino acid sequences were aligned using ClustalW (www.ebi.ac.uk/Tools/msa/clustalw2/) in BioEdit. Conserved and nonconserved regions were noted. Predict7 epitope prediction software (Cármenes *et al.*, 1989) was used to calculate the hydrophilicity (Kyte and Doolittle, 1982), surface probability (Chothia and Janin, 1975), flexibility (Karplus and Schulz, 1985) and antigenicity (Welling *et al.*, 1985) of each protease and prediction plots were constructed in MS Excel. Using both the alignment and Predict7 plots, peptides were selected to produce anti-peptide antibodies. Peptide *Tb/Tco*MSP:303-314 was selected since it shares high identity (>55%) with both *T. brucei* and *T. congolense* MSP-B which would result in an antibody that can detect both species. Peptides were also designed that are unique to *T. brucei* (*Tb*MSP:400-412) and *T. vivax* (*Tv*MSP:686-697), respectively. This was done by selecting peptides with sequence identity of less than 50% in comparison with full length proteases from other species. The peptides selected were used to produce antibodies in chickens (Section 3.2.3) and used to select scFvs from the Nkuku® phagemid library (Section 3.2.6).

3.2.3 Coupling Peptide to Carrier Protein and Chicken Immunisation

Four milligrams of each peptide was weighed out and dissolved in 50 µl of 100% (v/v) DMSO to which 450 µl reducing buffer [100 mM Tris-HCl, 1 mM EDTA, 0.02% (w/v) NaN_3, pH 8.0] was added. DTT (10 mM) was added to make a final volume of 1 ml. The solution was incubated at 37^{o}C for 90 min before loaded onto a Sephadex G-10 (1.5 cm x 11 cm) column equilibrated with 5 column volumes of MEC buffer [100 mM NaH_2PO_4, 0.02% (w/v) NaN_3, pH 7.0]. The eluted fractions (25 x 500 µl fractions) were collected and tested for reduction of peptide using Ellmans reagent [100 mM Tris-HCl , 1 mM EDTA, 0.1% (w/v) SDS, pH 8.0 supplemented with 2 mg Ellmans reagent]. Each fraction (10 µl) was mixed with Ellmans reagent (10 µl) and reduced peptide resulted in a yellow colour. All fractions that gave a yellow colour were pooled.

Rabbit serum albumin (RSA) was used as a carrier protein for attachment to each peptide via a terminal cysteine residue using MBS. Ratio of reduced peptide to rabbit albumin (RA) carrier used was 40:1 while the ratio of rabbit serum albumin to MBS was 1:40. The following equations were used to calculate the amount of carrier and MBS required for coupling:

$$Mass\ of\ RSA = mass\ of\ peptide\ \times \frac{(Mw\ of\ RSA\ \times 0.025)}{Mw\ of\ peptide}$$

$$Mass\ of\ MBS = mass\ of\ RSA\ \times \frac{(Mw\ of\ MBS\ \times 40)}{Mw\ of\ RSA}$$

*Mw: Molecular Weight

RSA was dissolved in 500 µl phosphate buffered saline (PBS; 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.2) while MBS was dissolved in 200 µl DMF and 300 µl PBS. The two solutions were mixed together and incubated at room temperature (RT) for 30 min before loaded on to a Sephadex G-25 (1.5 cm x 13 cm) column equilibrated with 3 column volumes MEC buffer. Fractions were collected (20 x 1 ml fractions) and absorbance at 280 nm was measured using a Nanodrop spectrophotometer (Thermoscientifc, USA).

Fractions with absorbance at 280 nm of greater than 0.3 were pooled. The reduced peptide fractions and MBS-activated RSA fractions were then mixed together and incubated at RT for 3 h. The peptide-RSA conjugate was aliquoted and stored at -20^{o}C until required for the immunisation of chickens.

The use of all animals for research was approved by the Biomedical Research Ethics Committee (053/15/Animal) at the University of KwaZulu-Natal. Two chickens per peptide were used to produce IgY antibodies. Each chicken was injected twice intramuscularly (once per breast; 400 µl per site) every two weeks, four times in total. The first immunisation was done using Freund's complete adjuvant followed by boosters using Freund's incomplete adjuvant. For immunisation, 1.5 ml of coupled peptide was emulsified in 1.5 ml Freund's complete/incomplete adjuvant. Each immunisation site was cleaned using 70% (v/v) ethanol and 750 µl of emulsified peptide adjuvant mixture was injected using a 2 ml syringe and 20G sterile needle. A non-immune egg was collected on the day of the first immunisation as a non-immune control. Eggs were collected daily for 16 weeks for the isolation of IgY antibodies.

3.2.4 Purification of Anti-peptide Antibodies

The purification of anti-peptide antibodies is divided into two steps: bulk isolation of all IgY antibodies from chicken egg yolk and affinity purification of specific anti-peptide antibodies.

3.2.4.1 IgY Isolation from Chicken Egg Yolks

The method used to isolate chicken IgY antibodies was done as described by Polson *et al.* (1985) and amended by Goldring and Coetzer (2003). Briefly, eggs were cracked open and all egg white removed by running the yolk under tap water. The yolk sac was punctured and the yolk volume determined. Working buffer [100 mM Na_2HPO_4, 0.02% (w/v) NaN_3, pH 7.6] was used to dilute the yolk in a 2:1 ratio. PEG (Mr 6000) was added at a 3.5% (w/v) concentration, gently dissolved and centrifuged (4400 x g, 30 min at 4^{o}C) to pellet the unwanted lipid fraction. The supernatant was filtered through cotton wool resulting in a clear filtrate. To the filtered supernatant 8.5% (w/v) PEG (Mr 6000) was added and dissolved to give a final concentration of 12% (w/v). To pellet precipitated proteins the solution was centrifuged (12000 x g, 10 min at 4^{o}C) and the supernatant discarded. The pellet was dissolved in the yolk volume using working buffer and 12% (w/v) PEG (Mr 6000) was added and dissolved. This solution was centrifuged (12000 x g, 10 min at 4^{o}C) and the final IgY pellet dissolved in 1/6 of the original yolk volume using final buffer [100 mM Na_2HPO_4, 0.1% (w/v) NaN_3, pH 7.6]. This was called total IgY and it was further purified using affinity purification.

3.2.4.2 Affinity Purification of Anti-Peptide IgY Antibodies

SulfoLink® resin was used to link peptides, via the terminal cysteine residue (as indicated in Figure 3.4), to the resin. These peptide affinity columns would allow for antigen-antibody

interaction for the purification of anti-peptide antibodies. Eight milligrams of peptide was dissolved in 100 µl of 100% (v/v) DMSO and 400 µl general buffer (50 mM Tris-HCl -HCl, 50 mM EDTA, pH 8.5). DTT (10 mM) was added as a reducing agent (500 µl) to the dissolved peptide and incubated for 90 min at 37^{o}C. The Sephadex G-10 column, equilibrated with general buffer, was used to separate reduced peptide from excess DTT. Fractions were collected (25 x 500 µl fractions) from the column and tested for reduction using Ellman's reagent as described in Section 3.2.3. The reduced peptide fractions were pooled and immediately added to 2 ml of resin (4 ml of the 50% slurry) and incubated at RT for 15 min on an end-over-end rotator, followed by 30 min without rotation. The column was washed with general buffer (3 x 5 ml). To block uncoupled reactive groups on the resin 50 mM L-cysteine was added to the column and incubated at room temperature for 15 min on an end-over-end rotator, followed by 30 min rotation. The column was then washed with general buffer (3 x 5 ml) and stored in phosphate buffer (100 mM Na_2HPO_4, pH 7.6) at 4^{o}C until used.

The SulfoLink® resin was equilibrated with PBS (3 x 5 ml) and total IgY, isolated from immunised chicken egg yolks, was added and incubated at 4^{o}C overnight. The resin was allowed to settle in the column and unbound IgY passed over the resin, by gravity, one more time. The column was then washed with PBS until the absorbance at 280 nm, read on the Nanodrop 6000 spectrophotometer (Thermoscientifc, USA), was less than 0.02. The bound anti-peptide antibodies were eluted by decreasing the pH with elution buffer [100 mM glycine, 0.02% (w/v) NaN_3, pH 2.8] and 10 x 1 ml fractions, containing 100 µl of neutralisation buffer [1 M Na_2HPO_4, 0.02% (w/v) NaN_3, pH 8.5], were collected

The absorbance at 280 nm for each fraction was determined using the Nanodrop spectrophotometer and the final concentration calculated using the Beer-Lambert law and an extinction coefficient for IgY of ε = 1.25. Fractions containing an absorbance of 0.1-0.5, 0.5-0.8 and greater than 0.8 were pooled and stored, in a final concentration of 0.1% (w/v) NaN_3, at 4^{o}C until needed.

3.2.5 Enzyme-Linked Immunosorbent Assay (ELISA)

Microtitre plates were coated overnight with either 1 µg/ml or 5 µg/ml peptide at 4^{o}C. The unoccupied sites were blocked with 200 µl/well of 0.5% (w/v) BSA-PBS for 60 min at 37^{o}C. The plate was washed with PBS Tween-20 [PBS with 0.1% (v/v) Tween-20] using a Biotek ELx50 plate washer (3 x 300 µl/plate). Primary IgY antibody (100 µl/well) was added and incubated for 60 min at 37^{o}C. The plates were again washed with PBS Tween-20 (3 x 300 µl/plate) using the plate washer and incubated with rabbit anti-chicken IgY-HRPO (1:10000) for 60 min at 37^{o}C (120 µl/well).

The last wash was performed as before and 150 µl/well substrate solution [0.05% (w/v) ABTS, 0.005% (v/v) H_2O_2 in 150 mM citrate-phosphate buffer, pH 5.5] added. Absorbance readings at 405 nm were taken either after 15 min, 30 min or 60 min using the FLUOstar Optima spectro-fluorometer (BMG Labtach, Germany).

3.2.6 Phage Display and Anti-Peptide scFvs

A number of steps (Figure 3.3) are used in phage display technology to obtain phagemids expressing specific anti-peptide scFvs which was shown by both polyclonal and monocolony ELISAs (Van Wyngaardt *et al.*, 2004). All the work conducted was in a sterile environment.

3.2.6.1 Culturing *Escherichia coli* TG1 cells

E. coli TG1 cells were streaked onto TYE agar plates [10 g tryptone, 5 g yeast extract, 8 g NaCl per litre with 1.5% (w/v) bacteriological agar] and grown overnight at 37^{o}C. A single colony was selected for growth in 2xYT medium (1.6% (w/v) tryptone, 1.0% (w/v) yeast extract, 86 mM NaCl) medium overnight at 37^{o}C. The overnight culture was used to make a 1:100 dilution with 2xYT medium and grown to an optical density (OD) of 0.5 at 600 nm (log phase). These cultures can be stored at 4^{o}C for one week.

3.2.6.2 Culturing and Titering M13K07 Helper Phage

M13KO7 helper phage was cultured and titred and stored at 1 x 10^{12} pfu/ml. The helper phage stock was used to make serial dilutions (10^{-2} - 10^{-10}) with 2xYT medium. Each phage dilution (100 µl) was incubated with log phase *E. coli* TG1 cells (100 µl) for 5 min at RT. The 200 µl of now infected TG1 cells was added to pre-warmed 1.5% (w/v) 2xYT agar and 3 ml liquid 0.7% (w/v) 2xYT agar (kept warm between 44-47 ^{o}C) was added on top and levelled. The top agar was allowed to set and agar plates incubated at 37^{o}C overnight.

A single plaque was cut out of the agar plate and added to 4 ml 2xYT medium. TG1 cells grown overnight were added (40 µl). The culture was incubated at 37^{o}C for 2 h at 100 rpm. This culture (40 ml) was used to inoculate 800 ml of fresh 2xYT medium and grown for 1 h (37^{o}C, 100 rpm). The culture was supplemented with kanamycin (final concentration of 50 µg/ml) and grown overnight (37^{o}C, 100 rpm). The resulting culture was centrifuged (10800 x g, 15 min, 4^{o}C) and the supernatant kept. Helper phage particles from the supernatant were precipitated by adding 1/4 supernatant volume of 20% (w/v) PEG (Mr 6000) in 2.5 M NaCl and incubating on ice for 30 min. The solution was centrifuged (10800 x g, 15 min, 4^{o}C) and the supernatant discarded. The pellet was centrifuged again (2000 x g, 2 min, 4^{o}C) to remove any excess liquid and the final pellet resuspended in 6 ml PBS and sterile-filtered through 0.45 µm filter.

This resulting wild type phage was titred again, pfu/ml calculated and diluted to 1 x 10^{12} pfu/ml using sterile PBS with a final concentration of 20% (v/v) glycerol. The helper phage was stored at -80^{o}C, for use later in rescuing the phagemid library (Section 3.2.6.3).

$$\text{pfu/ml} = \frac{\text{no.of plaques} \times \text{dilution factor}}{(0.5 \times 0.1 \text{ ml})}$$

3.2.6.3 Culturing and Titering Nkuku® Phagemid Library

The phagemid stock obtained from the Onderstepoort Veterinary Institute (Van Wyngaardt *et al.*, 2004) was used to amplify the library and new stocks made and titred. Phagemid stock (250 µl) was added to 500 ml 2xYT medium containing 100 µg/ml ampicillin and 2% (w/v) glucose. The initial OD of this culture at 600 nm should be 0.05. The inoculated medium was grown at 37^{o}C until the OD was 0.5. A 100 ml aliquot of this culture (the rest of the culture was used to make glycerol stocks) was transferred to a clean flask and M13K07 helper phage stock was added at a ratio of 20:1 (20 phage particles for 1 bacterial cell). An OD of 1 at 600 nm is equivalent to 8 x 10^{8} bacterial cells/ml (Van Wyngaardt *et al.*, 2004). This was incubated at 37^{o}C for 30 min standing and 30 min shaking (100 rpm) and this procedure is called rescuing. The culture preparation was centrifuged (3300 x g, 15 min, 4^{o}C) to pellet the cells, resuspended in 1 L of 2xYT medium containing 100 µg/ml ampicillin and 25 µg/ml kanamycin. The resuspended pellet was grown overnight at 30^{o}C (240 rpm). To make glycerol stocks, the rest of the culture was grown for another 3 h and cells were pelleted by centrifugation (3300 x g, 15 min, 4^{o}C). The pellet was resuspended in 2xYT medium (1:100 of original volume) and used to make stocks by adding glycerol to a final concentration of 20% (v/v).

After the rescued phages had grown overnight, the cells were pelleted by centrifugation (3300 x g, 20 min, 4^{o}C) and discarded. The supernatant was used to precipitate phages as before by adding 1/4 (of original volume) 20% (w/v) PEG (Mr 6000) in 2.5 M NaCl on ice for 1 h. The phagemid particles were pelleted by centrifugation (3300 x g, 30 min, 4^{o}C) and the supernatant discarded. The pellet was centrifuged again (2000 x g, 2 min, 4^{o}C) to remove any excess liquid and resuspended in 20 ml PBS (50 x concentrated) and centrifuged (11000 x g, 10-15 min, 4^{o}C) to remove any insoluble components. The supernatant was then sterile-filtered through a 0.45 µm filter. This represents the full Nkuku® library phagemid stock which needs to be titred and used for panning.

To titre the Nkuku® library, serial dilutions are made using 2xYT medium (10^{-2}-10^{-10}). Each dilution (100 µl) was incubated with log phase *E. coli* TG1 cells (100 µl) at RT for 5 min. Infected cells (200 µl) were plated onto TYE A/G agar.

Log phase TG1 cells (100 µl) and phagemid stock (100 µl) only were also plated as controls. Agar plates were incubated overnight at 37°C and cfu/ml is calculated as follows:

$$\text{cfu/ml} = \frac{\text{no.of cells} \times \text{dilution factor}}{(0.5 \times 0.1\ \text{ml})}$$

This is used to determine how much phagemid stock to add during the first round of panning (Section 3.2.6.4).

3.2.6.4 Biopanning

To select for specific phagemids that express scFvs that recognise and bind the peptides designed earlier, biopanning was used. Immunotubes are coated with a decrease in peptide antigen (50, 25, 10 or 5 µg/ml) for all four rounds of panning overnight at 4°C. Decreasing concentrations of peptide per round of panning was used as a selection mechanism to ensure only those phagemids expressing scFvs that bind most efficiently are selected for. The immunotubes were then washed with PBS (3 x) and blocked using 2% (w/v) non-fat milk-PBS for 1 h at RT. The extra immunotubes were stored at 4°C in blocking solution until required. During the blocking step, the phagemid stock wass also incubated with non-fat milk-PBS to remove any phagemids expressing scFvs that recognise the blocking solution (blocked phagemid). This was done by incubation of 10^{12} - 10^{13} cfu of phagemid stock (isolated and titred as described in Section 3.2.6.3) with 2% (w/v) milk-PBS Tween-20 [PBS with 0.1% (v/v) Tween-20] for at least 30 min (3.5 ml). After the blocking step was completed, the immunotubes were washed twice with PBS and the 'blocked' phagemid stock added to the immunotubes which were allowed to stand for 30 min at RT and then rotated for 90 min at RT using an end-over-end rotator.

Once the incubation step was completed, stringent washing was used to ensure only those phagemids expressing scFv with high binding efficiency remain. This was done by washing each immunotube 20 times with PBS Tween-20 followed by washing 20 times with PBS alone. To each immunotube, 3.5 ml log phase *E. coli* TG1 cells were added and incubated at 37°C for a maximum of 30 min to allow the bound phagemid to infect the cells. The *E. coli* solution was immediately centrifuged (3300 x g, 10 min, 4°C) or kept on ice until it can be centrifuged. The pelleted cells were resuspended in 1 ml 2xYT medium and 3 x 333 µl were plated on three TYE A/G agar plates (*E. coli* TG1 cells alone were also plated to ensure cells were not infected). The plates were grown overnight at 30°C. Three plates were used for round one and two and only one plate was used for rounds three and four.

To a single plate 5-10 ml of 2xYT medium was used to loosen the colonies and ~500 µl was added to 50 ml 2xYT medium (containing 100 µg/ml ampicillin and 2% (w/v) glucose) to make a culture of OD at 600 nm of less than 0.05.

This was grown at 37^{o}C until log phase was reached (OD at 600 nm = ~0.5). Only 5 ml of this culture was used to rescue phagemids while the rest was used to make glycerol stocks (culture containing 20% (v/v) glycerol in 1 ml aliquots) and stored at -80^{o}C. To the 5 ml removed from the log phase culture, M13K07 helper phage was added at a 20:1 ratio. Helper phage stocks made earlier were used and an OD at 600 nm of 1 was equivalent to 8 x 10^{8} bacterial cells/ml (Van Wyngaardt *et al.*, 2004). Helper phage was allowed to infect the *E. coli* cells by incubating the culture for 30 min at 37^{o}C with occasional agitation. At this stage some the *E. coli* cells contained both phagemid and helper phage DNA. They were pelleted by centrifugation (3300 x g, 10 min 4^{o}C twice and supernatant discarded each time), resuspended in 25 ml 2xYT medium containing 100 µg/ml ampicillin and 25 µg/ml kanamycin and grown overnight at 30^{o}C (240 rpm). Growth of the *E. coli* cells in both antibiotics selects for those that contain both phagemid and helper phage DNA.

The overnight culture was centrifuged (10000 x g, 10 min, 4^{o}C) and phages precipitated from the supernatant using 1/5 the volume 20% (w/v) PEG in 2.5 M NaCl at 4^{o}C for 60 min. The precipitated phage preparation was pelleted by centrifugation (10000 x g, 10 min, 4^{o}C) and the supernatant discarded. The pellet was resuspended in sterile PBS (1 ml) and centrifuged (11000 x g, 10 min, 4^{o}C) to remove any insoluble precipitate. A volume of 0.5 ml was used for the next round of panning while the other 0.5 ml was stored at 4^{o}C and used in a polyclonal ELISA. The steps listed above were repeated four times with the rest of the coated immunotubes.

3.2.6.5 Polyclonal and Monocolony phagemid ELISA

Polyclonal ELISA

Microtitre plates were coated with 10 µg/ml of peptide overnight at 4^{o}C. Unoccupied sites in the wells were then blocked with 2% (w/v) non-fat milk-PBS Tween-20 for 1 h at 37^{o}C. The 0.5 ml of precipitated phagemid from each round of panning, stored at 4^{o}C, was diluted 1:10 with 2% (w/v) milk-PBS Tween-20 and incubated at RT for at least 30 min (blocked phagemid). The full phagemid library was also included as a control in a separate well.

Following the blocking step, plates were washed with PBS Tween-20 using the BioTek Elx50 plate washer (3 x 300 µl/plate). Blocked phagemid (100 µl/well) were added to the plates and incubated at 37^{o}C for 2 h. The plates were washed as before (3 x 300 µl/plate) and then

incubated with 1:8000 diluted mouse anti-M13 primary antibody in 0.5% (w/v) BSA-PBS at 37^{o}C for 1 h (100 µl/well). To washed plates (as before) 1:10000 diluted goat anti-mouse IgG-HRPO secondary antibody (120 µl/well) was added for 1 h at 37^{o}C. Plates were washed as before and substrate solution [0.05% (w/v) ABTS, 0.005% (v/v) H_2O_2 in 150 mM citrate-phosphate buffer, pH 5.5] added (150 µl). Absorbance readings were taken either after 15 min, 30 min or 60 min using the FLUOstar Optima spectro-fluorometer at 405 nm.

Monocolony ELISA

Single colonies from round four of panning were used in a monocolony ELISA to select for clones that have the best binding efficiency and give the highest absorbance signal. First round four precipitated phage was serially diluted (10^{-2}-10^{-10}) with 2xYT medium and 200 µl plated on TYE A/G agar (as in Section 3.2.6.3). Colonies (96) were used to inoculate sterile microtitre plates which contained 100 µl/well 2xYT A/G medium. The inoculated plates were incubated at 30^{o}C overnight (220 rpm). Each well from the overnight culture (5 µl) was used to inoculate a fresh sterile microtitre plate containing 150 µl/well 2xYT A/G medium and incubated at 37^{o}C for 2.5 h (220 rpm). The rest of the culture within the plate was stored at -80^{o}C after adding 20% (w/v) glycerol.

After the new plate (containing 5 µl overnight culture with 150 µl fresh medium) was incubated at 37^{o}C for 2.5 h, 50 µl of 2×10^{9} pfu M13K07 helper phage was added and allowed to stand for 30 min at 37^{o}C. The plates were centrifuged (600 x g, 10 min, 4^{o}C) and supernatant removed. To each well 150 µl of 2xYT medium containing 100 µg/ml ampicillin and 25 µg/ml kanamycin was added and incubated at 30^{o}C overnight (220 rpm). The overnight plate was centrifuged and the supernatant used to perform an ELISA as described for the polyclonal ELISA. The supernatant (50 µl) was incubated with 4% (w/v) milk-PBS Tween-20 (50 µl) before use in the ELISA.

3.2.7 Parasite Purification and the Detection of MSP in Parasite lysate

Cryopreserved *T. brucei* LiTat 1.3 and *T. congolense* IL3000 parasites were used to infect mice intraperitoneally. Tail pricks were used to monitor parasitaemia using heomocytometer under light microscopy (Olympus, Germany) until the parasite level was 5×10^{4} cells/cm^{2}. Cardiac puncture was performed on anaesthetised mice and blood samples used for parasite purification.

All equipment used was sterile and the purification was done in a sterile environment. Parasite purification was adapted from the method outlined by Lanham and Godfrey (1970). A ~20 ml DE-52 column was set up using a 25 ml syringe with a plastic tap at the bottom.

Whatman no. 1 filter paper was used to plug the bottom of the syringe and a second disc placed on top of the resin. The column was washed with two column volumes sterile PBS and equilibrated with two column volumes sterile PBS with glucose (PSG; 57 mM Na_2HPO_4, 3 mM NaH_2PO_4, 42 mM NaCl, 50 mM glucose, 1 mM hypoxanthine, pH 7.8). The trypanosome-infected blood was diluted with an equal volume of sterile PSG and added to the column. The red blood cells were allowed to settle for 30 min. PSG was then used to elute unbound parasites and four 15 ml samples were collected. The parasites within the collected eluates were obtained by centrifugation (3500 x g, 15 min, 4^{o}C) and resuspended in 1 ml PSG containing 20% (v/v) glycerol. The final number of parasites was counted using a haemocytometer and the samples were stored at -80^{o}C until used.

The parasites were pelleted (2000 x g, 10 min, RT) and resuspended in PBS (150 µl) followed by centrifugation (2000 x g, 5 min, RT). This was repeated twice. The parasite cell pellet was resuspended in lysis buffer (20 mM Tris-HCl, 10 mM EDTA, 1% Triton™ X-100, 10 µM E64 at pH 7.2) and an equal volume of Laemmli's reducing treatment buffer (Section 4.2.7) was added followed by boiling for 10 min. A sample of this was subjected to SDS-PAGE and western blot as described in Chapter 4, Section 4.2.8. The western blot was probed with 5 µg/ml of primary antibody and 1:2500 rabbit anti-chicken secondary antibody conjugated to HRPO. The gel was visualised by ECL (Pierce, USA) using the Syngene, G:BOX (Vacutec, USA).

3.3 Results

3.3.1 Peptide Design

Of the MSPs, classes MSP-B and -C were used to select peptides for the production of anti-peptide antibodies. These two classes were selected because MSP-B is known to be anchored to the surface (LaCount *et al.*, 2003; Shimogawa *et al.*, 2015) while MSP-C is expected to be secreted by the parasite. Since MSP-A and -E resemble MSP-B while MSP-D resembles MSP-C, none of the other classes were used for this study. MSP-B and -C sequences from *T. brucei, T. congolense* and *T. vivax* were analysed and conserved and nonconserved regions were selected to undergo epitope prediction. The Predict7 software (Cármenes *et al.*, 1989) was used to determine which regions of the protein sequences are likely to be epitopes and, therefore, would be able to induce an immune response.

The program incorporates a number of prediction algorithms and four were selected: hydrophilicity (Kyte and Doolittle, 1982), surface probability (Chothia and Janin, 1975), flexibility (Karplus and Schulz, 1985) and antigenicity (Welling *et al.*, 1985). The latter prediction algorithm was developed using data from a small group of proteins for which

epitopes have been confirmed which leaves this parameter the least reliable out of the four. Regions that have these characteristics would be on the surface of the proteins in contact with the surrounding environment and, therefore, the immune system, forming epitopes.

Three peptides were selected for antibody production and the Predict7 plots are shown in Figure 3.4. The peptide, *Tb/Tco*MSP:303-314, was selected from the *T. congolense* MSP-B sequence and there are regions where the prediction characteristics peak above zero (Figure 3.4 A) which is suggestive of an immunogenic region. Comparison of the selected peptide region to leishmanolysin's tertiary structure shows that this peptide is likely to occur on the surface of the protein and is exposed to the surrounding environment (in yellow Figure 3.4 A). Figure 3.5 shows that peptide *Tb/Tco*MSP:303-314 shares 100% identity with *Tco*MSP-B and 77% identity with *Tb*MSP-B, but less than 45% identity with all MSP-Cs. This peptide was designed to detect both *T. brucei* and *T. congolense* MSP-B.

*Tb*MSP:400-412 was chosen from the *Tb*MSP-C sequences and the Predict7 plot again shows regions where all the epitope characteristics peak above zero (Figure 3.4 **B**), therefore, suggesting an immunogenic region. Comparison of the peptide region with the same region in leishmanolysin's tertiary structure shows it on the surface (in yellow, Figure 3.4 **B**) which supports the possibility that the peptide could have interacted with the immune system. *Tb*MSP:400-412 shows 100% identity with *Tb*MSP-C and 54% with *Tb*MSP-B, while, less than 50% identity with *Tco*MSP-B, *Tco*MSP-C and *Tv*MSP-C (Figure 3.5). This peptide was designed to detect *Tb*MSP-C only with a low possibility of detecting other MSP-B and -C.

The last peptide selected came from the C-terminal tail region seen in *Tv*MSP-C (Figure 2.5, p. 48) which has little sequence identity with other MSPs. The Predict7 plot of *Tv*MSP:686-697 shows a peak above zero for all epitope characteristics which is indicative of an immunogenic region. Again, the peptide was compared to leishmanolysin's tertiary structure; however, since the peptide selected is from a region not seen in leishmanolysin, the region could not be visualised on the leishmanolysin structure. There was a peptide within leishmanolysin which showed similarity to *Tv*MSP:686-697 and is shown in the tertiary structure in Figure 3.4 (**C**) in yellow. The sequence identity of *Tv*MSP:686-697 to *Tv*MSP-C was 100% while for the other MSP-B and -C it was less than 10%. This peptide was designed to detect *Tv*MSP-C only (Figure 3.5).

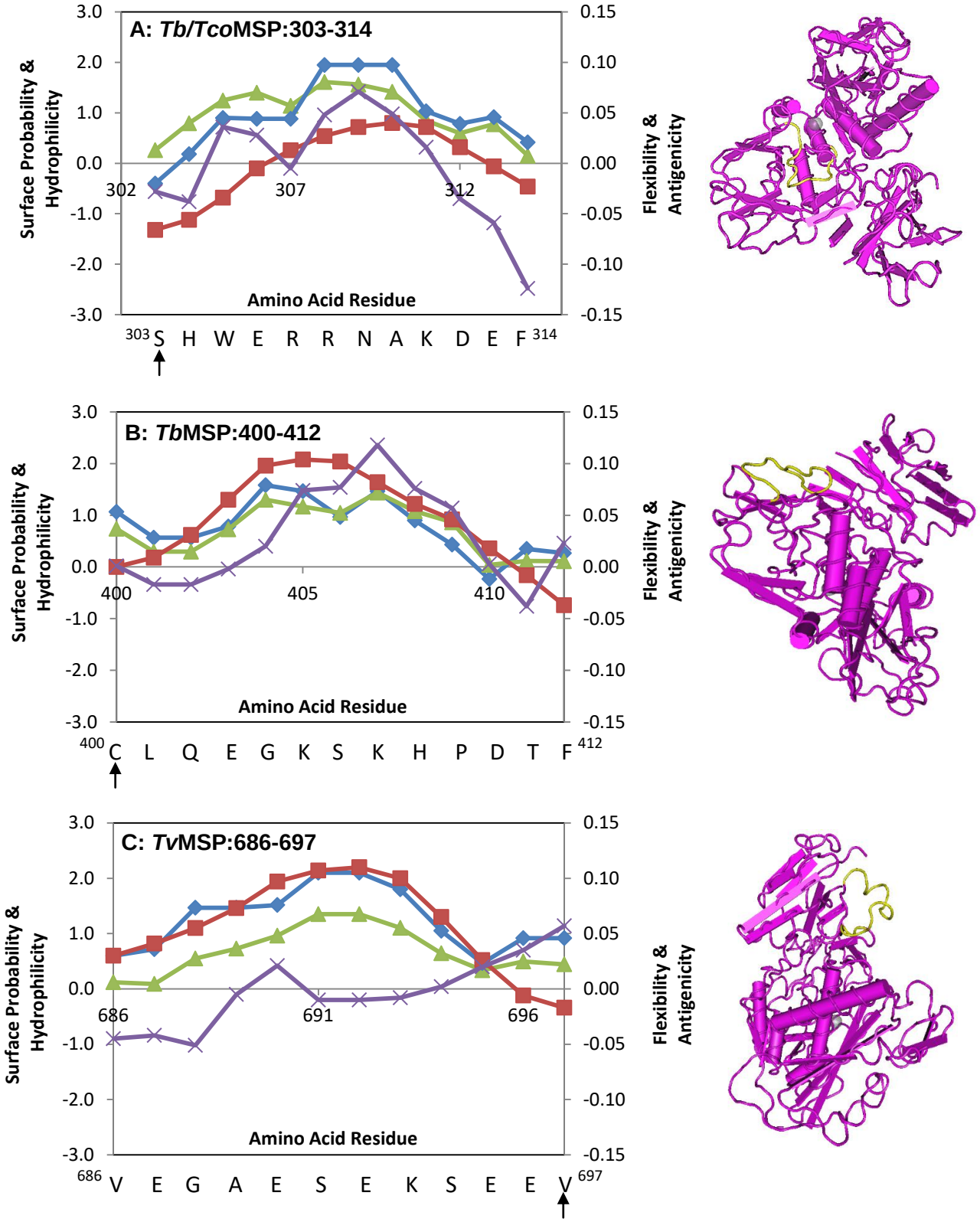
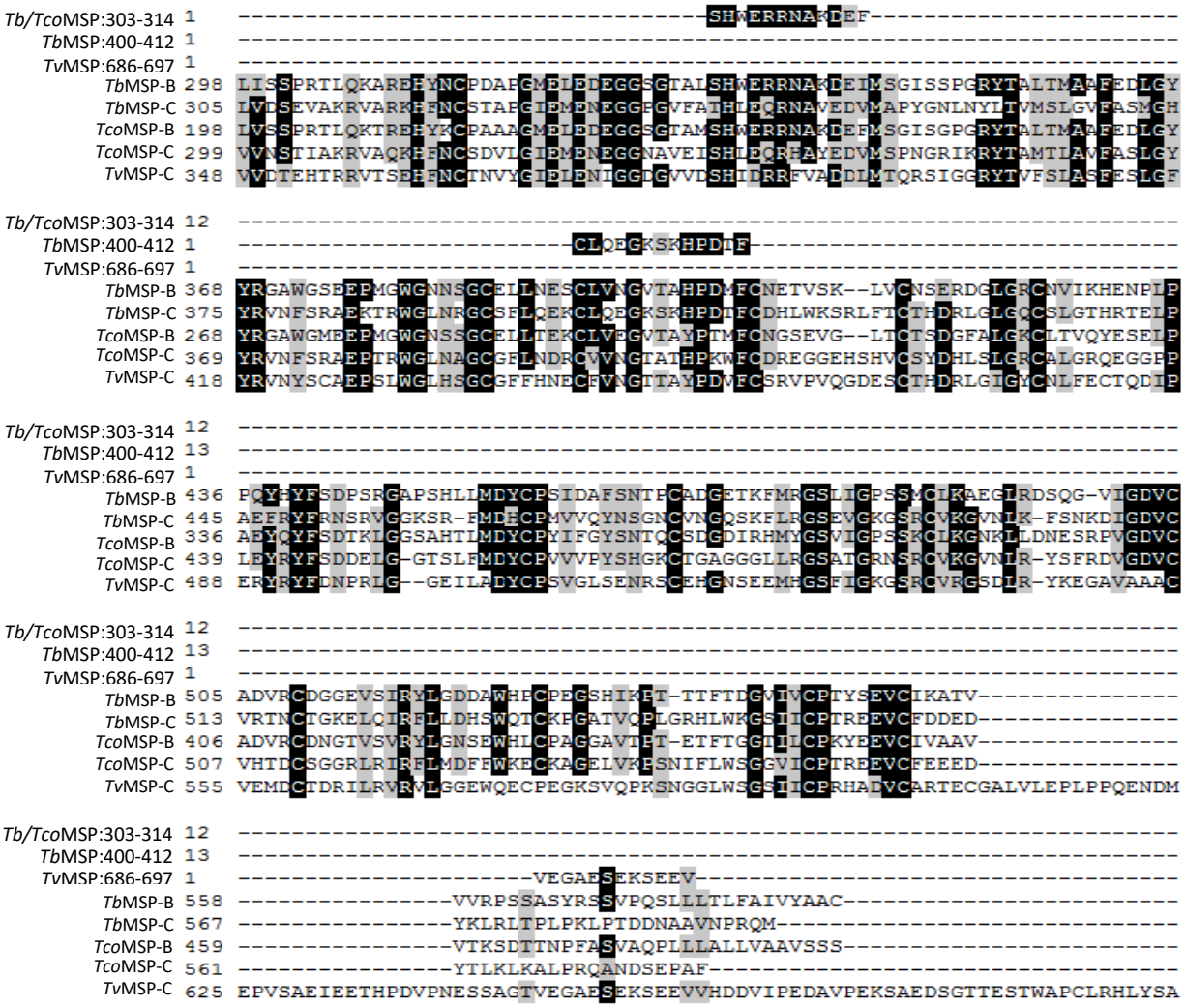


Figure 3.4: Predict7 plots of selected MSP peptides and their predicted positions in a homologous protease. (A) *Tb/Tco*MSP:303-314 is an interspecies peptide designed to detect both *T. congolense* and *T. brucei* MSP-B, (B) *Tb*MSP:400-412 peptide was designed to detect *Tb*MSP-C and (C) *Tv*MSP:686-697 peptide was designed to detect *Tv*MSP-C only. Hydrophilicity (◆), flexibility (■), surface probability (▲) and antigenicity (×) are shown. Arrows indicate the coupling terminus. Three dimensional representation of the predicted position of each peptide within the homolog, leishmanolysin (Schlagenhauf *et al.*, 1998), is shown on the right (yellow).



Sequence	Sequence identities of peptide with MSP		
	*Tb/Tco*MSP:303-314 peptide	*Tb*MSP:400-412 peptide	*Tv*MSP:686-697 peptide
*Tb*MSP-B	**77%**	*54%*	8%
*Tb*MSP-C	42%	**100%**	8%
*Tco*MSP-B1	**100%**	46%	8%
*Tco*MSP-C	42%	38%	0%
*Tv*MSP-C	42%	38%	**100%**

Figure 3.5: Sequence alignment of each peptide with full length MSP-B and –C from *T. brucei (Tb), T. congolense (Tco) and T. vivax (Tv)*. The table shows corresponding sequence identity. From the sequence identity, anti-*Tb/Tco*MSP:303-314 peptide antibodies (IgY or scFv) should detect both *Tb*MSP-B and *Tco*MSP-B; anti-*Tb*MSP:400-412 should detect only *Tb*MSP-C and anti-*Tv*MSP:686-697 should detect only *Tv*MSP-C.

3.3.2 Chicken Anti-Peptide Antibody Production

The peptides selected were used to produce antibodies in chickens. Since the peptide is only 12-14 residues long, it is too small to induce an immune response on its own which means it needs to be conjugated to a carrier protein. The carrier protein used here was rabbit serum albumin (RSA) because of its relatively large size (69 kDa) and its evolutionary distance from chickens (using a mammalian protein not an avian protein). MBS was used as the linker molecule via a lysine on RSA and a cysteine residue in the peptide.

For peptide *Tb/Tco*MSP:303-314, the flexibility is below zero from residue 303-307 which means it was conjugated was done to the N-terminus leaving the more immunogenic C-terminus available for immune system reaction (Figure 3.4, arrow). A cysteine residue was added to the N-terminus of the peptide to allow for conjugation. For *Tb*MSP:400-412 it was decided to conjugate to the N-terminus because there is a cysteine residue already available while the most immunogenic sequences are in the middle of the plot (Figure 3.4, arrow). Lastly, for *Tv*MSP:686-697, a cysteine was added to the C-terminus for conjugation; since, flexibility is lowest at this point (Figure 3.4, arrow).

To determine when antibody production peaks, a single representative egg from each chicken was collected each week and used. Total IgY purified from each egg and an enzyme-linked immunosorbent assay (ELISA) with the peptide antigen coated ELISA plate was conducted (Figure 3.6). Antibody production for all three peptides peaked from week 4 to 12 and the eggs that were laid in this peak period were used to batch purify IgY antibodies (Figure 3.6 grey area). The 200 µg/ml total IgY antibody gave higher signal compared to the 100 µg/ml, as expected.

Large scale IgY purification was done for eggs collected from weeks 4-7 and 8-12. Affinity chromatography was used to purify specific anti-peptide antibodies from total IgY. The eluted antibody fractions were analysed using absorbance at 280 nm (Figure 3.7). Affinity purified anti-*Tb/Tco*MSP:303-314 peptide antibodies produced and purified from eggs from both chickens to give a final yield of ~8.3 mg (calculated using the Beer-Lambert law; Figure 3.7). For the affinity purified anti-*Tb*MSP:400-412 peptide antibodies, four main absorbance peaks were seen for the two chickens used (Figure 3.7). The final yield of this antibody was ~7.4 mg. Affinity purified anti-*Tv*MSP:686-697 peptide antibodies gave only one prominent peak for chicken one (Figure 3.7). The final yield of this purified antibody was ~1.7 mg.

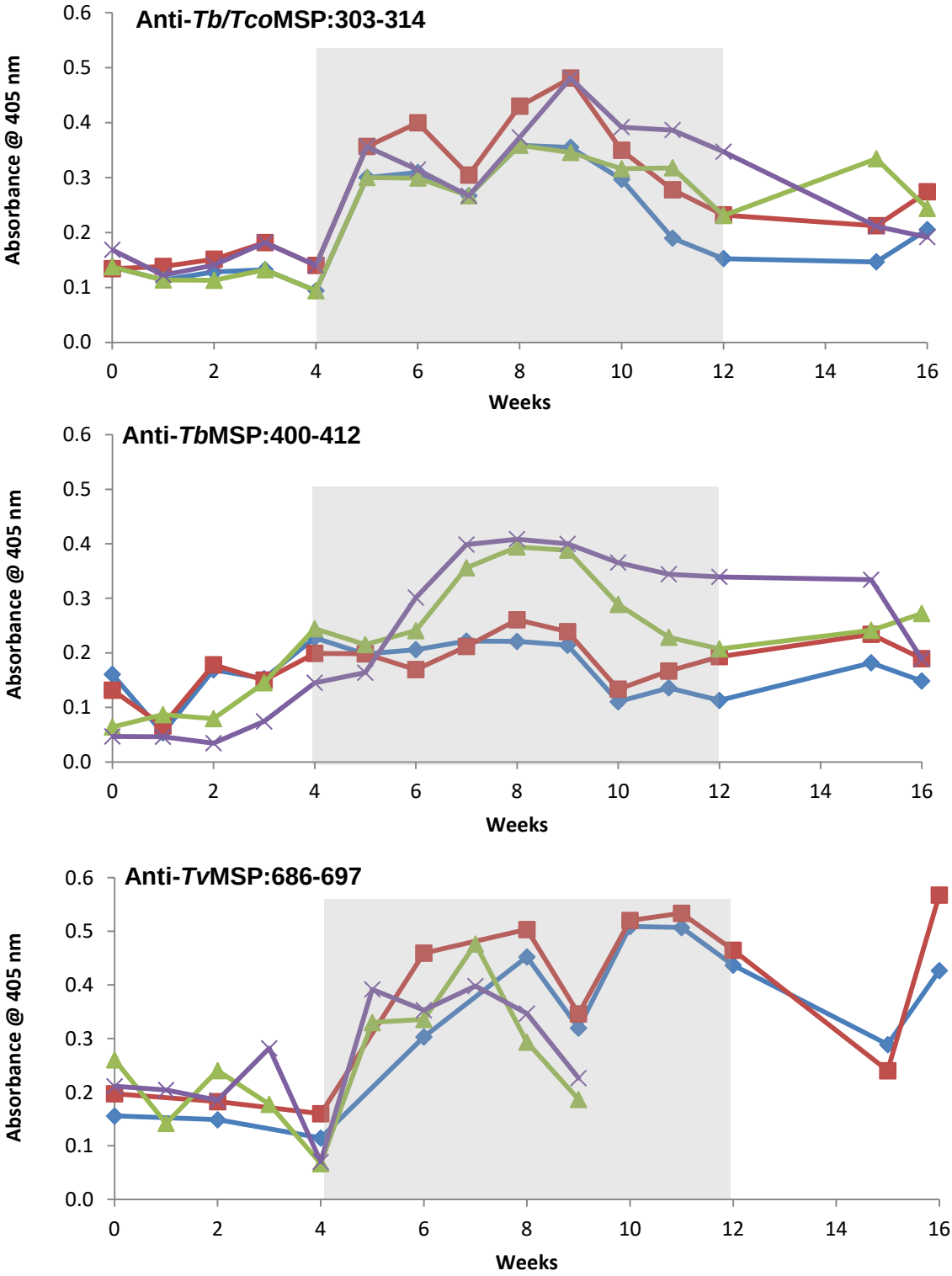


Figure 3.6: Analysis of antibody production in chickens against *Tb/Tco*MSP:303-314, *Tb*MSP:400-412 and *Tv*MSP:686-697 peptides by ELISA. Individual microtitre plates were coated with 1 µg/ml (*Tb/Tco*MSP:303-314 and *Tb*MSP:400-412) or 5 µg/ml (*Tv*MSP:686-697) of peptide. The primary antibody was total IgY isolated from single representative eggs collected from weeks 0-16 at the concentrations indicated. Immunisations were conducted at week 0, 2, 4 and 6. The secondary antibody was rabbit anti-chicken-conjugated to HRPO and the substrate ABTS/H_2O_2. Absorbance readings were recorded at 405 nm using the Optima spectro-fluorometer after 1 h corrected by subtracting negative controls. Grey represents peak selected for batch purification**.** Chicken 1: 100 µg/ml (◆), chicken 1: 200 µg/ml (■), chicken 2: 100 µg/ml (▲), chicken 2: 200 µg/ml (×) are shown.

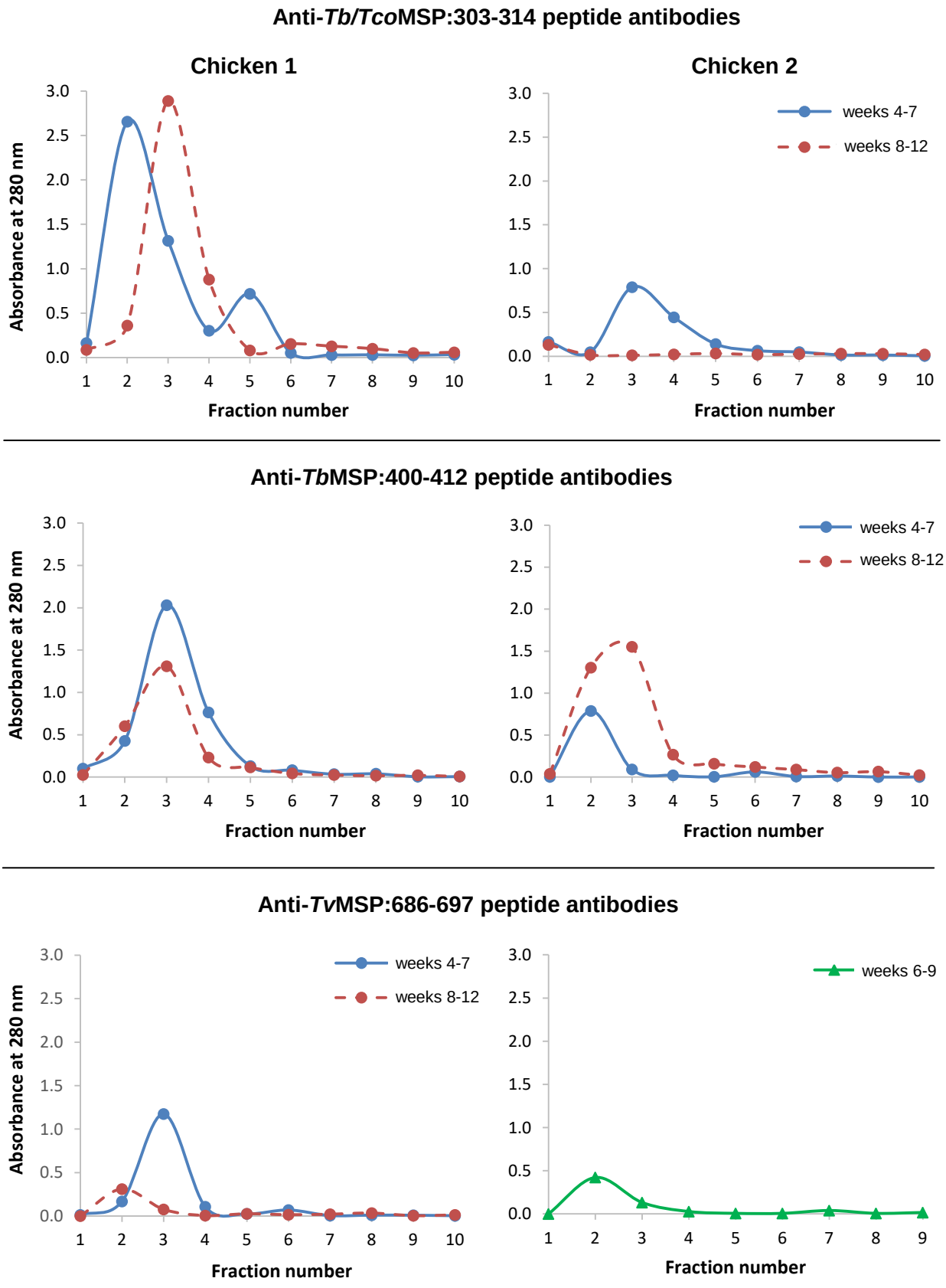


Figure 3.7: Elution profiles of anti-peptide antibodies from peptide-SulfoLink® affinity columns. Total IgY isolated from eggs collected during weeks 4-7 and 8-12 (unless otherwise indicated) was used to affinity purify antibodies on SulfoLink® resin coupled with specific peptide. Bound antibodies were eluted by decreasing the pH and 10 x 1 ml fractions were collected. Absorbance was read at 280 nm using the Nanodrop spectrophotometer.

To confirm that affinity purified anti-peptide antibodies detect the corresponding peptide, ELISAs were conducted (Figures 3.8-3.10). Purified antibody was compared to non-immune control, total IgY and non-specific antibodies that did not bind to the affinity column. All non-immune IgY fractions gave low absorbance values in the ELISA (Figures 3.8-3.10) and these were comparable to the signal when only PBS was used (results not shown).

Affinity purified Anti-*Tb/Tco*MSP:303-314 peptide antibodies gave higher absorbance values than total IgY before affinity purification and IgY that did not interact with the affinity column (Figure 3.8). There was little difference in the absorbance values for the different antibody concentrations used in the assay [(25-100 µ/ml); (Figure 3.8)], suggesting a wider range of antibody concentrations in the lower range (0.01-10 µg/ml) should have been tested. These results showed that specific anti-*Tb/Tco*MSP:303-314 peptide antibodies bound the column and were successfully purified. Colourimetric development of the signals for affinity purified antibodies was compared to total IgY antibodies over time. From Figure 3.8, absorbance after 15 min of total and purified antibody remained relatively constant; however, from 30-60 min the absorbances for purified antibody increased while total IgY stayed relatively the same. This is indicative of purified antibody binding antigen in a more efficient manner as compared to the total, non-affinity purified antibody mixture.

The purification of anti-*Tb*MSP:400-412 peptide antibodies (Figure 3.9) shows a similar trend to the purification of anti-*Tb/Tco*MSP:303-314 (Figure 3.8). A higher absorbance signal was obtained for the affinity purified antibodies when compared to the total IgY and IgY that did not interact with the column (Figure 3.9). Again, no titration of this antibody occurred the 25-100 µg/ml concentration range suggesting a prozone effect (Figure 3.9). Colourimetric development is more efficient by the purified antibodies when compared to total IgY (Figure 3.9).

Lastly, the purification process and antibody efficiency was studied for anti-*Tv*MSP:686-697 peptide antibodies (Figure 3.10). The affinity purified antibodies showed little difference in signal when compared to total IgY but higher signal when compared to the IgY that did not bind the column (Figure 3.10). Again, little difference in the absorbance values for the different antibody concentrations (25-100 µg/ml) was seen. Colourimetric development of purified antibody is similar to total IgY over time (Figure 8.10), indicating similar binding efficiency.

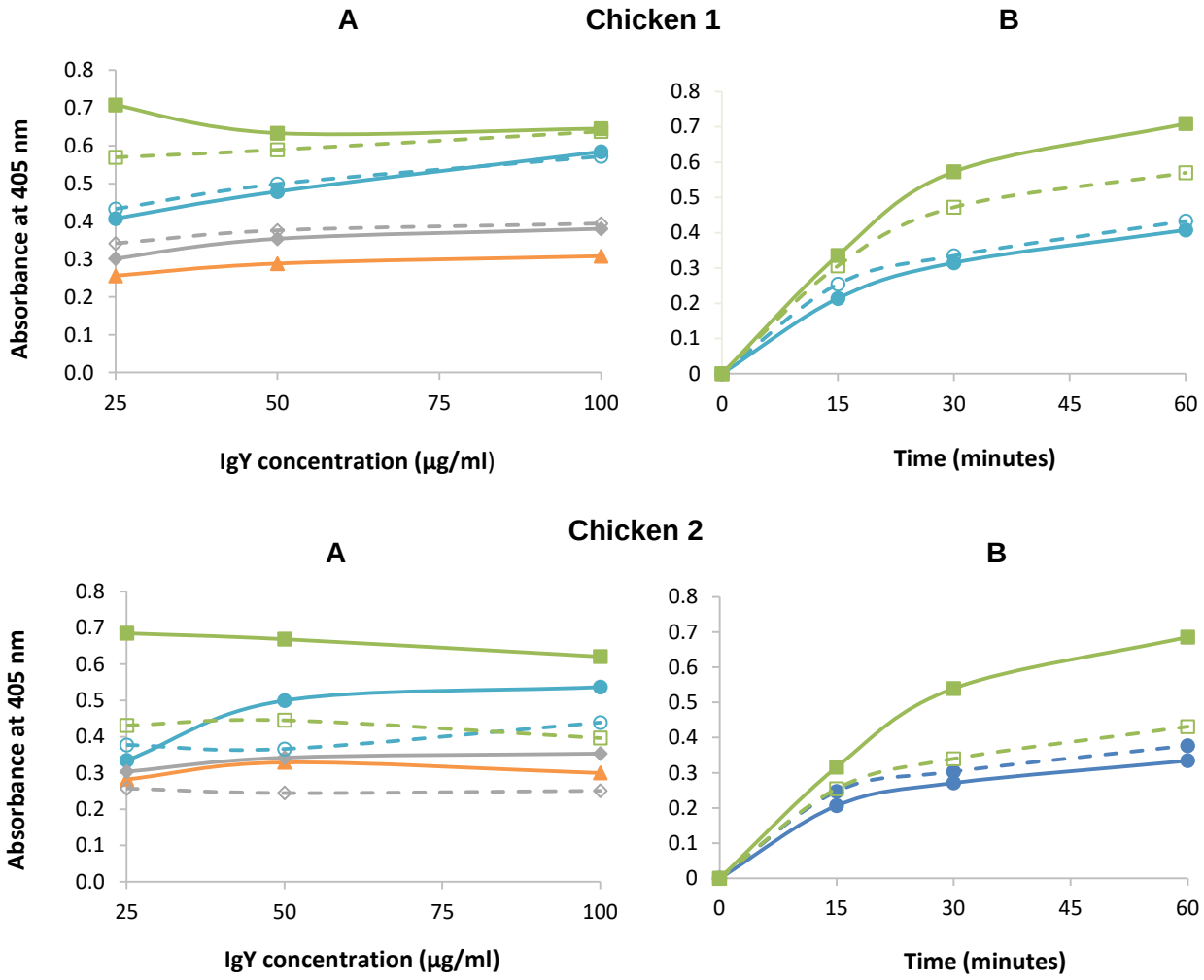


Figure 3.8: Recognition of *Tb/Tco*MSP:303-314 peptide by corresponding affinity purified chicken anti-peptide antibody in an ELISA. IgY isolated from eggs collected during weeks 4-7 (solid line) and weeks 8-12 (dotted line). Microtitre plates were coated with 5 µg/ml peptide and incubated with primary antibodies: non-immune control (▲), total IgY (●;○), affinity column unbound fraction (◆; ◇) and purified antibody (■;□), for 2 h. The secondary antibody was rabbit anti-chicken-conjugated to HRPO and the substrate ABTS/H_2O_2. **A)** Increasing primary antibody concentration (25-100 µg/ml) after 1 h in substrate and **B)** colourimetric development (25 µg/ml primary antibody) after 15, 30 and 60 min in substrate. Absorbance readings were recorded at 405 nm using an Optima spectro-fluorometer. Similar absorbance values to those of non-immune control were recorded when the primary antibody was replaced with PBS (not shown).

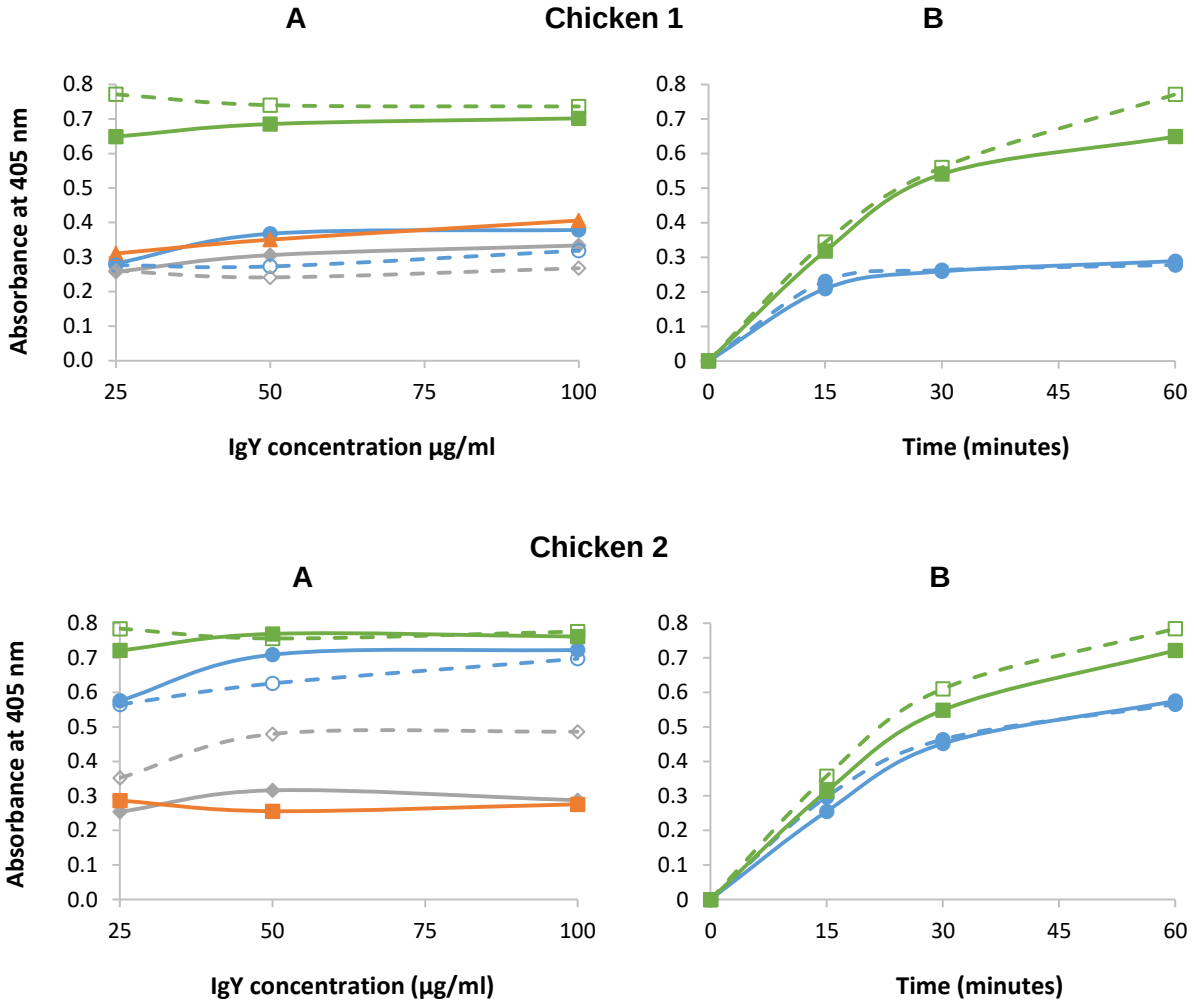


Figure 3.9: Recognition of *Tb*MSP:400-412 peptide by corresponding affinity purified chicken anti-peptide antibody in an ELISA. IgY isolated from eggs collected during weeks 4-7 (solid line) and weeks 8-12 (dotted line). Microtitre plates were coated with 5 µg/ml peptide and incubated with primary antibodies: non-immune control (▲), total IgY (●;○), affinity column unbound fraction (◆;◇) and purified antibody (■;□), for 2 h. The secondary antibody was rabbit anti-chicken-conjugated to HRPO and the substrate ABTS/H_2O_2. **A)** Increasing primary antibody concentration (25-100 µg/ml) after 1 h in substrate and **B)** colourimetric development (25 µg/ml primary antibody) after 15, 30 and 60 min in substrate. Absorbance readings were recorded at 405 nm using an Optima spectro-fluorometer. Similar absorbance values to those of non-immune control were recorded when the primary antibody was replaced with PBS (not shown).

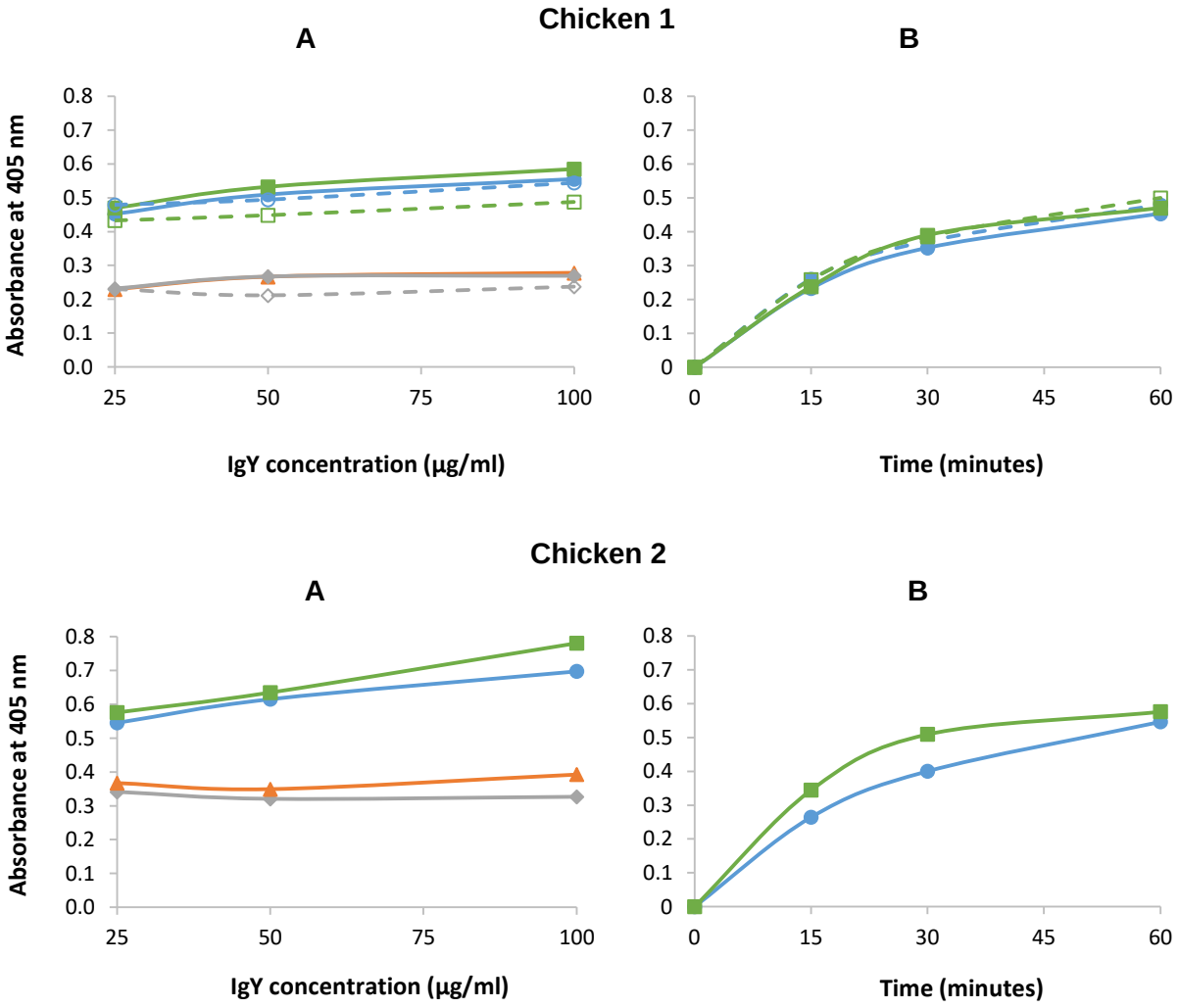


Figure 3.10: Recognition of *Tv*MSP:686-697 peptide by corresponding affinity purified chicken anti-peptide antibody in an ELISA. For chicken 1, IgY isolated from eggs collected during weeks 4-7 (solid line) and weeks 8-12 (dotted line). For chicken 2, IgY was isolated for weeks 6-9 (solid line). Microtitre plates were coated with 5 µg/ml peptide and incubated with primary antibodies: non-immune control (▲), total IgY (●;○), affinity column unbound fraction (◆;◇) and specific purified antibody (■;□), for 2 h. The secondary antibody was rabbit anti-chicken-conjugated to HRPO and the substrate ABTS/H_2O_2. **A)** Increasing primary antibody concentration (25-100 µg/ml) after 1 h in substrate and **B)** colourimetric development (25 µg/ml primary antibody) after 15, 30 and 60 min in substrate. Absorbance readings were recorded at 405 nm using an Optima spectro-fluorometer. Similar absorbance values to those of non-immune control were recorded when the primary antibody was replaced with PBS (not shown).

In all cases, peptide-specific antibodies were successfully produced by the immunised chickens and purified using affinity chromatography. These antibodies showed efficient binding of the respective antigens. *Tb/Tco*MSP:303-314 (8.3 mg) and *Tb*MSP:400-412 (7.4 mg) produced more antibodies, when compared to *Tv*MSP:686-697 (1.7 mg) and anti-*Tv*MSP:686-697 peptide antibodies showed the lowest sensitivity for detecting the peptide antigen. These purified antibodies were used to detect MSP in parasite lysate and in studies in Chapter 4.

3.3.3 scFv Anti-Peptide Antibody Production

Single chain variable regions (scFvs), from the Nkuku® library (Van Wyngaardt *et al.*, 2004), were also produced in this study. Biopanning was performed under stringent selection and washing conditions to select phagemids expressing scFv specific antibody scaffolds. Once panning was completed, a polyclonal ELISA was conducted on the phagemids isolated from each round as compared to the full phagemid Nkuku® library (Figure 3.11). Overall, an increase in signal is seen, after each round of panning, which shows the enrichment of specific scFvs and the removal of non-specific scFvs. The full Nkuku® library gave low absorbance readings against each MSP peptide as expected. scFvs selected to recognise *Tb/Tco*MSP:303-314 gave peak signals after round one and four; while, *Tb*MSP:400-412 gave a peak signal at round four only. *Tv*MSP:686-697 gave peak signals after round two and four (Figure 3.11).

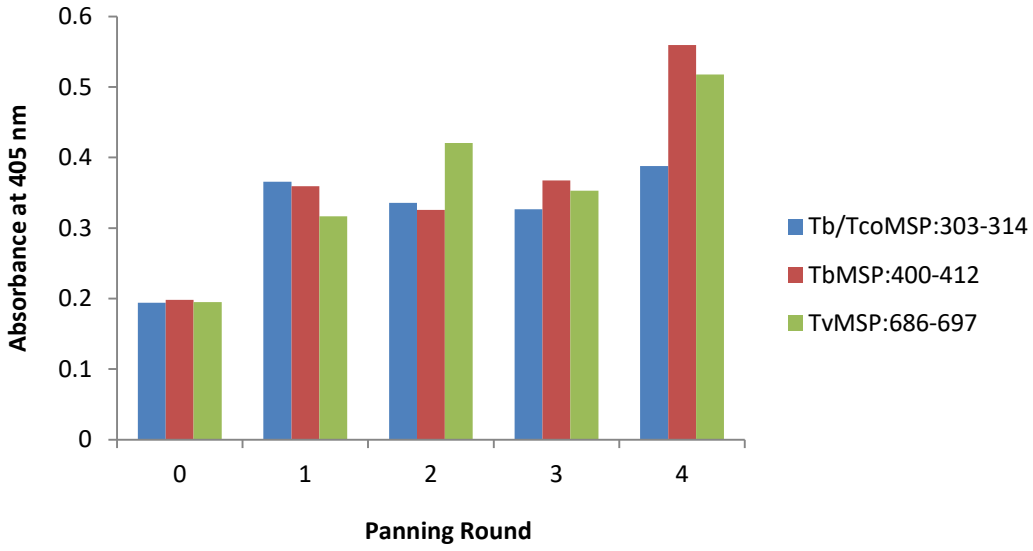


Figure 3.11: Polyclonal phagemid ELISA of panning rounds 1 to 4. Microtitre plates were coated with 10 µg/ml peptide and incubated with 1:10 dilution of phagemid stock isolated after each round of panning. Primary mouse anti-M13 antibody (1:8000) and secondary goat anti-mouse IgG antibody (1:10000) conjugated to HRPO were used. The substrate ABTS/H_2O_2 was added and readings were recorded at 405 nm using an Optima spectro-fluorometer**.**

Since a general trend of increasing signal over each panning round was seen, round four was used for the selection of monocolony scFvs. This is done to select the best signal producing clone for further scFv expression and purification.

Round 4 phagemid, selected for reactivity in ELISA tested against each peptide, was allowed to infect *E. coli* TG1 cells and titered to produce single colonies that can be individually selected for the production of single phenotypic scFvs. A total of 96 single

colonies that contain phagemid enriched to detect *Tb/Tco*MSP:303-314 were selected. After being cultured and rescued, soluble phagemid was allowed to react with antigen in an ELISA (Figure 3.12). A number of individual colonies gave strong signals which correspond to successful development and selection of single scFv phenotype producing clones that are able to detect the antigen. Colonies A1, A12, D2, D12, G1, G12, H1 and H12 gave the highest signal and could have been used to express and purify specific anti-*Tb/Tco*MSP:303-314 peptide scFvs (Figure 3.12).

Similar results were obtained from the 96 colonies selected to recognise *Tb*MSP:400-412 (Figure 3.13). Again, columns 1 and 12 gave the highest overall signal and colonies A1, A12, D5, G1, G12, H1, and H12 gave the highest signal (Figure 3.13). These colonies could have been selected for the expression and purification of phenotypically unique anti-*Tb*MSP:400-412 peptide scFvs.

The 96 colonies selected to detect *Tv*MSP:686-697 gave low absorbance readings (Figure 3.14). Only colonies F1, G1 and H11 gave a slight increase in signal over the others (Figure 3.14). This result was unexpected because in polyclonal environments the anti-*Tv*MSP:686-697 peptide scFvs performed well giving an adequate signal. It is possible that the colonies selected did not express the scFvs that bind *Tv*MSP:686-697 best or individual colonies did not produce enough scFv for detection.

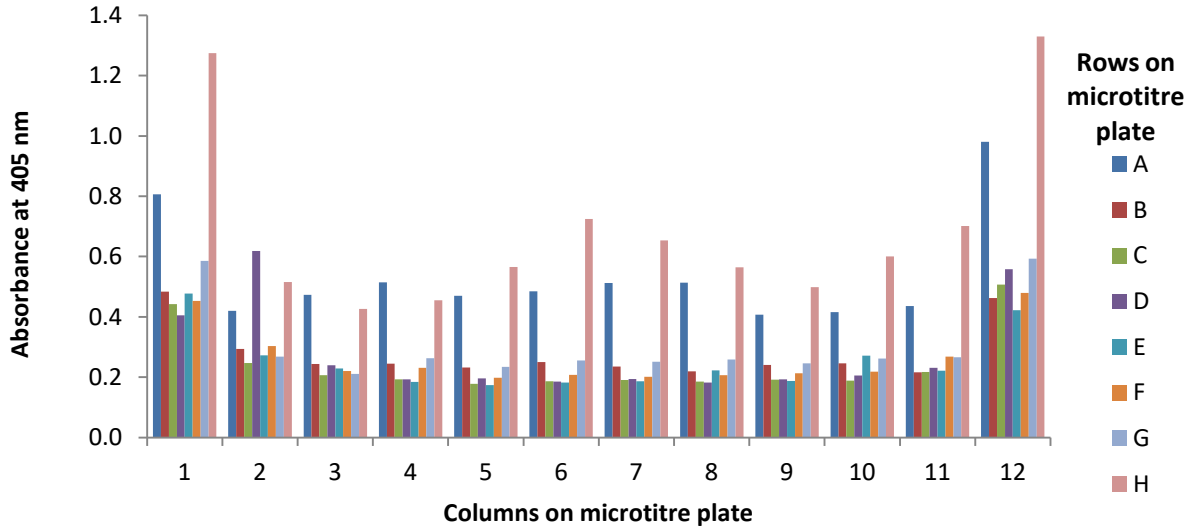


Figure 3.12: Monocolony scFv ELISA on 96 single colonies from round 4 of panning against *Tb/Tco*MSP:303-314. Single infected colonies cultured in a microtitre plate were rescued and phagemid isolated by centrifugation. Phagemid was allowed to react in a microtitre plate coated with 10 µg/ml *Tb/Tco*MSP:303-314. Primary mouse anti-M13 antibody (1:8000) and secondary goat anti-mouse IgG antibody (1:10000) conjugated to HRPO were used. The substrate ABTS/H_2O_2 was added and readings were recorded at 405 nm using an Optima spectro-fluorometer.

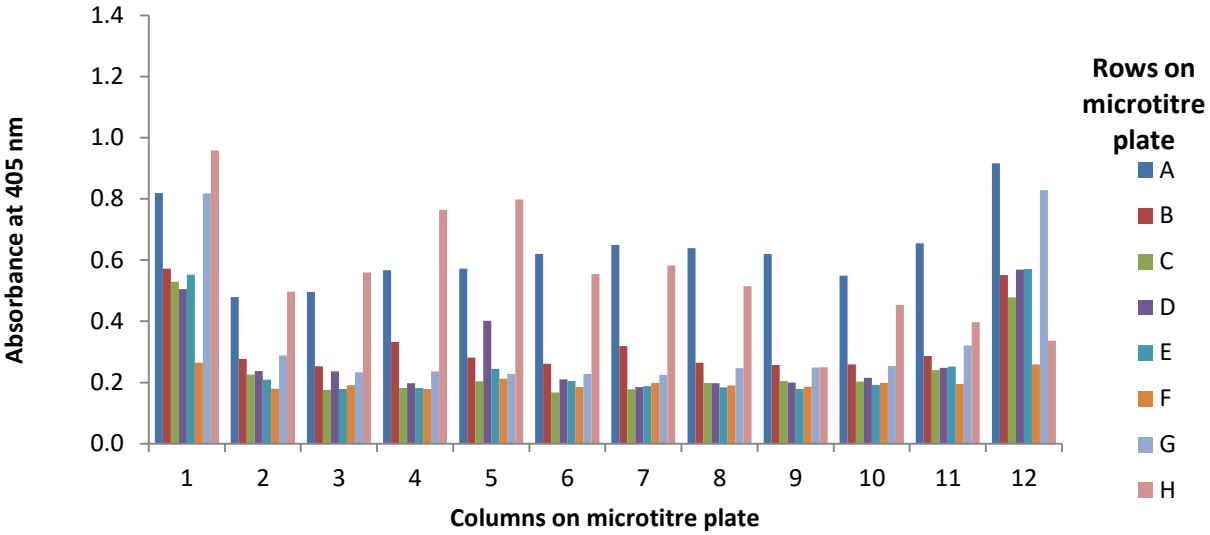


Figure 3.13: Monocolony scFv ELISA on 96 single colonies from round 4 of panning against *Tb*MSP:400-412. Single infected colonies cultured in a microtitre plate were rescued and phagemid isolated by centrifugation. Phagemid was allowed to react in a microtitre plate coated with 10 µg/ml *Tb*MSP:400-412. Primary mouse anti-M13 antibody (1:8000) and secondary goat anti-mouse IgG antibody (1:10000) conjugated to HRPO were used. The substrate ABTS/H_2O_2 was added and readings were recorded at 405 nm using an Optima spectro-fluorometer.

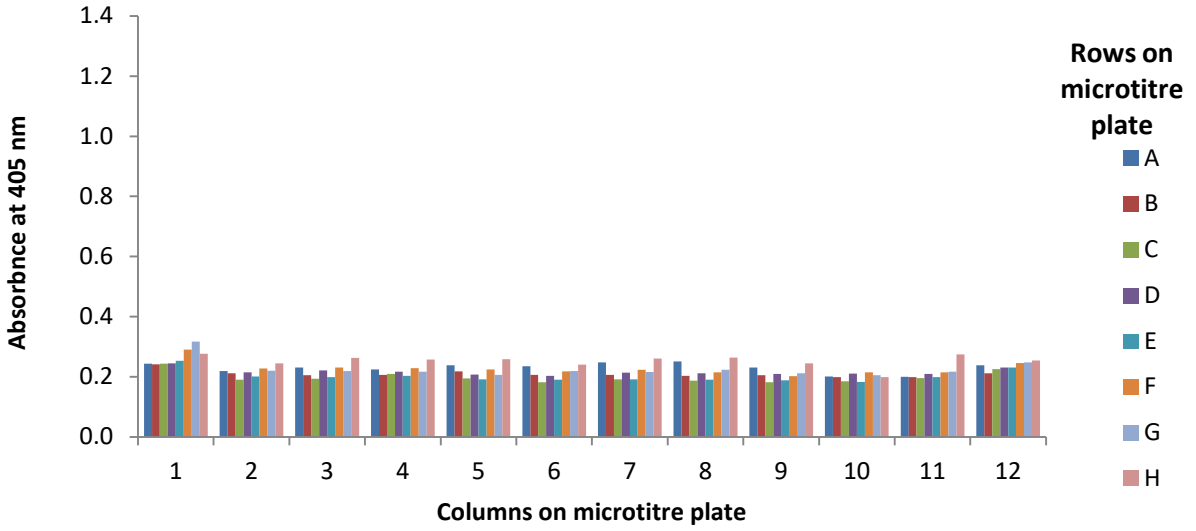


Figure 3.14: Monocolony scFv ELISA on 96 single colonies from round 4 of panning against *Tv*MSP:686-697. Single infected colonies cultured in a microtitre plate were rescued and phagemid isolated by centrifugation. Phagemid was allowed to react in a microtitre plate coated with 10 µg/ml MSP: 303-314. Primary mouse anti-M13 antibody (1:8000) and secondary goat anti-mouse IgG antibody (1:10000) conjugated to HRPO were used. The substrate ABTS/H_2O_2 was added and readings were recorded at 405 nm using an Optima spectro-fluorometer.

From the Nkuku® library, scFv phagemids were selectively enriched to detect the three peptides designed from the amino acid sequences of MSP-B and -C. The successful selection of individual colonies which would allow for the expression of specific scFv antibody scaffolds was obtained for two (*Tb/Tco*MSP:303-314 and *Tb*MSP:400-412) of the three peptides.

3.3.4 Detection of MSP in Parasite Lysate

The detection ability of the respective anti-peptide antibodies to detect proteins in blood stage form (BSF) parasite lysates of both *T. brucei* and *T. congolense* were assessed. Anti-*Tb/Tco*MSP:303-314 peptide antibody was able to detect a 63 kDa protein found in both species of *Trypanosoma* (Figure 3.15, panel B). Since this antibody was designed to detect MSP-B, from *T. brucei* and *T. congolense,* and the theoretical size of MSP-B is on average 63 kDa, it is evident that the protein detected here may be native MSP-B.

Detection of a 53 kDa protein by anti-*Tb*MSP:400-412 peptide antibody in lysate of both parasite species is also present (Figure 3.15, Panel C). Since, this peptide was designed to detect *Tb*MSP-C; the detection of a protein in the *T. congolense* lysate was unexpected. The size of this protein was smaller than the expected size of *Tb*MSP-B of 64 kDa, and that of *Tb*MSP-C of 66 kDa. Since these proteases have a proregion, which have to be removed for the protease to be active, it is suspected that the smaller size represents the active form of MSP. Anti-*Tb*MSP:400-412 showed no detection of the larger 63 kDa protein expected to be MSP-B. Therefore, it is suspected that the protein of 53 kDa is MSP-C. Furthermore, *Tb*MSP:400-412 shares a 100% identity with *Tb*MSP-C while only 53% identity with *Tb*MSP-B. The detection of 53 kDa protein in the *T. congolense* lysate was also observed, even though, the *Tb*MSP:400-412 peptide only shares 38% identity with *Tco*MSP-C. Anti-*Tb*MSP:400-412 also detected smaller unknown products of 38 kDa in both parasite lysates (Figure 3.15, Panel C, lane 1 and 2) and a 27 kDa protein in the *T. brucei* lysate. It could be possible that the 38 kDa protein is an active form of MSP while the smaller 27 kDa protein is the proregion of the protein (38 + 27 = 65 kDa). However, at residues 400-412, the peptide selected for antibody production is not present within the predicted proregion although, the predicted proregion is based on leishmanolysin and variations may occur.

Anti-*Tv*MSP:686-697 showed no detection of any protein in either parasite lysates (Figure 3.15, Panel D). *Tv*MSP:686-697 shares less than 8% identity with MSP-B and -C from *T. brucei* and *T. congolense* which means detection of these proteins was unlikely.

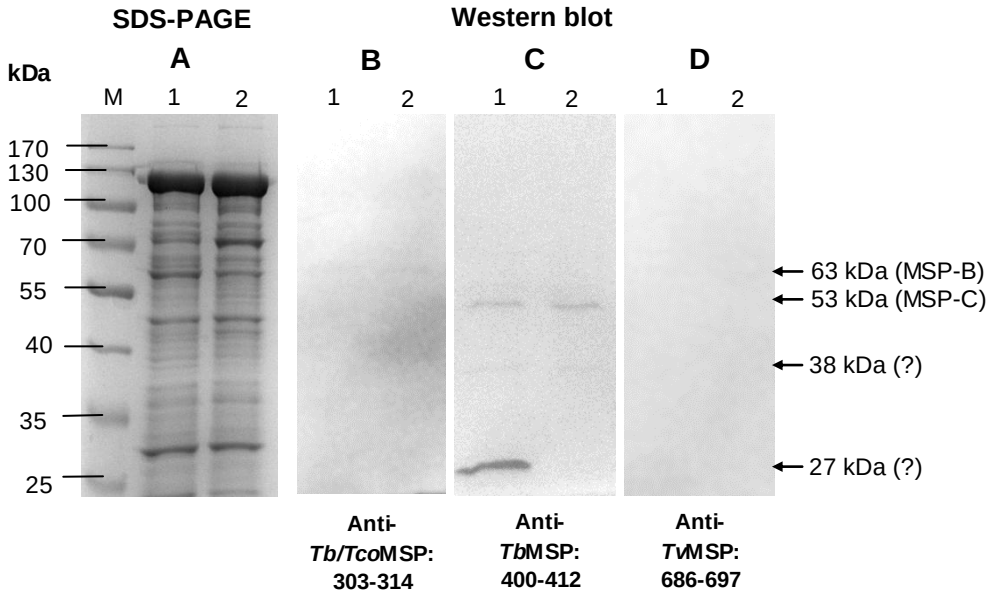


Figure 3.15: Detection of MSPs in parasite lysate, using anti-peptide antibodies. (A) 10% SDS-PAGE gel and western blots with **(B)** anti-*Tb/Tco*MSP:303-314, **(C)** anti-*Tb*MSP:400-412 or **(D)** anti-*Tv*MSP:686-697 peptide antibody (5 µg/ml) as indicated. Rabbit anti-chicken secondary antibody conjugated to HRPO (1:2500) and visualised by ECL luminescence using the syngene G:BOX. Lane M, ThermoScientific pre-stained marker; lane 1, BSF *T. brucei* parasite lysate; lane 2, BSF *T. congolense* parasite lysate. Arrows indicate detection of 63 kDa MSP-B, 53 kDa active MSP-C and 38 kDa and 27 kDa unknown products (?).

3.4 Discussion

To allow for the detection of specific MSPs and also assess the possible use of these proteases as diagnostic targets, peptides were designed that either detect the MSPs from more than one species or select a specific homologue of MSP. The results shown here indicated that these peptides can induce the production of anti-peptide detection molecules using both chickens (full antibody) and Nkuku® phagemid library (scFvs). Both methods gave successful production of detection molecules with the ability to detect the peptide antigens. Furthermore, native MSP was detected in parasite lysate using chicken anti-peptide antibodies.

In general, high homology in primary structure results in the production of antibodies with the ability to cross-react with different antigens (Aalberse *et al.*, 2001). An antibody's paratope bound to an antigen epitope covers about 15 amino acid residues (Benjamin and Perdue, 1996; Dougan *et al.*, 1998). It has been shown that five key residues can be responsible for the binding energy of the epitope (Benjamin and Perdue, 1996; Dougan *et al.*, 1998). This means that only 33% of an epitope region is responsible for recognition by the antibody. For this reason, peptides were designed based on both their predicted immunogenicity and their sequence identities with the full length proteases.

Sequence identity of over 50% between peptide and protein sequences was suspected to produce anti-peptide antibodies that have the ability to cross-react. *Tb/Tco*MSP:303-314 showed high sequence identity with both *Tco*MSP-B (100%) and *Tb*MSP-B (77%) and was chosen to produce antibodies to detect both species. The other two peptides were designed to detect specific species. *Tb*MSP:400-412 was designed to produce antibodies to detect only *T. brucei.* However, due the moderate identity between the peptide and *Tco*MSP-B and -C (46% and 38%), anti-*Tb*MSP:400-412 peptide antibodies may also be able to detect *T. congolense* (peptide may contain the important amino acids required for binding). *Tv*MSP:686-697 shows little sequence identity with any other protease species other than *T. vivax* and antibodies produced against this peptide should detect this species' MSP-C.

Phage display was used to isolate colonies expressing scFvs with the ability to detect *Tb/Tco*MSP:303-314 and *Tb*MSP:400-412, only. The *Tv*MSP:686-697 polyclonal ELISA results also gave a high signal after round two. Similarly, the scFvs selected against *Tb/Tco*MSP:303-314 also gave a signal after round one. This suggests that the original assumption that as selection rounds for specific scFvs continue the more specific the scFvs become and the higher the signal will be (Griep *et al.*, 2000; Van Wyngaardt *et al.*, 2004), may be incorrect. Therefore, to have the best monocolony scFv that gives the highest signal, clones should be selected from each round.

A nanobody library against *T. congolense* was recently developed that is able, *in vitro*, to detect experimental infection and can be used as a test of cure (Odongo *et al.*, 2016). Furthermore, an scFv library panned against the p24 HIV capsid protein resulted in highly specific antibody fragments which may be used in the diagnosis of viral load in HIV patients (de Haard *et al.*, 1998) and scFvs raised against Infectious Bursal Disease Virus (IBDV) antigens were reported to be able to distinguish between three strains of IBDV and even neutralise the virus (Sapats *et al.*, 2003). These reports of the use of a phagemid library in the diagnosis of *Trypanosoma* and other diseases highlight the benefits of this technology in diagnostics and even treatment.

The chicken anti-MSP peptide antibodies showed cross-reactivity during the detection of native MSP in parasite lysate. These proteases show different levels of conservation across species possibly resulting in this cross-reactivity. The anti-*Tb/Tco*MSP:303-314 peptide antibody was expected to cross-react. However, unexpectedly anti-*Tb*MSP:400-412 cross reacted with *Tco*MSP-C which only shared 38% identity with the peptide. This was, nevertheless, the first evidence of native MSP-C within *Trypanosoma* lysate.

Anti-*Tv*MSP:686-697 showed no cross reactivity being unable to detect MSP in parasite lysates. Either the design of the peptide needs to be reassessed to ensure identity is 10% or less or these proteases would have more potential as a pan species detection system.

The work described in this chapter showed that peptide epitopes were successfully used to produce antibody detection molecules. The specificity of the chicken anti-peptide antibodies was further assessed using *T. brucei* and *T. congolense* parasite lysate and a high level of cross-reactivity between the two species was seen. It is suggested that these MSPs can be used as a novel diagnostic marker for cross-trypanosomal species detection. These antibodies were used to further characterise recombinant MSP-C from both *T. brucei* and *T. congolense* as described in Chapter 4.

Chapter 4

Cloning of *T. brucei* MSP-C and Expression and Characterisation of *T. brucei* and *T. congolense* MSP-C

4.1 Introduction

Proteases within parasitic protozoa have been identified to play an important role in the life cycle of these organisms. Identifying different proteases and understanding the way they function have given insight into approaches for vaccine development and diagnostics for parasitic diseases (Klemba and Goldberg, 2002). The identification of the papain-like cysteine proteases and serine proteases within trypanosome species is a case in point (Grellier *et al.*, 2001; Lalmanach *et al.*, 2002; Bastos *et al.*, 2013). Some of these proteases are pathogenic factors which break down host proteins and contribute to the manifestation of disease symptoms (Bastos *et al.*, 2010).

The Major Surface Proteases (MSPs), novel metalloproteases which reside on the surface of parasites, have been identified in the kinetoplastid *Leishmania* (Etges, 1992). It functions to help these parasites to evade the host's immune system, invade host macrophages and help in the survival of parasites within these organelles (Wilson and Hardin, 1988; Brittingham *et al.*, 1995; Kulkarni *et al.*, 2006; Matte *et al.*, 2016). A number of MSPs have been identified in *Trypanosoma* species as well, including *T. brucei* and *T. congolense*. Since *Leishmania* is an intracellular parasite, some of the functions of these proteases do not apply to the African *Trypanosoma* homologues. These trypanosome populations have a highly evolved mechanism of evading the host immune system, which involves changing their VSG coat in a process called antigenic variation (Pays *et al.*, 2004). These parasites are also extracellular and, therefore, they do not invade any host organelles. So the question arises: why do trypanosomes express these proteases and what are their functions?

T brucei and *T. congolense* have several forms of these proteases (El-Sayed and Donelson, 1997a; Marcoux *et al.*, 2010). Even though little is known regarding their function, it is suspected that each class has unique functions due to their differential expression within the parasite life cycle (LaCount *et al.*, 2003). It has been shown that MSP-C is expressed in the blood stage form of *T brucei* parasites (LaCount *et al.*, 2003). It has also been predicted that *Tco*MSP-C is unlikely to be GPI-anchored to the surface of the parasite (Marcoux *et al.*, 2010). Studies have indicated that leishmanial species contain forms of the protease that are secreted (Jaffe and Dwyer, 2003). Therefore, it would be foreseeable if *Trypanosoma* species also contained forms of the protease that are secreted. It was deduced that MSP-C might be such a protease. If this was the case, it would be secreted straight into the blood

stream of its host where it may function as a pathogenic factor like the serine proteases prolyl oligopeptidase and oligopeptidase B [POP and OPB; Morty *et al.* (1998); Morty *et al.* (1999); Grellier *et al.* (2001)]. MSPs from *Leishmania* have been shown to cleave collagen and fibronectin (McGwire *et al.*, 2003a), uniformly to *Trypanosoma* POP and OPB which cleaves components of the extracellular matrix as well (Bastos *et al.*, 2013). If this protease is secreted into the host's blood stream, it may be available for detection and be a novel cross-species trypanosomal diagnostic marker. Alternatively, it may induce an immune response which could allow for the detection of serum antibodies as seen for other *Trypanosoma* diagnostic tests (Büscher *et al.*, 2014; Birhanu *et al.*, 2015; Fleming *et al.*, 2016).

Leishmanolysin, MSP from *Leishmania major,* has been shown to catalytically cleave the substrate N-Ala-Tyr↓Leu-Lys-Lys-C (Bouvier *et al.*, 1990; Schlagenhauf *et al.*, 1998). Using Schechter and Berger (1968) schematic layout of peptide substrate and active site interaction: P1 has a 35% chance of being Tyr while P2 can be either Ala and Leu (25% chance). P1′ also has a 35% chance of being Leu and P2′ and P3′ have a 45% chance of being Lys (Schomburg and Stephan, 2012). It has been shown in the analyses reported in this dissertation that each MSP contains key characteristics which would indicate that they have similar substrate specificities (Chapters 1 and 2). It was, therefore, viable to use the leishmanolysin substrate specificity and the possibility percentages of each residue within each P-site to select a peptide substrate for MSP-C cleavage.

In the work reported in this chapter, *T. brucei* LiTat 1.3 parasites were successfully purified and genomic DNA extracted. PCR was performed using primers designed to amplify *TbMSP-C* which was subsequently cloned into a pGEM®-T vector and subcloned into a pGEX 4T-1 expression vector. The *T. brucei* MSP-C protease was recombinantly expressed in *E. coli* and the anti-peptide antibody (Chapter 3) detection specificity assessed. However, due to significant mutations within the cloned insert and presence of a number of truncated products during expression, *TbMSP-C* and *TcoMSP-C* gene constructs were synthesised and respective recombinant protein successfully purified. These enzymes were also used to assess the detection specificity of the anti-peptide antibodies. Furthermore, they were characterised under non-denaturing conditions, analysed enzymatically and assessed for detection using infected cattle sera.

4.2 Materials and Methods

4.2.1 Materials

All general reagents were of the highest purity and purchased from Merck (RSA).

Molecular Biology: The pGEM® T-vector was purchased from Promega (USA) and the pGEX 4T-1 expression vector from Sigma-Aldrich (Germany). RNase, Dream™ Taq Enzyme Mix, High Fidelity PCR Enzyme Mix, dATP, dTTP, dGTP, dCTP, T4 DNA ligase Enzyme Mix, TransformAid™ cloning kit, GeneJet plasmid DNA isolation kit, EcoRI, NotI, 10 x 'O' buffer, Fast Alkaline Phosphatase (AP) enzyme mix, isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) were all from Life Technologies/ThermoScientific (Gauteng, RSA). Ethidium bromide was from Boehringer Mannheim/Roche (USA). O'GeneRuler 1 kb DNA Ladder and 6 x orange DNA loading dye were from Separation Scientific (USA). peqGOLD gel extraction kit was from PEQLAB biotechnologie (Germany) and Clean and Concentrate DNA Kit was purchased from ZymoResearch (USA).

***E. coli* cell lines:** *Escherichia coli (E. coli)* strains JM109 and BL21 (DE3) were purchased from New England Biolabs (USA).

Constructs: Two MSP-C (*T. brucei and T. congolense*) constructs within the pET-100D vector were synthesised by GeneArt (ThermoFisher, Germany) with codon optimisation for expression in *E. coli*. The expressed proteins were referred to as r*Tb*MSP-C and r*Tco*MSP-C

Expression: Lysozyme was from ThermoFisher (USA).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analyses: H_2O_2 and 2-mercaptoethanol were purchased from Merck (RSA) and ammonium persulfate from BioRad (USA). Rabbit anti-chicken IgY horse radish peroxidase (HRPO) and 4-chloro-1-naphthol were purchased from Sigma-Aldrich (USA). PageRuler pre-stained marker was purchased from Separation Scientific (USA) and BioTrace™ nitrocellulose was from PALL Life Sciences (USA). ECL reagent was from Pierce (USA). Anti-GST antibodies were made previously in chickens.

Purification: UltraLink® Hydrazide Resin was purchased from Separation Scientific (USA) and the Zeba™ desalting spin column was from Pierce (USA). Centricon® centrifugal concentrators were from Merck Millipore (Germany). *meta*-Periodate was purchased from Merck (RSA).

Protease activity: 1,10 phenanthroline and porcine gelatin were purchased from Sigma-Aldrich (USA). H-Suc-Leu-Tyr-AMC peptide substrate was purchased from Bachem (Switzerland) and Nunc® Black 96 well plates were from ThermoFisher (USA).

ELISA: Sera from infected (+28) and non-infected cattle (-7) were obtained from ClinVet International (PTY) LTD (Bloemfontein, South Africa). Rabbit anti-bovine IgG-HRPO (A8917) conjugate was purchased from Sigma Aldrich (USA).

4.2.2 Parasite Purification

This was done as described in Section 3.2.7 (Chapter 3).

4.2.3 Genomic DNA Isolation

T. brucei genomic DNA was isolated from both the procyclic Lister 427 and blood stage LiTat 1.3 strains. The purified parasites were pelleted by centrifugation (3000 x g, 10 min, 4^{o}C). The parasites were washed in 10 ml PBS and centrifuged again (3000 x g, 10 min, 4^{o}C). The supernatant was discarded and the pelleted parasites were gently resuspended in 2 x 150 µl TELT lysis buffer (50 mM Tris-HCl, pH 8 containing 62.5 mM EDTA and 2.5 mM LiCl). The resuspended pellet was incubated at room temperature (RT) for 5 min. Phenol:chloroform (ratio 1:1) was added (2 x 150 µl) and incubated for another 5 min at RT.

The sample was centrifuged (17000 x g, 10 min, 4^{o}C) to separate the solution into three distinct phases. The top aqueous phase, containing nucleic acids, was removed and the volume determined. To the top aqueous phase an equal volume of ice cold 100% (v/v) ethanol was added. This solution was incubated overnight at -20^{o}C. The precipitated nucleic acid was pelleted by centrifugation (17000 x g, 15 min, 4^{o}C), washed with 500 µl 100% (v/v) ethanol and dried by incubation at 37^{o}C for 10 min. The pellet was dissolved in 100 µl TE buffer (10 mM Tris-HCl -HCl, 1 mM EDTA, pH 8.0) containing 1 µg/ml RNase and incubated at 37^{o}C for 45 min to allow for the degradation of RNA. The genomic DNA sample was analysed for purity on a 1% (w/v) agarose gel containing 0.5 µg/ml ethidium bromide in 1 x TAE buffer (40 mM Tris-HCl , 20 mM acetate, 1 mM EDTA) and concentration determined using the Nanodrop 6000 spectrophotometer. The final purified genomic DNA from both *T. brucei* strains was stored at -20^{o}C until needed.

4.2.4 Primer Design and Polymerase Chain Reaction (PCR) of *TbMSP-C*

The nucleotide sequence coding for *T. brucei MSP-C* was obtained from NCBI (accession number: AAO72980.1) and used to design the primers shown in Table 4.1.

The primers were designed to amplify the full length 1773 bp *TbMSP-C* gene and to introduce the EcoRI [ECoRI (Roberts *et al.*, 2003)] and NotI site within the amplified product. The primers were synthesised by the University of Cape Town (South Africa).

Table 4.1: Primers designed for amplification of *Tb*MSP-C.

Primer	Tm (^{o}C)	%GC
Fwd: 5′ GG *GAATTC* **ATG** ACC CAA CTG TTA GGA ACC GCT ATC **3′** [a]	**59.7**	**48**
Rvs: 5′ GC *GCGGCCGC* CAT TTG ACG AGG GTT CAC GGC CGC GTT **3′** [b]	**64.3**	**59**

Restriction sites are underlined ([a]EcoRI and [b]NotI) and start codon is indicated in bold.

The *TbMSP-C* gene was amplified from *T. brucei* Lister 427 genomic DNA using Dream™ Taq polymerase (1 U) in 1 x Dream™ Taq buffer (2 mM $MgCl_2$), 0.8 µM dNTP mix, 0.5 µM forward and reverse primers and 1 ng genomic DNA made up to 20 µl reaction volume with MilliQ® deionised water.

The *TbMSP-C* gene was amplified from *T. brucei* LiTat 1.3 from genomic DNA using High Fidelity Taq polymerase (0.5 U) in 1 x HF buffer, 2 mM $MgCl_2$, 0.32 µM dNTP mix, 0.5 µM forward and reverse primers and 1 ng genomic DNA made up to 20 µl reaction volume with MilliQ® deionised water.

Both PCR reactions followed the same cycling conditions, using the T100™ Thermocycler (BioRad, USA), as indicated in Figure 4.1 and the products analysed on a 1% (w/v) agarose gel as before. The concentration of the insert was determined using the G:BOX imaging system and gene tool software (Syngene, UK).

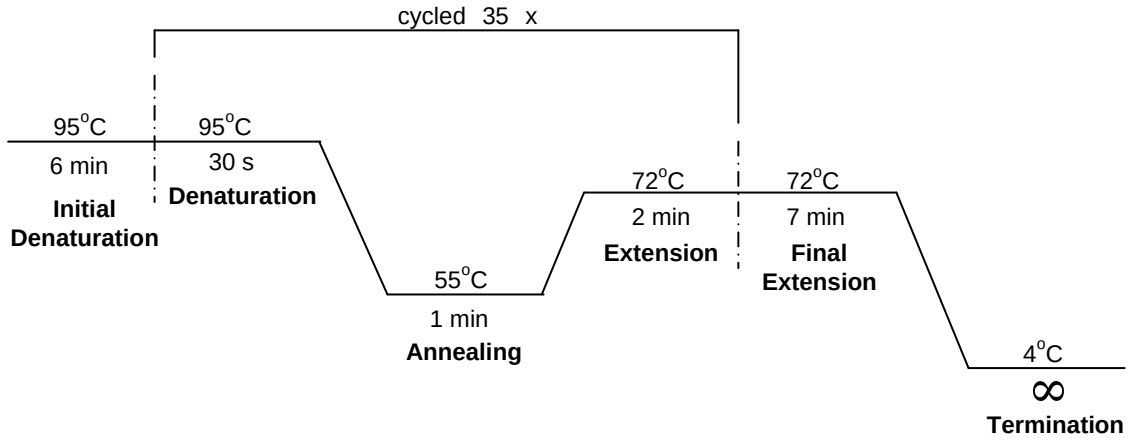


Figure 4.1: PCR cycling conditions used to amplify *TbMSP-C*.

4.2.5 Cloning of *TbMSP-C* into pGEM®-T and pGEX 4T-1 Vectors

Amplified *TbMSP-C* was ligated into the pGEM® T-vector (Appendix 7, Figure A7.1). This was done using T4 DNA ligase (1 U) in 1 x ligation buffer with 50 ng vector DNA and x ng insert DNA to a final volume of 10 µl. The insert DNA was used at a ratio of 4:1 with vector DNA. The amount of insert added was calculated using the following equation:

$$x\ ng = \frac{\text{amount of vector DNA (50 ng)} \times \text{size of insert DNA (kb)}}{\text{size of vector DNA (kb)}} \times \frac{4}{1}$$

Ligation was performed at 4°C overnight and transformation into *E. coli* JM109 cells completed the next day. The TransformAid™ kit was used according to the manufacturer's instructions (ThermoScientific) and the 40 µl final competent cell-ligation mixture was plated onto 2xYT media containing agar (1.5% (w/v) bacteriological agar) with 100 µg/ml ampicillin, 20 µg/ml X-gal and 0.1 mM IPTG for blue-white colony selection. The transformed cells were grown overnight at 37°C.

Once a white recombinant colony was obtained, the vector was tested to ensure the *TbMSP-C* insert was present. This was done by growing the white colony in 2xYT liquid medium containing 200 µg/ml ampicillin overnight at 37°C with shaking. From the overnight culture, plasmid DNA was isolated using the GeneJet plasmid DNA isolation kit (ThermoScientific) following the manufacturer's instructions. PCR using specific gene primers was performed, as described in Section 4.2.4, to determine if the correct gene was present.

Restriction digestion using EcoRI (1 U) and NotI (1 U) in 1 x 'O' buffer in a final volume of 10 µl was also done to excise the insert from the vector for analysis. Both the PCR and restriction digestion products was analysed on a 1% (w/v) agarose gel as before. Once the *MSP-C* gene was present within the pGEM® T-vector glycerol stocks were made by adding 25% (w/v) glycerol to an overnight culture and stored at -80°C.

Before cloning into a pGEX 4T-1 expression vector (Appendix 7, Figure A7.2), large scale restriction digestion was done on both the recombinant *TbMSP-C*-pGEM® T-vector and the pGEX 4T-1 expression vector and analysed on a 1% (w/v) agarose gel as before. The excised insert (containing EcoRI and NotI 'sticky ends') was gel extracted from a 1% (w/v) agarose gel containing 10 µg/ml crystal violet strain (Rand, 1996) using a peqGOLD Gel extraction kit (PEQLAB biotechnologie) following the manufacturer's instructions.

The concentration of extracted insert was determined using the Nanodrop 6000. After the digestion of the pGEX 4T-1 vector, it was dephosphorylated using Fast AP (1 U) in 1 x Fast AP buffer and 25 µl digested vector DNA made up to 30 µl with MilliQ® deionised water.

The enzyme was allowed to react for 10 min at 37°C and the reaction stopped by heat inactivation at 65°C for 15 min. The digested desphosphorylated pGEX 4T-1 vector was cleaned and concentrated using the ZymoResearch clean and concentrate kit as per the manufacturer's instructions. Concentration of pGEX 4T-1 vector was determined using the Nanodrop 6000.

Once both insert and expression vector contained EcoRI and NotI 'sticky ends', they were ligated using T4 DNA ligase as before. The ligation mix was transformed into BL21 (DE3) *E. coli* cells using the TransformAid™ kit according to the manufacturer's instructions and the 40 µl final competent cell-ligation mixture was plated onto 2xYT media containing agar supplemented with 100 µg/ml ampicillin. The plated cells were grown overnight at 37°C. Colonies that showed ampicillin resistance were tested for the presence of the insert by PCR and restriction digestion as before. Glycerol stocks of recombinant *E. coli* cells were made as described before. The recombinant pGEX 4T-1 vector was sequenced (University of Stellenbosch, South Africa) to ensure that the *MSP-C* gene was cloned and that no mutations occurred.

4.2.6 Transformation and Restriction Digestion of Recombinant (r)*Tb*MSP-C and r*Tco*MSP-C Constructs

The synthesised construct DNA (based on the *Tb*MSP-C sequence: AAO72980.1 and *Tco*MSP-C sequence: CCC93446.1) was resuspended in 50 µl MilliQ® dH_2O according to manufacturer's instructions. The construct DNA was transformed into BL21 (DE3) *E. coli* cells using $CaCl_2$ transformation (Sambrook *et al*., 2001). Briefly, a non-recombinant BL21 (DE3) *E. coli* colony was inoculated into 10 ml of 2xYT liquid medium and grown overnight, with shaking at 37°C. The overnight culture was diluted 1:100 with 90 ml 2xYT medium and grown until the OD at 600 nm was 0.3-0.4. The culture was transferred into ice cold, sterile centrifuge tubes and incubated on ice for 10 min. The cells were pelleted by centrifugation (5000 x g, 10 min, 4°C), resuspended in 40 ml of sterile $CaCl_2$ (60 mM $CaCl_2$, 10 mM HEPES, pH 7.0) and pelleted again (5000 x g, 10 min, 4°C). This was followed by resuspension in $CaCl_2$ solution (2 ml). The now competent cells (20 µl) were added to the construct preparation (1 µl) and incubated on ice for 30 min. Heat shock was performed by incubating cells at 42°C for 90 s followed by incubation on ice for 2 min.

The cells were added to 80 µl pre-warmed super optimal cataboliser with catabolite repression (SOC) medium [2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose] and incubated at 37^{o}C for 1 h with gentle shaking. The resulting mixture (100 µl) was plated onto 2xYT agar plates (supplemented with 100 µg/ml ampicillin) and cells grown overnight at 37^{o}C.

The resulting colonies were selected (five from each) and grown overnight with shaking at 37^{o}C in 2xYT medium containing 200 µg/ml ampicillin (10 ml). The overnight culture was used to prepare glycerol stocks [750 µl culture with 750 µl 50% (v/v) sterile glycerol] and the remainder used to perform plasmid DNA isolation using the GeneJet plasmid DNA isolation kit (ThermoScientific, USA), following the manufacturer's instructions. Restriction digestion of the r*Tb*MSP-C construct, using EcoRI (1 U) and NotI (1 U) in 1 x 'O' buffer in a final volume of 10 µl, was done to excise the insert from the vector and ensure its correct size (1800 bp). Similarly, restriction digestion of the r*Tco*MSP-C construct, using BamHI (1 U) and NotI (1 U) in 2 x tango buffer in a final volume of 10 µl, was also performed.

4.2.7 Recombinant Expression of *Tb*MSP-C, r*Tb*MSP-C and r*Tco*MSP-C.

The recombinant BL21 (DE3) *E. coli* cells were used to express *Tb*MSP-C within the pGEX 4T-1 vector with an N-terminal GST tag (Appendix 7, Figure A7.2)and r*Tb*MSP-C and r*Tco*MSP-C within the pET100-D vector with an N-terminal His tag (Appendix 7, Figure A7.3). Two methods were used in an attempt to express the proteins: (1) IPTG induction and (2) autoinduction. Recombinant *E. coli* cells were streaked on 2xYT agar containing 100 µg/ml ampicillin and incubated overnight at 37^{o}C. All *E. coli* culture work was performed in a sterile environment.

(1) A single *Tb*MSP-C colony was inoculated in 2xYT liquid medium (containing 100 µg/ml ampicillin) overnight at 37^{o}C. The overnight culture was used to perform a 1:100 dilution in fresh 2xYT medium (containing 100 µg/ml ampicillin) and the culture grown at 37^{o}C until an OD at 600 nm of 0.5-0.7 was reached. Expression was induced with 1 mM IPTG, supplemented with 100 µg/ml ampicillin and the culture grown for a further 4 h at 37^{o}C. Overnight culture (1 ml) was used as an uninduced control. Expressed culture (2 ml) was used as an induced sample. Cells were pelleted by centrifugation (5000 x g, 2 min, RT) and resuspended in 50 µl PBS and 50 µl reducing treatment buffer [125 mM Tris-HCl, pH 6.8 with 4% (w/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercaptoethanol with bromophenol blue]. Samples were stored at -20^{o}C and boiled for 2 min before being used.

(2) A single r*Tb*MSP-C or r*Tco*MSP-C colony was used to inoculate Terrific Broth autoinduction medium [1.2% (w/v) tryptone, 2.4% (w/v) yeast extract, 0.17 M KH_2PO_4, 0.72 M K_2HPO_4] and the culture grown at 30^{o}C overnight. Cells were pelleted by centrifugation (5000 x g, 2 min, RT) and resuspended in 50 µl PBS and 50 µl reducing treatment buffer. Samples were stored at -20^{o}C and boiled for 2 min before being used.

4.2.8 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western Blot Analyses.

Samples were analysed by SDS-PAGE as described by Laemmli (1970) under reducing and non-reducing conditions. A 4% stacking and 10% running gel were used and gels electrophoresed at 20 mA in tank buffer [250 mM Tris-HCl , 192 mM glycine, 0.1% (m/v) SDS, pH 8.3] using a BioRad casting system (USA). Gels were either stained with Coomassie blue [0.125% (m/v) Coomassie blue R-250, 50% (v/v) methanol, 10% (v/v) acetic acid], silver stained (Blum *et al.*, 1987) or transferred to nitrocellulose for analyses by western blot.

Proteins from the polyacrylamide gel were transferred to nitrocellulose using a Semi-phor® Semi-Dry blotter (Sigma, USA) at 20 V for 50 min or using a Biorad overnight transfer system (USA), at 20 mA, in blotting buffer [25 mM Tris-HCl, 190 mM glycine, 20% (v/v) methanol with 0.1% (w/v) SDS]. Transfer success was determined by staining with Ponceau S [0.1% (w/v) Ponceau S, 15% (v/v) acetic acid]. Stain was removed with a 0.5 M NaOH solution and protein free areas blocked using 5% (w/v) non-fat milk powder in Tris buffered saline (TBS; 20 mM Tris-HCl, 200 mM NaCl, pH 7.4) for 1 h. Primary antibody in blocking solution was added and the blot incubated at 4^{o}C, overnight with gentle mixing. Anti-peptide antibody was used at 2 µg/ml, anti-GST antibody at 0.75 µg/ml and anti-His antibody (Sigma, USA) at 1:1500, unless otherwise indicated. The blot was washed 3 x 5 min with TBS (washing). Rabbit anti-chicken IgY conjugated to horse radish peroxidase (HRPO) antibody [1:5000 dilution in 0.5% (w/v) bovine serum albumin (BSA) in TBS] was added and the blot incubated for 1 h. Following the washing step, protein bands were either visualised using 4-chloro-1-naphthol substrate solution [0.06% (m/v) 4-chloro-1-naphthol, 0.0015% (v/v) H_2O_2 in TBS] or using ECL with luminal substrate solution [Pierce, USA]. All SDS-PAGE gels and western blots were visualised using the Syngene G:BOX (Vacutec, UK)

4.2.9 Purification of Truncated (t)*Tb*MSP-C, r*Tb*MSP-C and *rTcoMSP*-C

Large scale expression of t*Tb*MSP-C was performed in 1000 ml 2xYT media induced with 1 mM IPTG for 4 hours at 37^{o}C. Large scale expression of both r*Tb*MSP-C and r*Tco*MSP-C was performed in 250 ml Terrific Broth overnight at 30^{o}C.

The cells were pelleted (6000 x g, 10 min) and resuspended in equilibration buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 7.0) containing 1 mg/ml lysozyme and 1 % (v/v) Triton™ X-100 and incubated at RT for 1 hour. The cells were sonicated 10 x on ice (30 s sonication, 30 s rest) and centrifuged (10000 x g, 15 min) to obtain the soluble and insoluble fractions. The insoluble t*Tb*MSP-C expression fraction was used to solubilise the protein by resuspending the pellet in PBS with 300 mM NaCl of the same volume as the supernatant and the solution was sonicated 10 x on ice (1 min sonication, 1 min rest). The solubilised protein was obtained by centrifugation (10000 x g, 5 min, 4^{o}C) and the supernatant was filtered twice through Whatman no. 1 filter paper before purification. The soluble fractions from the r*Tb*MSP-C and r*Tco*MSP-C expressions were clarified by filtering through 0.45 µm filter before purification.

Purification of t*Tb*MSP-C was performed using anti-*Tb*MSP:400-412 peptide antibodies (2 mg) specific for *Tb*MSP-C (Chapter 3) linked to UltraLink® Hydrazide resin. Purification of r*Tb*MSP-C and r*Tco*MSP-C was performed using UltraLink® Hydrazide resin linked to anti-*Tb/Tco*MSP:303-314 (4 mg) cross-species anti-peptide antibodies (Chapter 3).

The anti-peptide antibodies were linked to the resin following the manufacturer's instructions. Firstly, the antibody was incubated with 10 mM *meta*-periodate in the dark for 30 min at RT. The Zebu™ spin desalting column (10 ml, ThermoFisher) was prepared following the manufacturer's instruction. The column containing the resin was centrifuged (1000 x g, 2 min, 4^{o}C) and washed 3 x with coupling buffer (0.1 M NaH_2PO_4, pH 7.0) by centrifugation (1000 x g, 2 min, 4^{o}C). The oxidised antibody preparation was added to the desalting column and desalted oxidised antibody recovered by centrifugation (1000 x g, 2 min, 4^{o}C). The absorbance at 280 nm was recorded for this sample.

Secondly, UltraLink® Hydrazide resin (2 mg antibody per 1 ml) was added to a plastic column and allowed to settle for 15 min. The liquid was drained and resin washed with 5 column volumes coupling buffer. Finally, the oxidized antibody was added to the resin. The solution was mixed into the resin by gentle inversion and coupling was completed by incubating the column overnight at 4^{o}C using an end-over-end rotator. The resin was allowed to settle for 15 min and the liquid drained. It was washed with one column volume coupling buffer and the absorbance at 280 nm recorded. The resin was washed with 5 column volumes 1 M NaCl followed by 5 column volumes of phosphate storage buffer [0.1 M NaH_2PO_4 with 0.1% (w/v) NaN_3] and was stored at 4^{o}C until used.

The respective anti-peptide antibody affinity columns were equilibrated with 5 column volumes equilibration buffer (PBS) and the filtered soluble or the solubilised protein fraction

added. The columns were incubated overnight at 4°C on an end-over-end rotator to allow for antibody-antigen interaction.

The resin was allowed to settle for 15 min and the liquid drained. The flow-through was cycled over the column once more followed by washing with equilibration buffer until the absorbance at 280 nm was less than 0.02. The bound antigen was eluted by decreasing the pH with elution buffer [100 mM glycine-HCl, 0.02% (w/v) NaN_3, pH 2.8] and 10 x 1 ml fractions, containing 100 µl of neutralisation buffer [1 M Na_2HPO_4, 0.02% (w/v) NaN_3, pH 8.5], were collected and stored at -20°C until use. Samples (10 µl) were combined with 10 µl reducing treatment buffer and boiled for 2 min before analysis using SDS-PAGE and western blot.

4.2.10 Protein Concentration Determination

All fractions eluted from the hydrazide immunoaffinity resin were pooled and concentrated using Centricon® centrifugal concentrators (6000 x g, 5 min, 4°C). The Bradford protein assay was performed to determine the protein concentration as described by Bradford (1976). In order to obtain a standard curve (Figure 4.2), 1 mg/ml ovalbumin stock was used to prepare protein standards (1000 µg/ml - 125 µg/ml) in triplicate. Protein standards (10 µl) were mixed with an equal volume of Bradford reagent [10 µl; 0.12 % (w/v) Coomassie brilliant blue G-250 dissolved in 500 ml of a 2 % (v/v) perchloric acid solution]. The absorbance at 595 nm was recorded after 10 min using a Nanodrop 6000.

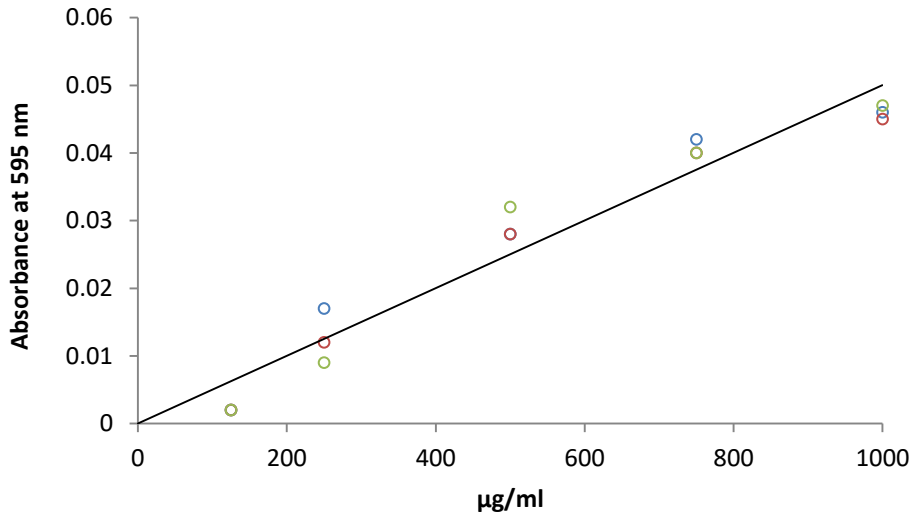


Figure 4.2 Bradford standard curve used for protein quantification. Triplicate ovalbumin standards (1000 µg/ml - 125 µg/ml) were added to Bradford reagent and the resulting absorbances read at 595 nm using a Nanodrop 6000 spectrophotometer. The equation of the fitted line is y=0.00005x - 0.0003.

4.2.11 Gelatin Zymography of t*Tb*MSP-C

Enzyme activity of t*Tb*MSP-C was visualised using zymography (substrate SDS-PAGE) according to Heussen and Dowdle (1980). Gelatin [0.1% (w/v)] was incorporated into a 10% running gel and SDS-PAGE completed as normal. The samples were prepared under non-reducing conditions (mixed with non-reducing treatment buffer ([125 mM Tris-HCl -HCl, pH 6.8 with 4% (w/v) SDS, 20% (v/v) glycerol with bromophenol blue] and were not boiled).

The gel was washed with assay buffer (50 mM Tris-HCl , 5 mM $CaCl_2$, 1 µM $ZnCl_2$, pH 7.4) containing 2.5% Triton™ X-100 for 2 x 30 min at RT, briefly rinsed with dH_2O and incubated in assay buffer (50 mM Tris-HCl , 5 mM $CaCl_2$, 1 µM $ZnCl_2$, pH 7.4) for 6 h or overnight at 37^{o}C and stained with Coomassie blue as described by Hassani *et al.* (2014).

4.2.12 Molecular Exclusion Chromatography of r*Tb*MSP-C and r*Tco*MSP-C

Purified r*Tb*MSP-C and r*Tco*MSP-C were subjected to molecular exclusion using the HiPrep™ 16/16 Sephacryl™ S200 HR resin (16 x 600 mm) with the ÄKTA purifier®. The packed and calibrated column was equilibrated with 60 ml MilliQ dH_2O at 1 ml/min and, thereafter, with 180 ml MEC buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) at 1 ml/min. After equilibration, r*Tb*MSP-C and r*Tco*MSP-C were added, in separate experiments, to the column and eluted with MEC buffer in 2 ml fractions. The elution profile was compared to that obtained by separating standard proteins of known molecular weight to determine the size of MSP-C under non-denaturing conditions. Absorbance peaks were subjected to analysis by SDS-PAGE under reducing conditions.

4.2.13 Hydrolysis of H-Suc-Leu-Tyr-AMC Peptide Substrate by r*Tb*MSP-C and r*Tco*MSP-C

The activity of r*Tb*MSP-C and r*Tco*MSP-C was assessed by measuring its hydrolysis of the fluorogenic peptide substrate H-Suc-Leu-Tyr-AMC. Protein, diluted in assay buffer (0.5-1.5 µg/ml), was preincubated at 37^{o}C for 10 min before addition of substrate (7.5-10 µM) within a black NUNC® microtitre plate. The plate was incubated at 37^{o}C for 10 min intervals and free 7-amino-4 methylcoumarin (AMC) excited at 360 nm and emission read at 460 nm using the FLUOstar Optima spectro-fluorometer (BMG Labtech, Germany).

4.2.14 Inhibition of r*Tb*MSP-C and r*Tco*MSP-C

Recombinant protein in assay buffer (0.5 ug/ml) was incubated with EDTA or 1,10 phenanthroline (0.01-1 mM) at 37^{o}C for 10 minutes within a black NUNC® microtitre plate, after which the H-Suc-Leu-Tyr-AMC peptide substrate (7.5 µM) was added.

This reaction mix was incubation for 10 min at 37^{o}C with substrate before free 7-amino-4 methylcoumarin (AMC) was excited at 360 nm and fluorescence emission at 460 nm read using the FLUOstar Optima spectro-fluorometer (BMG Labtech, Germany).

4.2.15 pH Profile of r*Tb*MSP and r*Tco*MSP-C

Protein (0.5 µg/ml) was preincubated with constant ionic strength buffer AMT [acetic acidic (100 mM), MES (100 mM) and Tris (200 mM), pH 2-11] buffer at 37^{o}C for 10 min, after which, 7.5 µM of H-Suc-Leu-Tyr-AMC substrate was added within black NUNC® microtitre plates. Plates were incubated at 37^{o}C for 10 min and free AMC excited at 360 nm and emission read at 460 nm using a FLUOstar Optima spectro-fluorometer.

4.2.16 Detection of antibodies against r*Tco*MSP-C with Infected Cattle Sera

T. congolense infected (CVB 151 +28) and non-infected (CVB 186 -7) cattle sera were used to detect r*Tco*MSP-C antigen, using ELISA (Section 3.2.5). NUNC® Maxisorb 96 well microtitre plates were coated with r*Tco*MSP-C (10-50 µg/ml) overnight at 4^{o}C. The plates were blocked with 0.5% (w/v) BSA in PBS for 1 h at 37^{o}C, followed by washing 3 x 300 µl with PBS Tween-20 using the Biotek ELx50 plate washer. Infected or non-infected sera, diluted with 0.5% (w/v) BSA-PBS (1:200), were added to the microtitre plate (100 µl/well) and incubated for 2.5 h at 37^{o}C, followed by washing with PBS Tween-20 (3 x 300 µl). Secondary rabbit anti-bovine IgG conjugated to HRPO (1:5000 in BSA-PBS) was added (100 µl/well) and incubated for 1 h at 37^{o}C after which the plate was washed 3 x 300 µl with PBS Tween-20. Addition of ABTS/H_2O_2 substrate solution (Section 3.2.5; 150 µl/well) was followed by absorbance readings at 405 nm every 15 min for 1 h using the FLUOstar Optima spectro-fluorometer (BMG Labtach, Germany).

4.3 Results

4.3.1 Genomic DNA Isolation

Genomic DNA was isolated from both procyclic Lister 427 and blood stage LiTat 1.3 *T. brucei* strains using phenol:chloroform followed by standard ethanol precipitation (Figure 4.3). Genomic DNA from both *T. brucei* strains was detected at a size larger than 10000 bp (Figure 4.3).

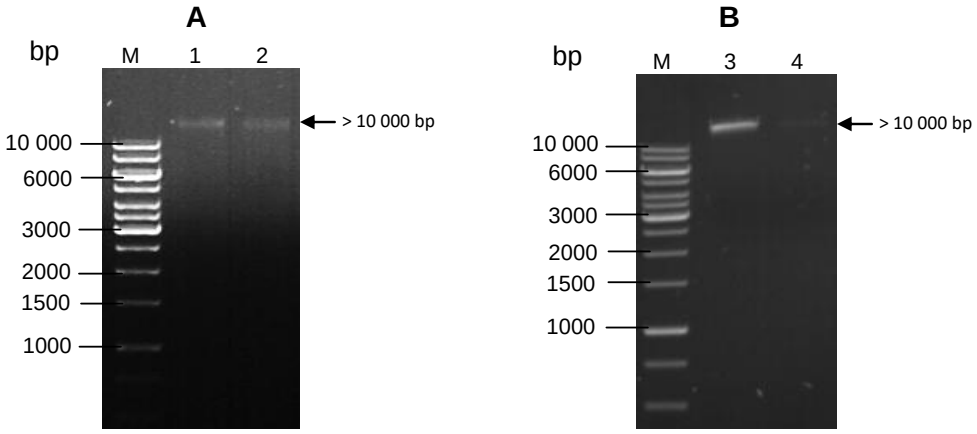


Figure 4.3: Genomic DNA extraction from procyclic (Lister 427) and blood stage (LiTat 1.3) *T. brucei* strains, evaluated on 1% (w/v) agarose gels. Gels were stained with ethidium bromide and viewed under ultra-violet light. Lane M, O'GeneRuler 1 kb DNA ladder. Genomic DNA from **(A)** procyclic (lane 1-2) and **(B)** blood stage form (lane 3-4) parasites are shown. Arrows indicates presence of genomic DNA at >10000 bp.

4.3.2 PCR of *TbMSP-C* and Cloning into pGEM®-T and pGEX 4T-1 Vectors

Polymerase chain reaction (PCR) was conducted on the genomic DNA from both parasite forms. Figure 4.4 shows the successful amplification of a DNA product of 1800 bp in size which corresponds to the calculated size of the *TbMSP-C* gene. Other DNA products at 3300 bp and 1400 bp were also present. Band intensity was used to calculate the concentration of the DNA for the three bands shown in lane 7 (Figure 4.4). For the bands at 1400 bp and 3300 bp, the concentration was less than 10 ng/µl, while the 1800 bp band had a concentration of 293 ng/µl. The concentration ratio of the expected *TbMSP-C* gene (1800 bp band) to non-specific DNA products (1400 bp and 3300 bp bands) is ~29:3:1 suggesting that the *TbMSP-C* gene would be cloned in preference.

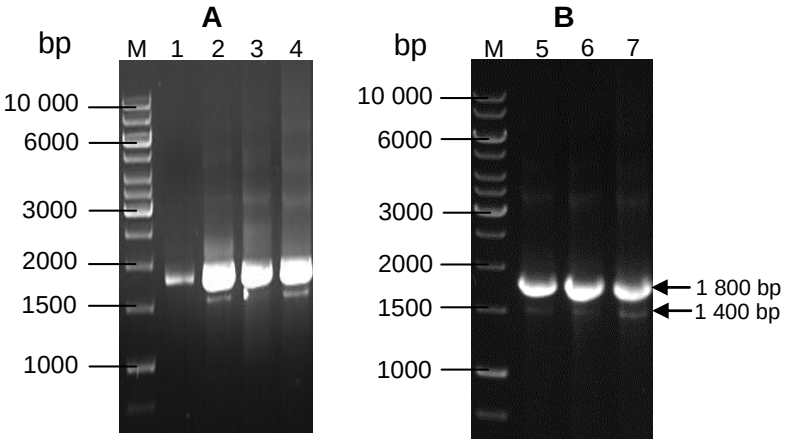


Figure 4.4: Polymerase chain-reaction of *Tb*MSP-C on both procyclic and blood stage form *T. brucei* genomic DNA, evaluated on 1% (w/v) agarose gels. Lane M, O'GeneRuler 1kb DNA ladder. PCR performed on **(A)** procyclic form (lanes 1-4) or **(B)** blood stage form (lanes 5-7) genomic DNA. Arrows indicate sizes of DNA bands obtained.

Once a substantial amount of amplified *TbMSP-C* DNA was obtained, it was cloned into the pGEM®-T vector and transformed into *E. coli* JM109 cells. Analysis of the recombinant T-vector (Figure 4.5, lane 1) showed at least three forms of DNA i.e. linear, nicked and supercoiled (Lodge *et al.*, 2007). The tightly compacted, smaller, supercoiled form is most abundant as reported in the literature (Lodge *et al.*, 2007). In comparison, digestion of recombinant pGEM®-T-*Tb*MSP-C with the restriction enzymes, EcoRI and NotI, shows the excised insert of 1800 bp (lane 2). The band at 3000 bp represents the linear pGEM®-T vector, while the larger band at 4750 bp represents undigested T-vector. The undigested pGEX 4T-1 expression vector appeared at a size 3000 bp (Figure 4.5, lane 3). However, on digestion, with EcoRI and NotI, linearisation occurred resulting in a 4900 bp product (lane 4).

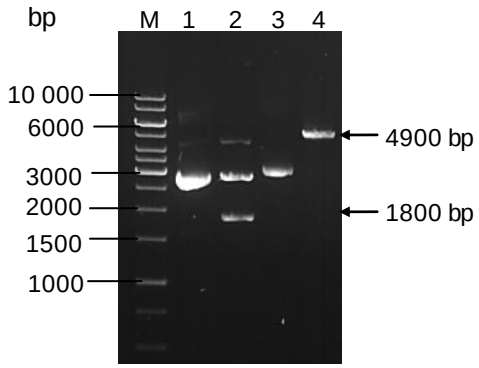


Figure 4.5: Undigested and digested pGEM®-T-*TbMSP-C* and expression vector pGEX 4T-1 using EcoRI and NotI, analysed on 1% (w/v) agarose gel. Lane M, O'GeneRuler 1kb DNA ladder; lane 1, undigested recombinant pGEM®-T-*TbMSP-C*; lane 2, digested recombinant pGEM®-T-*TbMSP-C*; lane 3, undigested pGEX 4T-1; lane 4, digested pGEX 4T-1. Arrows show sizes of linear pGEX 4T-1 (4900 bp) and *TbMSP-C* (1800 bp).

The digested *TbMSP-C* gene was sub-cloned into the pGEX 4T-1 expression vector and transformed into *E. coli* BL21 (DE3) cells. Analysis of a recombinant colony by PCR shows the amplification of a gene at 1800 bp (Figure 4.6, lane 1). Restriction digestion of recombinant pGEX 4T-1, with EcoRI and NotI, shows the release of the insert (1800 bp) from the vector (4900 bp; lane 3), in comparison with the undigested vector (lane 2). It was concluded that the *TbMSP-C* gene was successfully ligated into the expression vector and the transformed *E. coli* cells could be used to express *Tb*MSP-C recombinantly.

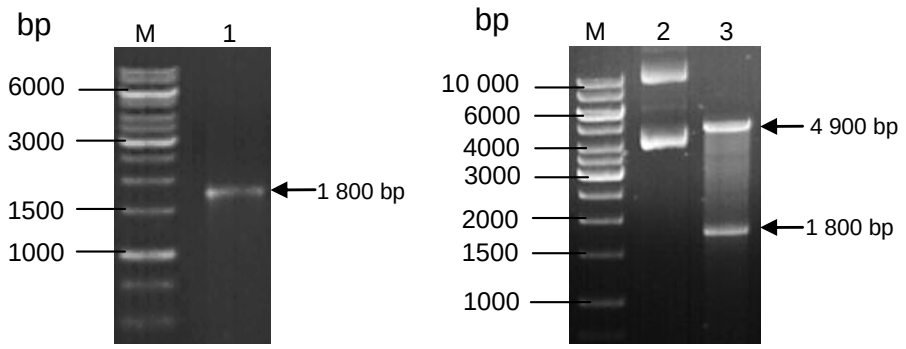


Figure 4.6: Determining whether the expression vector is recombinant as analysed on 1% (w/v) agarose gel. Lane M, O'GeneRuler 1kb DNA ladder; lane 1, PCR using *T. brucei* MSP-C specific primers on pGEX 4T-1; lane 2, undigested recombinant pGEX 4T-1-*TbMSP-C* DNA; lane 3, restriction digestion of recombinant pGEX 4T-1-*TbMSP-C* DNA using EcoRI and NotI. Arrows show linear pGEX 4T-1 (4900 bp) and *Tb*MSP-C (1800 bp).

4.3.3 Expression of *Tb*MSP-C (t*Tb*MSP-C*)* in *E. coli*

Taking into account the GST tag of 26 kDa, *Tb*MSP-C would be expressed at 92 kDa. Over expression, using IPTG, of *E. coli* cells containing the recombinant pGEX 4T-1 vector was compared to expression of *E. coli* cells containing the non-recombinant pGEX 4T-1 vector. The expressed proteins were resolved on SDS-PAGE and western blotted using anti-GST antibodies (Figure 4.7). The *E. coli* lysate of uninduced cells showed no induced proteins (panel B, lanes U1 and U2). Expression of the *E. coli* cells containing non-recombinant pGEX 4T-1 vector resulted in a single protein of 26 kDa corresponding to the GST tag (panel B, lane 1). Expression of the *E. coli* cells containing recombinant pGEX 4T-1-*TbMSP-C* vector resulted in numerous protein bands and the four main bands are labelled (panel B, lane 2). The smallest 26 kDa band is suspected to be GST. The bands at 54 kDa and 39 kDa are suspected to be truncated forms of the fusion protein, while the 92 kDa band would correspond to full length GST-*Tb*MSP-C fusion protein.

The detection of a single 26 kDa band after non-recombinant pGEX 4T-1 expression suggests that the numerous bands detected in the western blot with anti-GST antibodies, after recombinant *Tb*MSP-C expression, are due to the presence of the *TbMSP-C* gene and are likely truncated forms of the protein. Since there are these unexpected truncated forms of the protein present during expression in *E. coli*, the expressed protein from cloned recombinant pGEX 4T-1-*TbMSP-C* will be referred to as truncated (t)*Tb*MSP-C.

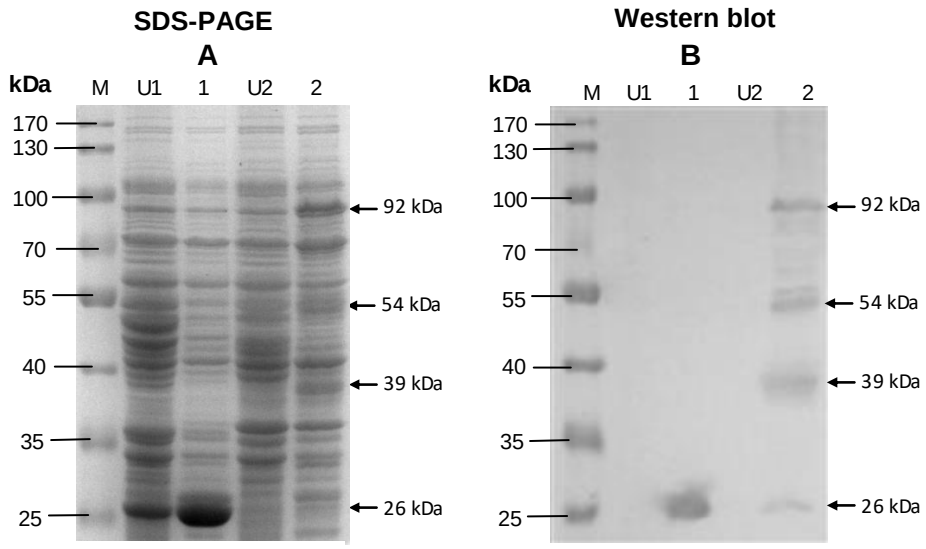


Figure 4.7: Expression of (t)*Tb*MSP-C (recombinant pGEX 4T-1) in comparison with non-recombinant pGEX 4T-1. (A) 10% SDS-PAGE gel and **(B)** western blot using chicken anti-GST primary antibody, rabbit anti-chicken secondary antibody conjugated to horseradish peroxidase and addition of 4-chloro-1-naphthol/H_2O_2 chromogenic substrate. Lane M, ThermoScientific pre-stained marker; lane U1 and U2, uninduced non-recombinant and recombinant pGEX 4T-1; lane 1, induced non-recombinant pGEX 4T-1; lane 2, induced recombinant pGEX 4T-1. Arrows indicate full length and truncated forms of *Tb*MSP-C.

4.3.4 Gene Sequencing Analyses

Gene sequencing and translated amino acid sequence alignments are shown in Appendix 6. Four point mutations which caused the translated protein sequence to change occurred during the cloning procedure. Mutations one (F^{10}-L^{10}), three (N^{207}-S^{207}) and four (V^{226}-I^{226}) do not result in a change in amino acid class. However, mutation two (H^{27}-L^{27}) results in a change from a positively charged amino acid to a hydrophobic amino acid. Mutations that result in a change in amino acid class can result in a change in final tertiary structure and therefore enzyme activity (Lodish *et al.*, 2000). A deletion also occurred within the recombinant *Tb*MSP-C sequence which resulted in a frameshift at amino acid residue 504 and a premature stop codon at amino acid residue 520 (Appendix 6).

4.3.5 Anti-Peptide Antibody Specificity against t*Tb*MSP-C

The recombinantly expressed protein was shown to be detected with the anti-peptide antibodies (Figure 4.8) produced as described in Chapter 3. Anti-*Tb*MSP:400-412 peptide antibodies designed to detect *Tb*MSP-C were able to detect full length t*Tb*MSP-C, at 92 kDa. However, they were unable to detect the truncated forms (Figure 4.8, panel C, lane 1 and 3). Furthermore, a protein band at 80 kDa was also seen, which is detected by anti-GST antibodies as well (compare Figure 4.8 C to Figure 4.7 B).

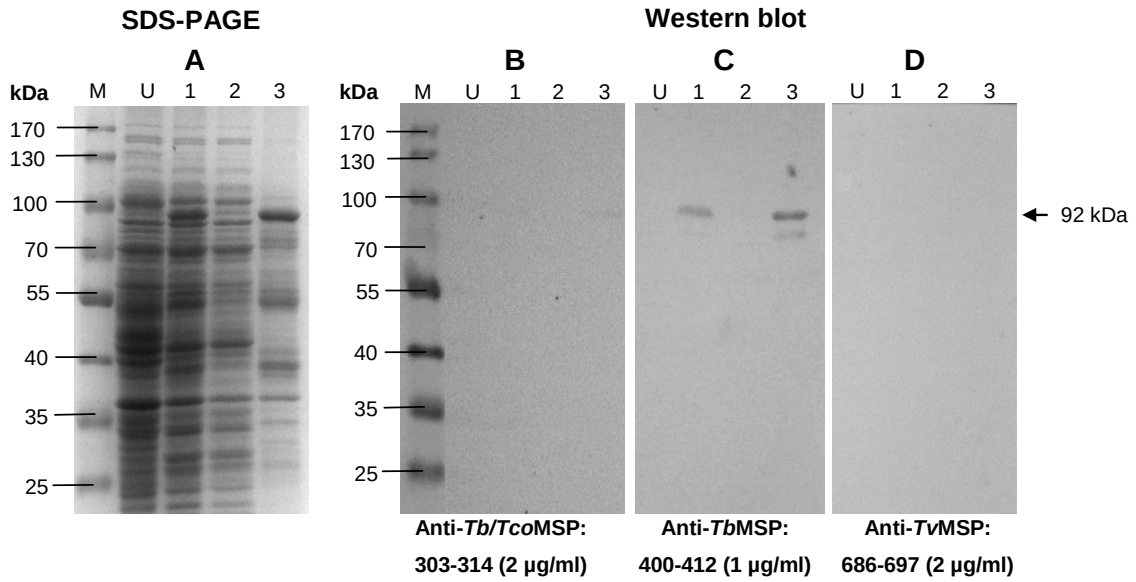


Figure 4.8: Detection of t*Tb*MSP-C with anti-peptide antibodies. (A) 10% SDS-PAGE gel and western blots with **(B)** anti-*Tb/Tco*MSP:303-314, **(C)** anti-*Tb*MSP:400-412 or **(D)** anti-*Tv*MSP:686-697 peptide antibodies as indicated. Rabbit anti-chicken secondary antibody conjugated to horseradish peroxidase was used and addition of 4-chloro-1-naphthol/H_2O_2 chromogenic substrate. Lane M, ThermoScientific pre-stained marker; lane U, uninduced control; lane 1, IPTG induced culture; lane 2, soluble fraction; lane 3, insoluble fraction. Arrow indicates t*Tb*MSP-C.

The t*Tb*MSP-C protein was detected in inclusion bodies in the insoluble fraction (Figure 4.8, lane 3). Unexpectedly, detection of full length t*Tb*MSP-C by anti-*Tb/Tco*MSP:303-314 peptide antibodies was also seen (panel B, lane 3). No protein was detected by anti-*Tv*MSP:686-697 peptide antibodies (panel B) which was expected. The fact that the expressed t*Tb*MSP-C protein was detected by anti-*Tb*MSP:400-412 peptide antibodies, infers that the protein being expressed is *Tb*MSP-C. These antibodies were used to purify t*Tb*MSP-C in the hope of diminishing the presence of truncated forms of the protein.

4.3.6 Purification and Activity of t*Tb*MSP-C

Detection of t*Tb*MSP-C in the insoluble fraction meant the protein had to be solubilised before purification. After solubilisation, the 92 kDa t*Tb*MSP-C from the insoluble fraction (Figure 4.8, lane 1) moved to the soluble fractions (Figure 4.9, lanes 2-4). This fraction was then used to purify t*Tb*MSP-C using anti-*Tb*MSP:400-412 peptide antibodies (Figure 4.10).

The soluble 92 kDa t*Tb*MSP-C is seen in both the soluble and unbound fractions (Figure 4.10, lanes 1-2) with purification of a smaller 70 kDa protein seen (lanes 3-12). Small amounts of full length t*Tb*MSP-C are, however, seen (Figure 4.10, panel A, lanes 5-9) and a number of lower molecular weight contaminating proteins are also present. No protein of 70 kDa is seen or detected in the soluble fraction used for purification (Figure 4.10, panel B, lane 1) and so it is suspected that the protein underwent degradation or enzymatic processing when bound in its native form to the antibody. This may also be due to the low

pH used when eluting the protein from the resin. Since the protein was purified using specific anti-peptide antibodies, it is suspected that the smaller protein is a truncated form of 92 kDa fusion protein.

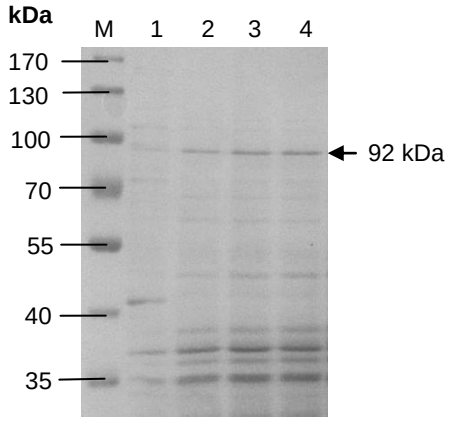


Figure 4.9: Solubilisation of t*Tb*MSP-C as analysed on a 10% SDS-PAGE gel. Lane M, ThermoScientific pre-stained marker; lane 1, insoluble fraction; lane 2, solubilised fraction; lane 3 and 4, filtrate 1 and 2. Arrow indicates full length t*Tb*MSP-C.

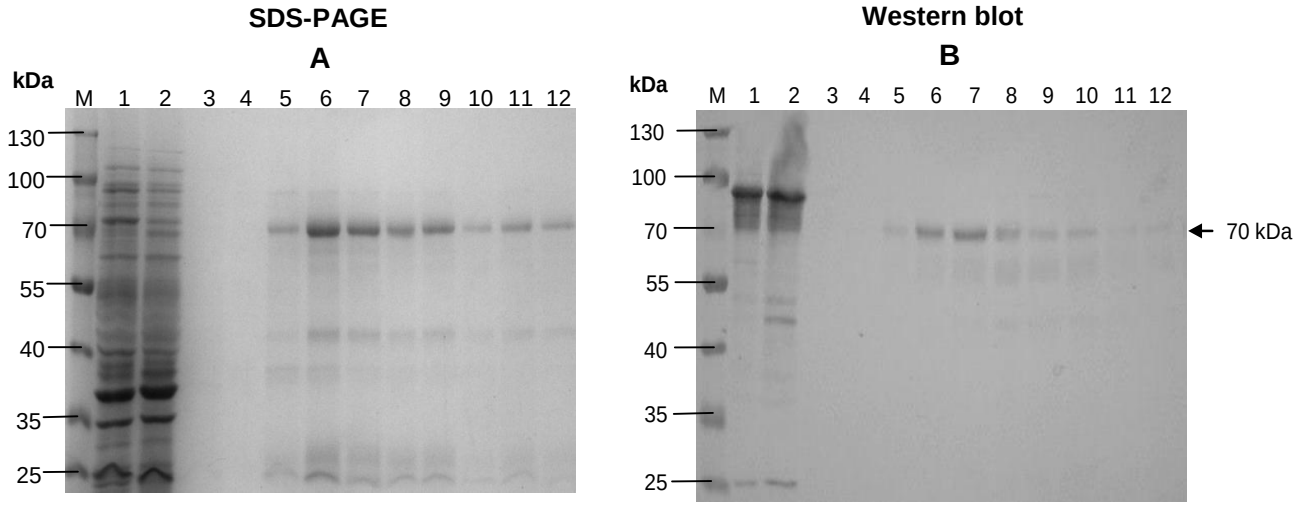


Figure 4.10: Purification of recombinantly expressed and solubilised t*Tb*MSP-C. (A) 10% SDS-PAGE and **(B)** western blot using chicken anti-*Tb*MSP:400-412 peptide antibody, rabbit anti-chicken secondary antibody conjugated to HRPO with addition of 4-chloro-1-naphthol/H_2O_2 chromogenic substrate. Lane M, ThermoScientific pre-stained marker; lane 1, solubilised fusion protein loaded onto the column; lane 2, unbound protein from the column; lane 3-12, purified protein fractions 1-10 (1 ml each). Arrow indicates purified t*Tb*MSP-C.

To ensure that the purified protein was an active protease, gelatin zymography was performed (Figure 4.11). When low amounts of purified t*Tb*MSP-C (0.25 µg) were present a single clear band was seen at 55 kDa. However, when a larger amount of t*Tb*MSP-C (1 µg)

was present smear of digestion was seen at ~55 kDa and digestion products were also seen at 70 and 92 kDa.

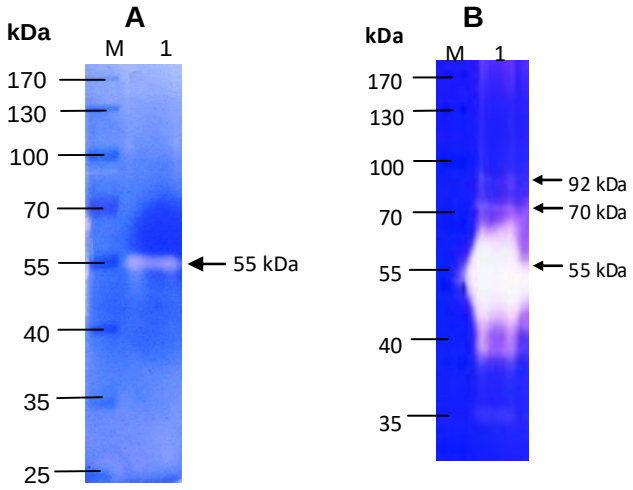


Figure 4.11: Protease activity of purified t*Tb*MSP-C as analysed on a 0.1% (w/v) gelatin zymogram stained with Coomassie blue. (A) 0.25 µg protein and **(B)** 1 µg protein. Lane M, ThermoScientific pre-stained marker; lane 1, purified t*Tb*MSP-C. Arrow at 55 kDa indicates active t*Tb*MSP-C and at 70 and 92 kDa latently active t*Tb*MSP-C.

This smaller 55 kDa protein is present in small quantities in the purification of t*Tb*MSP-C (Figure 4.10, panel A and B, lanes 6-9) and may represent a fully active *Tb*MSP-C, without the proregion. The larger 70 kDa and 92 kDa protein showed little activity and could represent a latently active enzyme in the presence of the proregion. This 55 kDa active protein is probably present due to the numerous truncations seen when t*Tb*MSP-C is expressed in *E. coli* (Figure 4.7).

Due to the presence of numerous truncations of the t*Tb*MSP-C protein, the mutations present in the cloned gene, the unexplained reduction of size during purification and the presence of an active smaller 55 kDa protein, it was decided to have full length *Tb*MSP-C and *Tco*MSP-C constructs synthesised and optimised for expression in *E. coli*. These will be called recombinant (r) *Tb*MSP-C and *Tco*MSP-C.

4.3.7 Transformation of Synthesised Constructs: r*Tb*MSP-C and r*Tco*MSP-C

The construct plasmid r*Tb*MSP-C and r*Tco*MSP-C, obtained from GeneArt, were transformed into *E. coli* BL21 (DE3) cells. Recombinant cells were used to obtain plasmid DNA and restriction digestion performed to ensure the correct size gene product was present. Undigested r*Tb*MSP-C and r*Tco*MSP-C DNA again showed the presence of three main bands of DNA i.e. linear, nicked and supercoiled (Figure 4.12, lane 1 and 3). Restriction digestion of r*Tb*MSP-C and r*Tco*MSP-C resulted in the release of a 1800 bp

product indicative of the MSP-C gene (Figure 4.12, lane 2 and 4), comparable to the PCR product obtained previously (Figure 4.4).

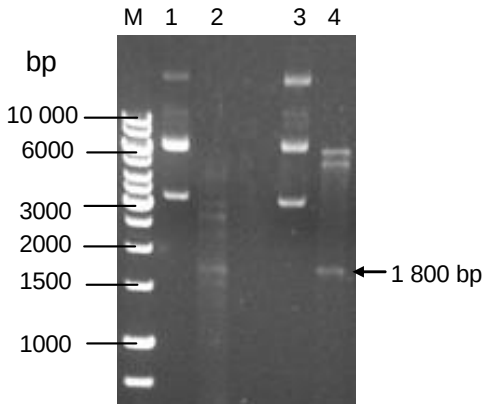


Figure 4.12: Restriction digestion of r*Tb*MSP-C and r*Tco*MSP-C plasmid DNA obtained from recombinant *E. coli* cells as analysed on 1% (w/v) agarose gel. Lane M, O'GeneRuler 1kb DNA ladder; lane 1 and 3, undigested plasmid DNA from *Tb*MSP-C and *Tco*MSP-C, respectively; lane 2, digested *Tb*MSP-C construct with NotI and EcoRI; lane 4, digested *Tco*MSP-C construct with NotI and BamHI. Arrow shows presence of MSP-C gene (1800 bp).

4.3.8 Expression of r*Tb*MSP-C and r*Tco*MSP-C

Recombinant *E. coli* BL21 (DE3) cells were used to express r*Tb*MSP-C and r*Tco*MSP-C protein in Terrific Broth overnight and the solubility of each expressed protein was assessed. The theoretical size of *Tb*MSP-C is 70 kDa while *Tco*MSP-C is slightly smaller at 68 kDa. Expression was assessed by SDS-PAGE and western blot (Figure 4.13, lane 1) and a 71 kDa protein was observed for both plasmids as detected with both anti-*Tb/Tco*MSP:303-314 peptide antibodies [(Figure 4.13, panels B1 and B2); (designed to detect *Tb* and *Tco*MSP-**B**)] and anti-His antibodies (Figure 4.14, panels C1 and C2). The detection of MSP-C by the anti-*Tb/Tco*MSP:303-314 peptide antibody was unexpected since, the peptide only shared 42% identity with the protein (Figure 3.5, Section 3.3.1; see Section 4.3.11 for more information). The expressed proteins contained the N-terminal His-tag as evidenced by their detection with anti-His tag antibodies (Figure 4.13, panels C1 and C2).

The presence of smaller protein products was also unexpectedly observed and may be truncated protein. The r*Tb*MSP-C construct expressed a smaller 60 kDa protein (Figure 4.13, panel B1), while, the r*Tco*MSP-C construct expressed two smaller proteins of 65 and 43 kDa (Figure 4.13, panel B2). In the case of r*Tco*MSP-C expression, smaller protein products were only detected with anti-*Tb/Tco*MSP:303-314 peptide antibodies and not anti-His antibodies which suggests the loss of size occurred at the N-terminus of the protein. On the other hand, the r*Tb*MSP-C 60 kDa protein was detected by both

anti-*Tb/Tco*MSP:303-314 and anti-His antibodies indicating that the reduction in size occurred at the C-terminus. The theoretical size of active MSP-C, without the proregion, is 53-55 kDa and, therefore, these unexpected products are not the active form of the protein. The protein was shown to be soluble (Figure 4.13, lane 2) with low amounts of protein detected in the insoluble fraction (lane 3). The soluble fraction was used for the purification of recombinant protein.

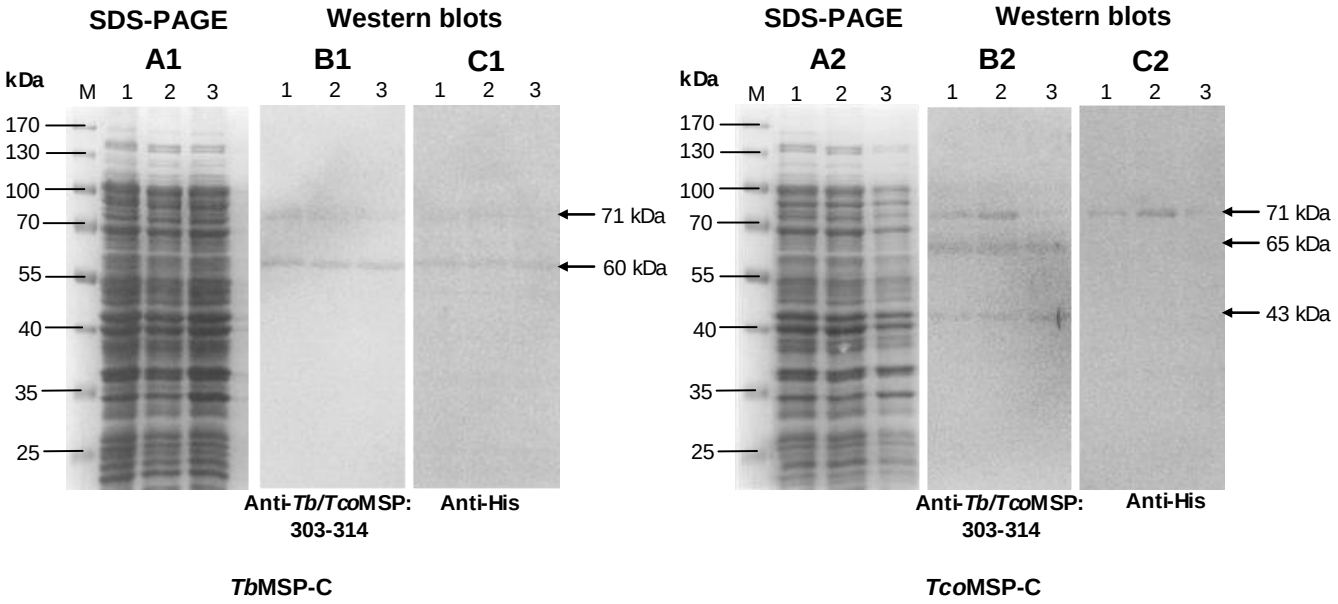


Figure 4.13: Expression and solubility of r*Tb*MSP-C (1) and r*Tco*MSP-C (2). Protein expression was performed overnight at 30^{o}C in Terrific Broth and analysed using a 10% SDS-PAGE gel **(A1 and 2)** and western blot with anti-*Tb/Tco*MSP:303-314 **(B1 and 2)** and anti-His **(C1 and 2)** primary antibody, rabbit anti-chicken secondary antibody conjugated to HRPO and visualised by ECL luminescence using the Syngene G:BOX. Lane M, ThermoScientific pre-stained marker; lane 1, overnight culture; lane 2, soluble supernatant fraction and lane 3, insoluble pellet fraction. Arrows indicate detection of 71 kDa MSP-C, and smaller protein bands.

4.3.9 Purification of r*Tb*MSP-C and r*Tco*MSP-C

Recombinantly expressed r*Tb*MSP-C and r*Tco*MSP-C was successfully detected by anti-*Tb/Tco*MSP:303-314 peptide antibodies. It was decided to use these antibodies, coupled to hydrazide resin, to purify the recombinant protein as metal chelate affinity purification proved to be inefficient (data not shown). The soluble fraction obtained after expression was loaded onto the antibody conjugated resin after which bound antigen was eluted. The successful purification of a 71 kDa protein, of both proteins, as detected by both anti-*Tb/Tco*MSP:303-314 peptide (Figure 4.14, panels C and D, lanes 2-11) and anti-His antibodies (Figure 4.14, panels E and F, lanes 2-11). was achieved. No protein was detected in the unbound sample (Figure 4.14, panels C, D, E and F, lane 1).

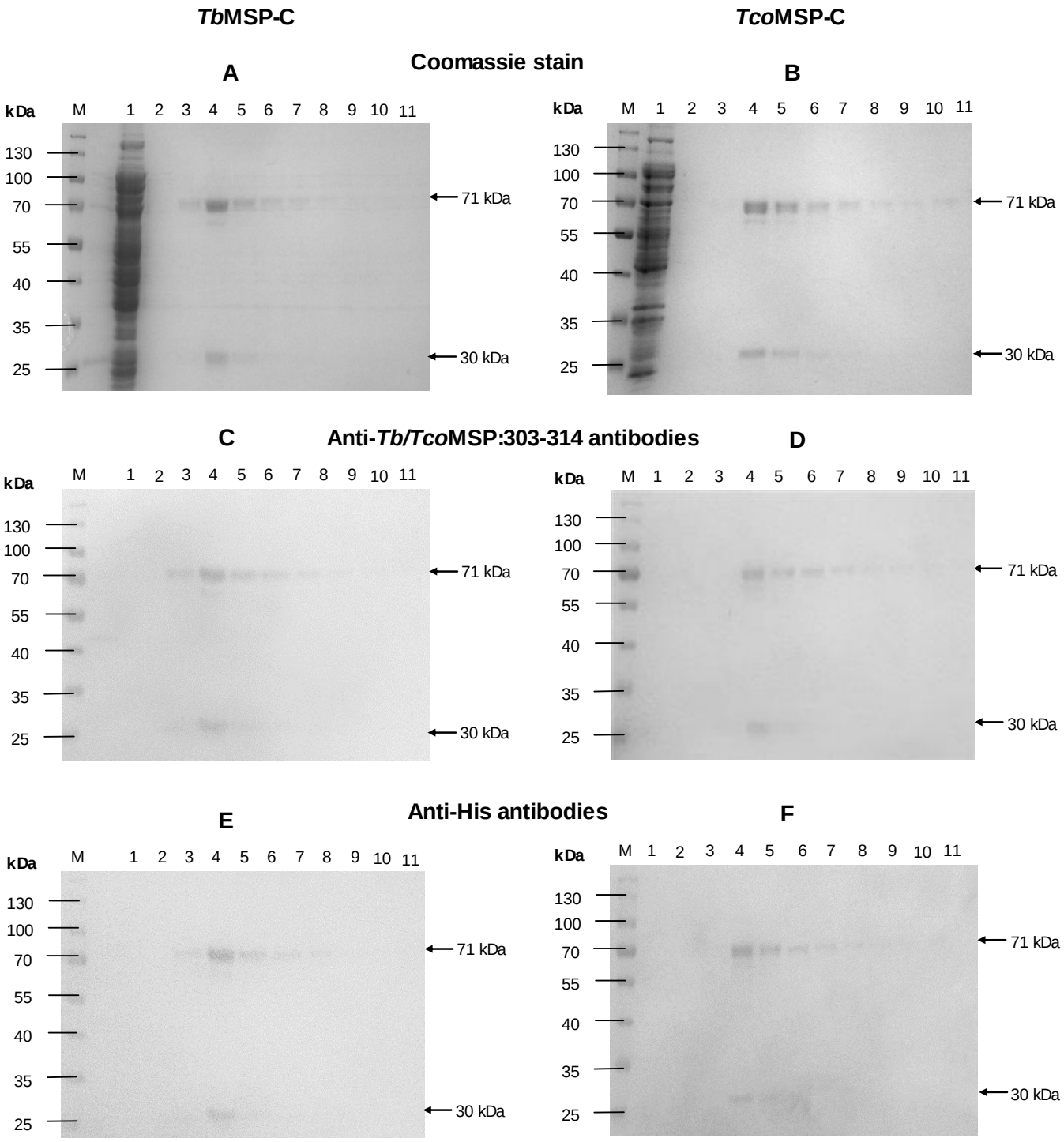


Figure 4.14: Purification of r*Tb*MSP-C (A, C and E) and r*Tco*MSP-C (B, D and F). (A and B) 10% SDS-PAGE gel and western blot using **(C and D)** anti-*Tb/Tco*MSP:303-314 peptide antibody and **(E and F)** anti-His antibody with rabbit anti-chicken secondary antibody conjugated to HRPO and visualised by ECL luminescence using the Syngene G:BOX. Lane M, ThermoScientific pre-stained marker; lane 1, unbound protein; lane 2-11, purified protein fractions 1-10 (1 ml each).

Purification of a smaller 30 kDa protein was also seen (Figure 4.14, panels, A, B, C, D, E and F, lane 2-11). This 30 kDa product was not observed in expression samples (Figure 4.13) and may be degradation of the smaller protein bands observed; however, a similar 27 kDa protein was detected in *T. brucei* parasite lysate (Figure 3.15, section 3.3.4). Since the purification was done using specific antibody-antigen interaction, it is suspected that the smaller product is some part of the full length MSP-C protein. The successful purification of MSP-C, from both species, allowed for further analysis of its multimerisation and its enzymatic properties

4.3.10 Analysis of the Multimerisation of r*Tb*MSP-C and r*Tco*MSP-C.

To assess the possible multimerisation of MSP-C from the two trypanosomal species, the protein was studied under non-denaturing conditions using non-reducing SDS-PAGE and molecular exclusion chromatography. Under non-reducing conditions, r*Tb*MSP-C and r*Tco*MSP-C appear as protein multimers of larger than 170 kDa (Figure 4.15, panel A). This larger protein was detected by the anti-*Tb/Tco*MSP:303-314 antibodies confirming its identity (Figure 4.15, panel B).

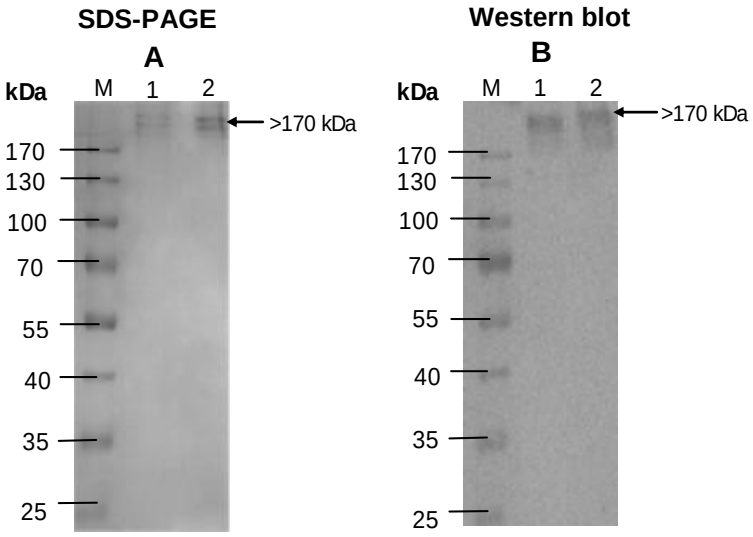


Figure 4.15 Analysis of r*Tb*MSP-C and r*Tco*MSP-C under non-denaturing conditions. (A) Non-reducing SDS-PAGE (10%) and **(B)** western blot using anti-*Tb/Tco*MSP:303-314 peptide antibodies, rabbit anti-chicken secondary antibody conjugated to HRPO and visualised by ECL luminescence using the Syngene G:BOX. Lane M, ThermoScientific pre-stained marker; lane 1, r*Tb*MSP-C and lane 2, r*Tco*MSP-C. Arrows show multimerisation of protein.

To confirm this data, molecular exclusion chromatography was performed and showed that these proteins aggregate and elute at the void volume of the column (Figure 4.16, panel A and C). The elution peak from molecular exclusion chromatography was subjected to reducing SDS-PAGE which showed the expected 71 kDa monomer (Figure 4.16, panel B

and D). There were, however, smaller degradation products present in the sample as well (Figure 4.16, panel B and D). The Sephacryl S200 column has a separation range of 5-250 kDa, therefore, the enzyme aggregate is expected to be greater than 250 kDa. Since the multimerisation of the protein was disrupted by denaturation, it suggests that the aggregation is dependent on disulfide bonds (or other covalent bonds) between the protein molecules.

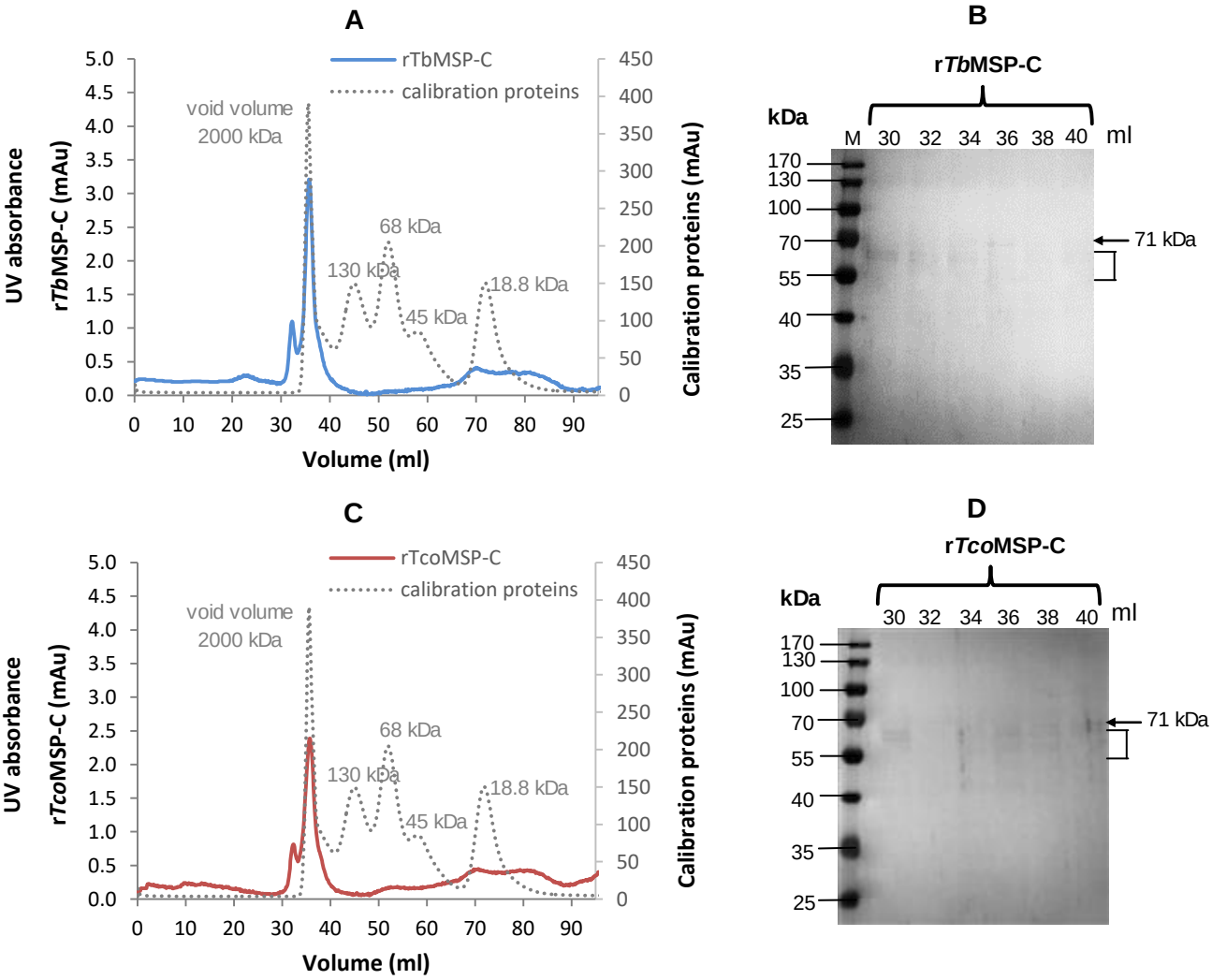


Figure 4.16: Molecular exclusion chromatography of r*Tb*MSP-C (A and B) and r*Tco*MSP-C (C and D). (A and C) Elution profile of r*Tb*MSP-C (blue) and r*Tco*MSP-C (orange) from an S200 sephacryl molecular exclusion chromatography column. Elution profile of calibration proteins is included (grey). **(B and D)** Reducing SDS-PAGE (10%) showing the presence of 71 kDa protein (arrow). Smaller degradation products are also present (square box).

4.3.11 Hydrolysis of H-Suc-Leu-Tyr-AMC Peptide Substrate

The enzyme activity of r*Tb*MSP-C and r*Tco*MSP-C was determined by assessing the hydrolysis of the H-Suc-Leu-Tyr-AMC peptide substrate. It has been shown that

leishmanolysin recognises -Leu/Ala-Tyr↓Leu-Lys-Lys- and a similar recognition site was used here.

Both r*Tb*MSP-C and r*Tco*MSP-C proteins were able to hydrolyse the peptide substrate as evidenced by the release of fluorescence over time (Figure 4.17). A maximum fluorescence signal, of 30000-40000, was seen for lower concentration of protein (0.5 and 1 µg/ml) after 30 min incubation with the substrate. In contrast, a lower fluorescence of 15000-25000 was observed at higher protein concentration (5 µg/ml). This is likely due to enzymatic competition for the small amount of substrate in the presence of large amount of enzyme. The fluorescence released by 0.5 µg/ml of protein showed an increase in fluorescent intensity from 20000 after 20 min to 30000 after 30 min. A similar trend is observed for 1 µg/ml and 5 µg/ml of protein with maximum overall fluorescence obtained from r*Tco*MSP-C, at 1 µg/ml, of 35000. This is indicative of peptide substrate over time.

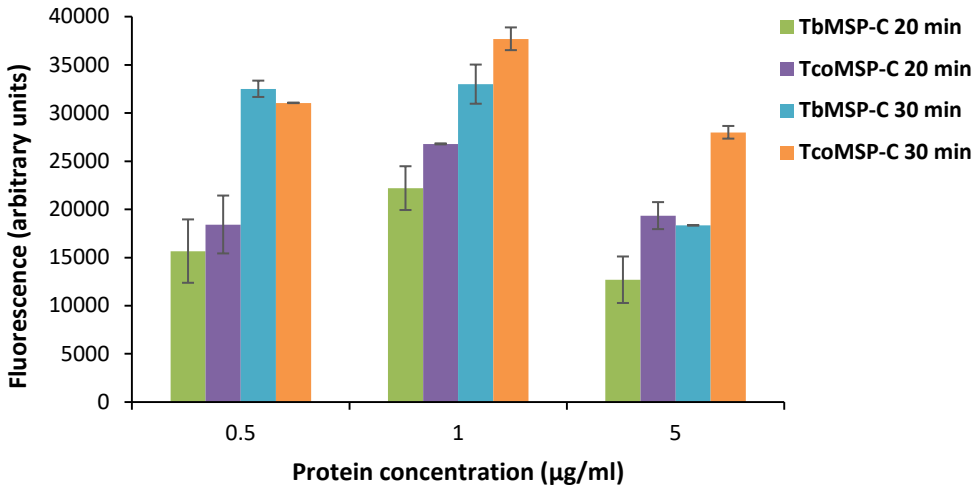


Figure 4.17: Hydrolysis of peptide substrate H-Suc-Leu-Tyr-AMC by r*Tb*MSP-C and r*Tco*MSP-C. Varying concentrations (0.5 µg/ml, 1 µg/ml, 5 µg/ml) of r*Tb*MSP-C (green and blue) and r*Tco*MSP-C (purple and orange) were used to assess the hydrolysis of 10 µM substrate by measuring fluorescence released ($Ex_{360\ nm}$ and $Em_{460\ nm}$) after incubation at 37^{o}C for 20 (green and purple) and 30 min (blue and orange) incubation times, using an Optima spectro-fluorometer. Results are the mean of duplicate readings with standard deviation.

The activity of both r*Tb*MSP-C and r*Tco*MSP-C were shown to be inhibited by the two metalloprotease inhibitors, EDTA and 1,10 phenanthroline. The fluorescent signal measured was shown to decrease as the concentration of inhibitor increased (Figure 4.18). The activity of r*Tb*MSP-C and r*Tco*MSP-C was inhibited by 1,10 phenanthroline best, with a larger reduction in fluorescent signal when comparing 10 µM to 1 mM. The fluorescent intensity decreases from ~55000 (10 µM) to ~25000 (1 mM).

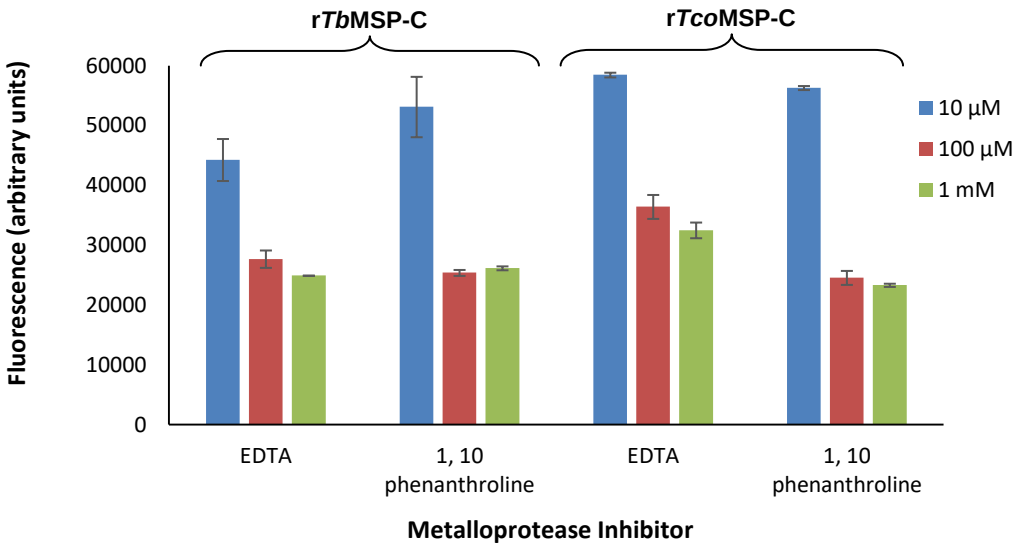


Figure 4.18: Inhibition of r*Tb*MSP-C and r*Tco*MSP-C by metalloprotease inhibitors, EDTA and 1,10 phenanthroline. r*Tb*MSP-C and r*Tco*MSP-C (0.5 µg/ml) was incubated with varying concentrations (0.01-1 mM) of metalloprotease inhibitor (EDTA and 1,10 phenanthroline) for 10 min at 37^{o}C, after which, 7.5 µM H-Suc-Leu-Tyr-AMC peptide substrate was added. Fluorescence intensity ($Ex_{360\ nm}$ and $Em_{460\ nm}$), after incubation at 37^{o}C for 10 min, was read using an Optima spectro-fluorometer. Results are the mean of duplicate readings with standard deviation.

To assess the pH optimum of r*Tb*MSP-C and r*Tco*MSP-C, activity was determined from pH 2 to pH 11 (Figure 4.19). It can be seen that both r*Tb*MSP-C and r*Tco*MSP-C enzymes prefer an acidic pH to that of a neutral one and basal signal is detected in alkaline pH of less than 20000. r*Tco*MSP-C favoured a more acidic pH of 4.5 (Figure 4.19, panel B), compared to r*Tb*MSP-C of pH 5 (Figure 4.19, panel A).

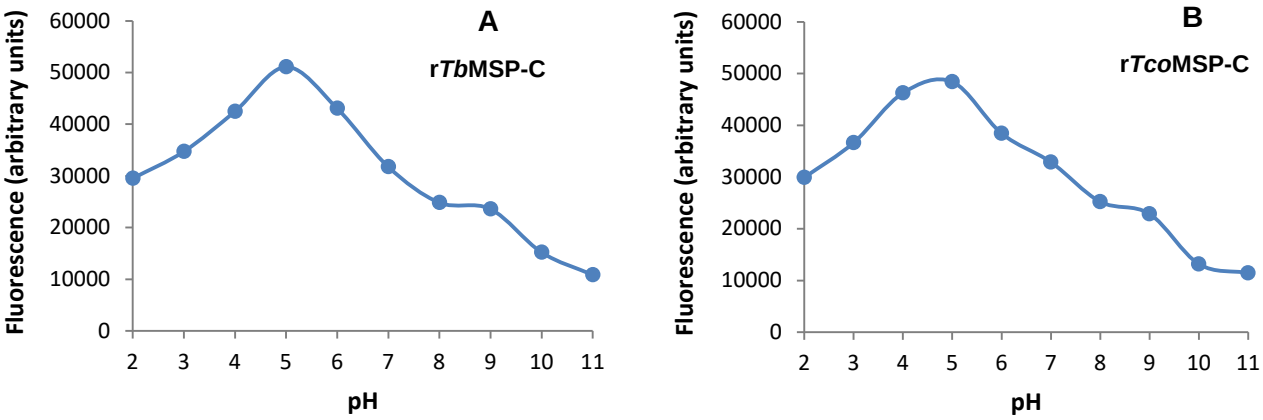


Figure 4.19: pH profile of r*Tb*MSP-C (A) and r*Tco*MSP-C (B). Protein (0.5 µg/ml) was incubated with AMT buffer (pH 2-11) at 37^{o}C for 10 min before addition of H-Suc-Leu-Tyr-AMC peptide substrate (7.5 µM). Enzyme activity was determined by measuring fluorescence released ($Ex_{360\ nm}$ and $Em_{460\ nm}$) after 10 min at 37^{o}C, using an Optima spectro-fluorometer. Results are the mean of triplicate readings.

4.3.12 Anti-peptide Antibody Specificity against r*Tco*MSP-C

To assess the detection ability of the anti-peptide antibodies against r*Tco*MSP-C, as was done for t*Tb*MSP-C (Section 4.3.5); purified r*Tco*MSP-C was probed with each antibody (Figure 4.20).

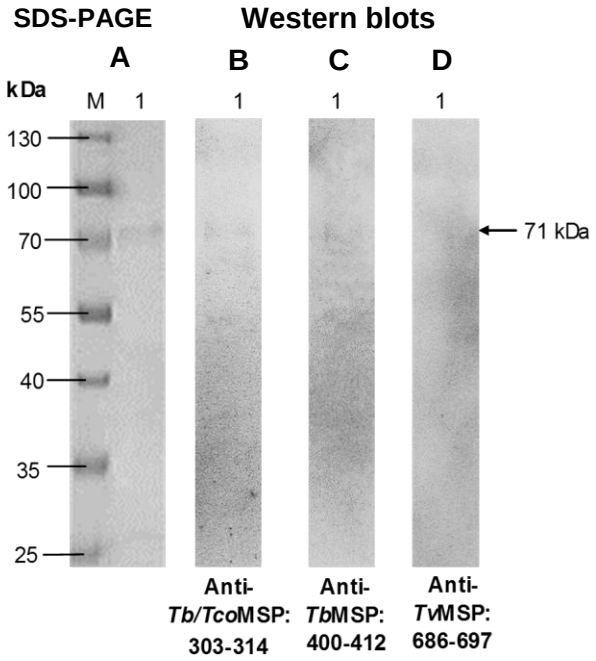


Figure 4.20: Detection of r*Tco*MSP-C with anti-peptide antibodies. (A) SDS-PAGE (10%) and western blots with **(B)** anti-*Tb/Tco*MSP:303-314, **(C)** anti-*Tb*MSP:400-412 or **(D)** anti-*Tv*MSP:686-697 peptide antibodies (2 µg/ml) as indicated. Rabbit anti-chicken secondary antibody conjugated to HRPO was used and visualised by ECL luminescence using the Syngene G:BOX. Lane M, ThermoScientific pre-stained marker; lane 1, purified r*Tco*MSP-C. Arrow indicates r*Tco*MSP-C.

r*Tco*MSP-C was detected by anti-*Tb/Tco*MSP:303-314 antibodies (designed to detect *Tb* and *Tco*MSP-**B**), at 71 kDa (Figure 4.20, panel B). The detection of r*Tco*MSP-C by this antibody was unexpected, since it only shares 42% identity with the peptide selected for antibody production. By comparison, t*Tb*MSP-C (Figure 4.8, panel B) and r*Tb*MSP-C (Figure 4.13, panel B1) were also detected by the antibody and also shares 42% identity with the peptide (Section 3.3.1, Figure 3.5). Anti-*Tb*MSP:400-412 antibodies (designed to detect ***Tb***MSP-C) showed faint detection of r*Tco*MSP-C (Figure 4.20, panel C), while no detection was seen by anti-*Tv*MSP:686-697 antibodies (Figure 4.20, panel D). *Tco*MSP-C shares 38% identity with the *Tb*MSP:400-412 peptide and 0% identity with the *Tv*MSP:686-697 peptide (Figure 3.5, Section 3.3.1).

4.3.13 Detection of r*Tco*MSP-C with Infected Cattle Sera

To determine whether MSP-C induces an immune response in cattle, *T. congolense* infected cattle sera were used to detect r*Tco*MSP-C in an ELISA format (Figure 4.21). Increasing concentrations of r*Tco*MSP-C as coating antigen resulted in an increase in absorbance signal with time, with 50 µg/ml protein giving the highest signal of 0.75 after 60 min. However, high absorbance signal was obtained for the non-infected sera control (Figure 4.21, orange), comparable with the infected sera (green). A comparison of the different protein concentrations with the signal released after 60 min shows higher signal in the presence of more protein (compare blue, red and green). The signal increases from 0.45 for 10 µg/ml of protein to 0.55 for 25 µg/ml of protein to 0.75 for 50 µg/ml of protein. This result suggests that there are serum antibodies present that detect r*Tco*MSP-C in infected sera and that during *T. congolense* infection MSP-C may induce an immune response in cattle.

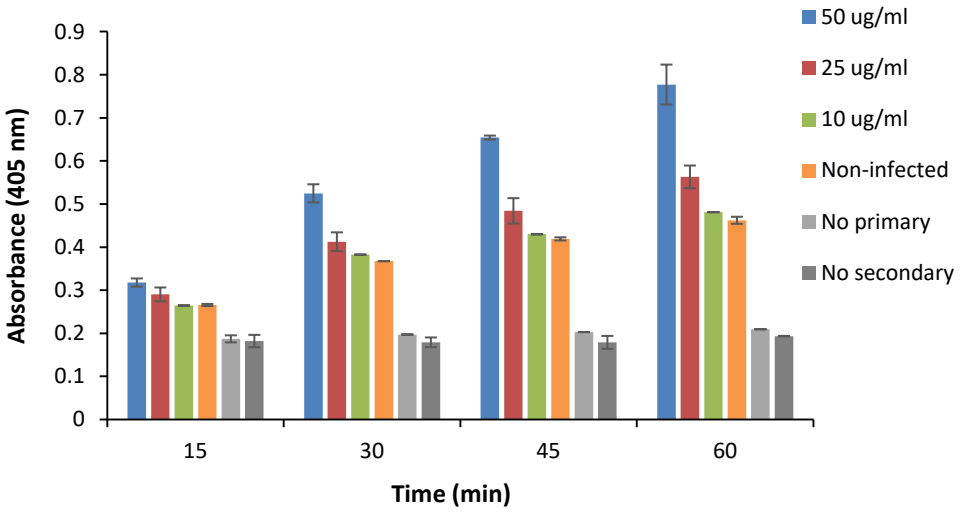


Figure 4.21: Detection of r*Tco*MSP-C with infected cattle sera. Decreasing protein concentration (50, 25 and 10 µg/ml) were used to coat a 96 well microtitre plate and infected or non-infected cattle sera (1:200) were used to detect the protein. Secondary anti-bovine IgG (1:5000) conjugated to HRPO was used to visualise detected protein in the presence of substrate ABTS/H_2O_2. Absorbance was read at 405 nm at 15 min intervals for one hour using an Optima spectro-fluorometer. Results are the mean of duplicate readings with standard deviation. Controls are shown in grey.

4.4 Discussion

The results in this chapter reported the analyses of the MSP-C protease from both *T. brucei* and *T. congolense.* The *TbMSP-C* gene was detected in both Lister 427 and LiTat 1.3 strains of *T. brucei* confirming previous results that showed that these proteases are conserved across species, genera and families of parasite protozoa [(Etges, 1992; El-Sayed and Donelson, 1997a; Cuevas *et al.*, 2003; Jesudhasan *et al.*, 2007; Marcoux *et al.*, 2010; Oladiran and Belosevic, 2012) and Chapter 2].

After the PCR, cloning and expression of the *TbMSP-C* gene, numerous truncated protein products resulted. It is known that when expressing eukaryotic genes within prokaryotes, codon bias may result. Codon bias is caused when a codon present within the cloned gene is used by *E. coli* at a rate of 1% or less [rare codon; (Kane, 1995)]. This may result in truncation of the expressed protein or incorrect amino acid incorporation (Rosano and Ceccarelli, 2014), due to the fact that tRNA of their amino acid are less abundant (Buchan and Stansfield, 2007). During protein synthesis, if many rare codons are present within the cloned gene, the tRNA molecules can become depleted, thus resulting in ribosomal pausing and subsequent disconnection of the RNA strand (Buchan and Stansfield, 2007). By using rare codon predictors (http://nihserver.mbi.ucla.edu/RACC/; http://www.genscript.com/cgi-bin/tools/rare_codon_analysis), the *TbMSP-C* gene was analysed for the presence of rare codons. Both predictors revealed between 11-38 rare codons. This is substantial evidence that explains the presence of the numerous truncated forms of the protein seen during expression.

It was because of this truncated expressed protein as well as the presence of mutations within the cloned gene, the unexplained purification of a smaller 70 kDa protein (expected a 92 kDa fusion protein) and the activity of a 55 kDa protein, that the synthesis of *TbMSP-C* and *TcoMSP-C* constructs with *E. coli* optimised genes were considered. These constructs were used to express and purify full-length MSP-C, of 71 kDa, from both species. The purified protein still contained the proregion; hence the larger size than when compared to MSP-C in parasite lysate (Figure 3.15, Section 3.3.4). Purification of the protein was performed using anti-*Tb/Tco*MSP:303-314 peptide antibodies coupled to a hydrazide resin. It was unexpectedly seen that both r*Tb*MSP-C and r*Tco*MSP-C were detected and further purified with this antibody. As mentioned in chapter 3 (Section 3.4), it was accepted that 33% identity in the selected MSP peptide epitopes would be the minimum requirement for cross-reactivity (Benjamin and Perdue, 1996; Dougan *et al.*, 1998). Similarly, *Tb*MSP-C and *Tco*MSP-C share 42% identity with *Tb/Tco*MSP:303-314 peptide and the cross-reactivity of this antibody was not unforeseen. Furthermore, *Tv*MSP-C also shares 42% identity with the peptide; therefore, this anti-peptide antibody would be expected to detect the *T. vivax* species as well. However, this antibody was unable to detect native MSP-C in neither *T. brucei* nor *T. congolense* parasite lysates (Figure 3.15, section 3.3.4).

The anti-*Tb*MSP:400-412 peptide antibody showed a low level detection of r*Tco*MSP-C (with ECL only) but a high level of detection of t*Tb*MSP-C, although this cross-reactivity was expected, since this antibody showed detection of both native *Tb*MSP-C and *Tco*MSP-C in parasite lysate (Figure 3.15, section 3.3.4). Anti-*Tv*MSP:686-697 peptide antibodies showed no detection of either t*Tb*MSP-C or r*Tco*MSP-C, and remains a *T. vivax* specific antibody.

The possible multimerisation of MSP-C was assessed and it was found that the enzyme from both *T. brucei* and *T. congolense* was shown to be greater than 170 kDa on a non-reducing SDS-PAGE and greater than 250 kDa in molecular exclusion chromatography. This is evidence that the protein forms multimeric complexes. At a size of ~70 kDa for each monomer, the multimer comprises of 3.5 monomers or greater. Since the multimer could be visualised within a non-reducing SDS-PAGE gel and there is evidence that the range for a 10% bis-Tris polyacrylamide gel is 20-300 kDa (Spectra Multicolor High Range Protein Ladder example), this suggests that MSP-C is present as a tetramer (280 kDa) or pentamer (350 kDa) but is unlikely larger than this. Leishmanolysin has been shown to be active as a monomer or dimer while oligomeric forms have been detected in parasite membranes (Soteriadou *et al.*, 1988). Furthermore, other metalloproteases like meprin and aminopeptidase, have been detected as multimers as well (Bertenshaw *et al.*, 2003; Appolaire *et al.*, 2014). Interestingly, GPI-PLC, an enzyme that has a similar function to MSP-B in *Trypanosoma*, has been shown to form tetramers (Armah and Mensa-Wilmot, 2000).

The MSP-C enzyme showed activity using the H-Suc-Leu-Tyr-AMC peptide substrate and this activity was inhibited by the metalloprotease inhibitors: EDTA and 1,10 phenanthroline as expected (Chaudhuri *et al.*, 1989; Bangs *et al.*, 1997; Cuevas *et al.*, 2003; Grandgenett *et al.*, 2007). The detection of enzyme activity, in the presence of the proregion, was not unexpected since leishmanolysin was shown to have latent activity in the presence of its proregion (Macdonald *et al.*, 1995) and it was shown here that t*Tb*MSP-C had both a fully active 55 kDa form and a latently active 70 kDa form. Further work to assess the effects of the removal of the proregion is required. The removal of the proregion is possible with chemicals such as APMA (4-Aminophenylmercuric acetate) or organomercurials (Stricklin *et al.*, 1983; Van Wart and Birkedal-Hansen, 1990; Macdonald *et al.*, 1995; Shapiro *et al.*, 1995; Galazka *et al.*, 1996), which, disrupt the cysteine switch mechanism suspected to be employed by these surface metalloproteases for inhibition (Macdonald *et al.*, 1995). The testing of different peptide substrates to determine P-site preference of these enzymes will also be considered in future studies.

The pH optimum for r*Tco*MSP-C was pH 4.5 while for r*Tb*MSP-C was pH 5. It has been reported that leishmanolysin is an acidic protease (Chaudhuri and Chang, 1988; Chaudhuri *et al.*, 1989). Conversely, literature has shown that this enzyme possesses neutral to alkaline pH optima (Etges *et al.*, 1986; Ip *et al.*, 1990). Later, it was shown that the pH optimum for leishmanolysin is substrate dependent (Tzinia and Soteriadou, 1991). Furthermore, the soluble form of leishmanolysin was also shown to be active at acidic pH as well (Ilg *et al.*, 1993), similar to MSP-C.

Further assessment of trypanosomal MSP-C activity with different substrates is required to determine whether this enzyme has a pH optimum at a neutral pH as well. However, considering the possibility that the enzyme may be secreted into the host blood stream (Marcoux *et al.*, 2010), activity at physiological pH (7.2-7.4), for certain substrates, is probable. Substrates found within the host such as extracellular matrix components like collagen and fibronectin should be assessed for hydrolysis by MSP-C at a range of pH-values.

To assess the possibility that *Tco*MSP-C induces an immune response in cattle infected with *T. congolense,* infected cattle sera were used to detect r*Tco*MSP-C in an ELISA, as was done before for other trypanosomal proteases (Mendoza-Palomares *et al.*, 2008; Pillay *et al.*, 2013; Mo *et al.*, 2014; Fleming *et al.*, 2016). Successful detection of the recombinant protein by sera was seen. However, non-infected sera showed high levels of non-specific detection as well. Whether this is due to the presence of high concentrations of protein and antibodies that resulted in protein-protein interaction or due to the presence of anti-*Tco*MSP-C IgG within non-infected sera is uncertain. Further optimisation of the ELISA is required to maximise antigen detection but minimise non-specific background detection before determining the sensitivity and specificity of the MSP-C antigen as a diagnostic target. However, the requirement for diagnostic tests against trypanosomiasis, to be able to distinguish between current and post infection is a necessity. It would, therefore, be more beneficial to focus on an antigen detecting test rather than an antibody detecting test as is already available for different trypanosomiasis infections (Büscher *et al.*, 2014; Birhanu *et al.*, 2015; Jamonneau *et al.*, 2015).

In conclusion, MSP-C from both *T. brucei* and *T. congolense*, were successfully expressed and purified and were subjected to further analyses. The enzymes were shown to form multimers under non-denaturing conditions and were shown to have acidic pH optima. Furthermore, it is suspected that MSP-C is able to induce an immune response during trypanosome infection in cattle. A number of additional analyses are required to further understand the activity and processing of these two enzymes.

Chapter 5

General Discussion

Human African trypanosomiasis (HAT) and animal African trypanosomiasis (AAT) are two of the many neglected tropical diseases on the continent negatively affecting health, agriculture and the economy. These diseases are caused by protozoan parasites particularly *Trypanosoma brucei, T. congolense and T. vivax*. Trypanosomal parasites have a dual life cycle: within a mammalian host and within the tsetse fly vector (*Glossina* spp). Within its host, it is an extracellular parasite which multiplies in the blood causing disease symptoms and eventually death, if untreated.

The number of HAT cases are declining, which suggests that disease transmission is under control (Tiberti *et al.*, 2013). As infected populations decline; more sensitive diagnosis is required to accurately diagnose HAT and eradicate it from the continent. On the other hand, bovine trypanosomiasis (a form of AAT) control has been less effective as the parasites have spread to previously trypanosome free areas (Majekodunmi *et al.*, 2013; Abdoulmoumini *et al.*, 2015; Odeyemi *et al.*, 2015).

Diagnostic tests are often required to follow the WHO ASSURED criteria [(**A**ffordable, **S**ensitive, **S**pecific, **U**ser-friendly, **R**apid and robust, **E**quipment-free and **D**eliverable to end-users); (Peeling *et al.*, 2006)] to be viable which is particularly applicable for HAT and AAT due to the environments where the diseases occur. Often these types of tests involve immunochemical reactions for the detection of antibodies (Büscher *et al.*, 2014; Birhanu *et al.*, 2015; Jamonneau *et al.*, 2015). Diagnostic tests that can distinguish between post and current infection are still necessary. Both diseases are still treated with drugs that have been on the market for over 25 years and are becoming ineffective with increasing drug resistance (Barrett *et al.*, 2007; Alsford *et al.*, 2013; Munday *et al.*, 2015). Therefore, there is a need to identify novel virulence factors and diagnostic targets to help in the fight against HAT and AAT.

Good diagnostic targets are those that are expressed in the host blood stage form of the parasite. These targets either need to be expressed at high enough levels to be detected or need to be immunogenic enough to induce an immune response in the host to allow detection of the resulting antibodies. Proteases from these parasites are often virulence factors (Grellier *et al.*, 2001; Lalmanach *et al.*, 2002; LaCount *et al.*, 2003; Marcoux *et al.*, 2010) that elicit high levels of antibodies.

One such factor may be the MSPs of trypanosomes. These zinc M8 metalloproteases were first found in leishmanial parasites (Russell and Wilhelm, 1986). Later, a number of putative gene and protein sequences of these MSPs were identified in trypanosome species, including *T. brucei* and *T. congolense* (El-Sayed and Donelson, 1997a; Marcoux *et al.*, 2010). MSPs are separated into five classes (-A, -B, -C, -D and -E) within trypanosome species. It has been shown by mRNA northern blot analyses that *T. brucei* contain classes A, B and C only (LaCount *et al.*, 2003), while, *T. congolense* contains all five classes (Marcoux *et al.*, 2010). These proteases are expressed in both the procyclic and blood stage form of the parasites where they have been identified as a virulence factor (LaCount *et al.*, 2003; Grandgenett *et al.*, 2007; Marcoux *et al.*, 2010). Some of these MSP genes are present as tandem repeats. (El-Sayed and Donelson, 1997a; LaCount *et al.*, 2003) which may indicate that they are required in high levels by the parasite while others are expected to be secreted.

This study set out to investigate MSPs from the African trypanosomes: *T. brucei, T. congolense* and *T. vivax* to determine their potential as diagnostic markers and highlight characteristics of the enzyme to deduce possible novel functions. The sequenced *T. vivax* genome was examined and putative MSP sequences were used to perform a number of bioinfomatic analyses. This subsequently allowed for the identification of conserved and non-conserved regions from the two MSP classes, MSP-B and -C, from the three trypanosome species. Epitope prediction software was used to select three immunogenic peptides from protein sequences of MSPs for the production of two forms of anti-peptide detection molecules: full IgY antibody and single chain variable fragments (scFvs). The anti-peptide antibodies, raised in chickens, were used to detect native MSP in parasite lysates. Furthermore, the *Tb*MSP-C gene was cloned and expressed in *E. coli*. However, due the numerous truncated proteins present during expression (t*Tb*MSP-C), r*Tb*MSP-C and r*Tco*MSP-C constructs were commercially synthesised. Both rMSP-C proteases were recombinantly expressed and purified and it was determined that these proteases form multimers under non-denaturing conditions and are optimally active at acidic pH. The sera of *T. congolense* infected cattle were shown to detect r*Tco*MSP-C in an ELISA format.

Before assessing these MSPs as diagnostic markers, various analyses of *T. vivax* MSPs were performed. Putative M8 metalloprotease genes were identified previously within the *T. vivax* genome (Jackson *et al.*, 2012). In the present study, nine putative metalloprotease sequences for *T. vivax* were found which were then phylogenetically classed into MSP-A, -C, -D and -E, based on their identity with other MSPs from *T. brucei, T. congolense, T. cruzi* and leishmanolysin from *L. major* (using BLASTp). The absence of a sequence resembling

MSP-B was unforeseen. These sequences were subjected to bioinfomatic analyses as a parallel to the work done by Marcoux *et al.* (2010) for *T. congolense*.

All the *MSP* gene sequences identified contained the HEXXHXXGF truncated active site motif unique to M8 metalloprotease and the third histidine ligand required for zinc binding ~60 amino acid residues downstream. *Tv*MSP-A and -E were predicted to be GPI-anchored to the surface of the parasite while *Tv*MSP-D is unlikely to be GPI-anchored as seen in other trypanosome species (Cuevas *et al.*, 2003; LaCount *et al.*, 2003; Marcoux *et al.*, 2010). It was unexpected that *Tv*MSP-C was predicted to be GPI-anchored to the surface of the parasite, unlike other MSP-C proteases (Marcoux *et al.*, 2010). There were a number of putative *T. vivax* MSP-C sequences identified during phylogenetic classification and one sequence was selected to represent *Tv*MSP-C. These putative sequences may represent a tandemly repeated gene within the *T. vivax* genome. However, no other MSP-Cs were tandemly repeated and all appeared as a single gene within the species' genome (Voth *et al.*, 1998; LaCount *et al.*, 2003; Marcoux *et al.*, 2010). GPI-anchor prediction (http://gpcr.biocomp.unibo.it/predgpi/) was done on the unused sequences (CCD18122.1, CCD19977.1, CCD21328.1, CCD20103.1) from the *T. vivax* genome and it was found that two were weakly probable and two were highly probable (like *Tv*MSP-C) to have an anchor. These characteristics of *Tv*MSP-C (and sequences CCD20103.1 and CCD21328.1) resemble MSP-Bs from other species and so it was hypothesised that '*Tv*MSP-B' may have higher identity with MSP-Cs from other species, rather than MSP-Bs as assumed. Sequences CCD18122.1 and CCD19977.1 would then represent the true MSP-C class from *T. vivax* which lacks the GPI-anchor as expected. Nevertheless, this is speculative and it is possible that *T. vivax* lost the MSP-B gene and uses MSP-A and -E as a replacement. These forms have been shown to have similar properties to MSP-B [work done here and (Marcoux *et al.*, 2010)]. To date, the *Tv*MSP-B gene remained unidentified.

Three-dimensional modelling of the four putative sequences was done to identify key characteristics that MSPs may have, in comparison to leishmanolysin. It was found that all *T. vivax* MSPs have three domains. The N-terminal domain contained helix A and helix B with the active site motif. The central domain contained a Met-turn, helix C and the third histidine ligand while the C-terminal domain contained two anti-parallel β-sheets leading into a β-sheet sandwich. All these structural elements are present in leishmanolysin (Schlagenhauf *et al.*, 1998), as well as a number of metzincin metalloproteases (Stocker *et al.*, 1993).

The three histidines were in optimal position for zinc binding. *Tv*MSP-A and -E were predicted to be compact enzymes with a conformation closely resembling that of

leishmanolysin, while *Tv*MSP-C and -D contain extra regions making their conformations unique. The propeptide region was also included during modelling to assess the possible mechanism of enzyme inhibition. *Tv*MSP-A and -E show propeptide folding as a lid over the active site as expected by analogy with leishmanolysin which would explain the latent activity seen for leishmanolysin in the presence of its proregion (Macdonald *et al.*, 1995). On the other hand, *Tv*MSP-C and -D proregion showed little active site interference which was attributed to the method the three-dimensional modelling software uses. This bioinformatic study reinforced what was known previously i.e. that these proteases are highly conserved across species, genera and families which indicates that they must have some significant function within all these protozoan parasites (Chaudhuri and Chang, 1988; Etges, 1992; Ramamoorthy *et al.*, 1992; El-Sayed and Donelson, 1997a; Cuevas *et al.*, 2003; Jesudhasan *et al.*, 2007; Marcoux *et al.*, 2010; Oladiran and Belosevic, 2012).

To establish whether these MSPs could be novel diagnostic markers, a number of preliminary studies were done. The first was the production of anti-peptide antibodies (whole IgY and scFvs) which were assessed for specificity against different MSPs from different trypanosomal species. The production of antibodies that can detect parasitic antigen is the basis for an antigen detection diagnostic test.

This was done by determining conserved and non-conserved regions within *T. brucei, T. congolense and T. vivax* MSP sequences. These regions were subjected to epitope prediction, using Predict7 software (Cármenes *et al.*, 1989), to determine peptides with high immunogenicity for their use in anti-peptide antibody production. Three peptides were selected to produce antibodies to detect 1) *Tb*MSP-B and *Tco*MSP-B (peptide *Tb/Tco*MSP:303-314; cross-species), 2) detect only *Tb*MSP-C (peptide *Tb*MSP:400-412) and 3) detect only *Tv*MSP-C (peptide *Tv*MSP:686-697). They were used to generate two-types of detection molecules; complete IgY anti-peptide antibodies and scFvs, with the capability of detecting their respective peptide in an ELISA format.

The ability of the IgY anti-peptide antibodies to detect native MSP was assessed by probing blood stage form *T. brucei* and *T. congolense* parasite lysate, in a western blot. Their specificity was further assessed on recombinantly expressed t*Tb*MSP-C, r*Tb*MSP-C and r*Tco*MSP-C. The anti-*Tb/Tco*MSP:303-314 peptide antibody was shown to detect a 63 kDa protein, in both *T. brucei* and *T. congolense* lysate, expected to be native MSP-B. Furthermore, these antibodies were able to detect t*Tb*MSP-C, r*Tb*MSP-C and r*Tco*MSP-C.

By comparison *Tv*MSP-C, like *Tb*MSP-C and *Tco*MSP-C, shares 42% sequence identity with the peptide and may also be detected with this antibody. This evidence suggests that the

anti-*Tb/Tco*MSP:303-314 peptide antibody is a true cross-trypanosomal species detecting antibody allowing detection of all three African *Trypanosome* species.

Anti-*Tb*MSP:400-412 showed detection of a smaller 53 kDa protein in both *T. brucei* and *T. congolense* parasite lysates expected to be active MSP-C with the proregion removed. As expected t*Tb*MSP-C was detected with anti-*Tb*MSP:400-412 peptide antibodies as well. But, this antibody also detected r*Tco*MSP-C. The cross-reactivity seen here was not expected since the peptide only shares 38% identity with the corresponding *Tco*MSP-C sequence. However, detection of r*Tco*MSP-C was only seen using ECL, which is a highly sensitive technique and, therefore, the antibody showed preference for t*Tb*MSP-C. This cross-reactivity of the anti-*Tb*MSP:400-412 peptide antibody implies that it is not species specific as originally suspected and the high level of identity between the MSPs made finding species specific antibodies difficult.

Anti-*Tv*MSP:686-697 peptide antibody detected no protein in either species lysate nor for t*Tb*MSP-C or r*Tco*MSP-C, as expected. Although the specificity of this antibody against *T. vivax* proteins remains to be tested, it appears to be *T. vivax* specific, showing no cross-reactivity with *T. brucei* or *T. congolense* proteins. It shares less than 10% identity with the *Tb*MSP and *Tco*MSPs. Therefore, for further species specific peptide design, percentage identity of 10% or less should be considered. A peptide specific for *T. brucei* MSPs needs to be redesigned and a peptide designed for *T. congolense* MSPs needs to be selected from the unique regions already identified.

The development of scFvs using the Nkuku® phage display library against each peptide was also done. It was possible to produce individual colonies using this technology that express scFvs that can detect the peptide antigen. However, further work to express and purify individual scFvs needs to be done (Van Wyngaardt *et al.*, 2004). The detection abilities and specificity of these scFvs also needs to be tested, as was done for the IgY anti-peptide antibodies. Similarly, a nanobody library was developed against *T. congolense* that is able, *in vitro*, to detect experimental infection and can be used as a test of cure (Odongo *et al.*, 2016). These antibody molecules can also be selected to produce detection fragments with improved thermal stability (Brockmann *et al.*, 2005; Miller *et al.*, 2010). Thermal stability studies, comparing IgY antibodies to scFvs, will be done to identify a better detection system in hot semi-arid climate, where trypanosomiasis is prevalent.

An alternative diagnostic test detects host antibodies within infected blood or sera (Büscher *et al.*, 2014; Birhanu *et al.*, 2015; Jamonneau *et al.*, 2015). To establish whether MSP antigens could be used as a diagnostic marker in this diagnostic test, *T. congolense* infected cattle sera were used to detect r*Tco*MSP-C in an ELISA format. The sera were able to detect

r*Tco*MSP-C, however, high background signal was obtained from the non-infected sera as well. This showed that *Tco*MSP-C induces an immune response during a trypanosome infection. Further optimisation of this assay is required before assessing the specificity and sensitivity of this antigen as a diagnostic target, as was done previously for *Tco*MSP-D, Tc38630, GM6 and p310 antigens (Marcoux *et al.*, 2010; Mochabo *et al.*, 2013; Pillay *et al.*, 2013; Fleming *et al.*, 2016). The ability of *T. congolense* infected sera to detect r*Tb*MSP-C will also be assessed, however, due to the high cross-reactivity seen for the anti-peptide antibodies, it is expected that cross-reactivity will occur here too.

Although, the potential of MSPs to be novel diagnostic markers has been shown in preliminary studies, further work as highlighted here, needs to be done. The highly conserved nature of these proteases, within *Trypanosoma,* makes them more suited as a cross-trypanosomal species detecting agent. More analyses to determine the possibility of species-specific detection is required, although anti-peptide antibodies against *T. vivax* MSP-C showed no cross reactivity with other species' MSPs.

These proteases have been identified as virulence factors, however, little is known about their function within trypanosome parasites, especially the MSP-C class. The latter class of protease remains of particular interest, since it is not detected on the surface of the parasite (Shimogawa *et al.*, 2015), but has been suggested to be secreted (Marcoux *et al.*, 2010) and gene expression of *Tb*MSP-C is only present within the blood stage form of the parasite (LaCount *et al.*, 2003). It was for these reasons that research into this class of MSP was done by performing initial characterisation on the recombinantly expressed protein.

The cloning and expression of *Tb*MSP-C yielded numerous truncated forms of the protein thought to result from the presence of rare codons within the sequence (Rosano and Ceccarelli, 2014). Therefore, in the present study it was called truncated (t)*Tb*MSP-C. Most importantly, expression of the cloned gene resulted in a full length fusion protein of 92 kDa (26 kDa GST tag + 66 kDa *Tb*MSP-C). However, after immunoaffinity purification using anti-*Tb*MSP:400-412 peptide antibodies, a 70 kDa protein was purified. Additionally, full-length 92 kDa and smaller protein products were also present. This purified t*Tb*MSP-C was assessed for activity using gelatin zymography. It was interesting to observe both latently active 70 kDa and fully active 55 kDa protein in the zymogram.

This is expected when the enzyme is in the presence or absence of its proregion, respectively, as seen for leishmanolysin previously (Macdonald *et al.*, 1995). The active form of the protein is expected to be present due to the numerous truncated proteins formed during expression and is unlikely due to an autocatalytic activity of this enzyme. Leishmanolysin shows no autocatalytic activity (*in cis* or *in trans*) and is believed to be

inhibited by its proregion via the cysteine switch mechanism (Macdonald *et al.*, 1995). This is where a conserved cysteine's sulfhydryl group binds the zinc cofactor, blocking the active site. All MSPs have a conserved cysteine residue (except *Tco*MSP-B1) and are suspected to be inhibited in a similar manner.

Due to a number of shortcomings in the expression and purification of t*Tb*MSP-C that could not be explained, *Tb*MSP-C and *Tco*MSP-C constructs were commercially synthesised for expression in *E. coli*. These two enzymes (recombinant (r)TbMSP-C and rTcoMSP-C) were successfully expressed and purified and additional characterisation was done. Unexpectedly both r*Tb*MSP-C and r*Tco*MSP-C formed multimers in non-reducing conditions which were deduced to be tetramers or pentamers. Older literature reports the identification of oligomeric forms of leishmanolysin (Soteriadou *et al.*, 1988), but no other evidence exists that this protease forms multimeric complexes. Further studies on the native protease are required to determine whether these multimeric complexes form naturally or whether this assembly of protein complexes is due to the recombinant expression environment of the proteases. Prokaryotic expression systems lack the ability to perform glycosylation which may be required for proper protein folding. This may result in soluble or insoluble protein aggregates (De Marco *et al.*, 2005).

These multimers did not affect the activity of the enzyme as both r*Tb*MSP-C and r*Tco*MSP-C were able to hydrolyse the H-Suc-Leu-Tyr-AMC peptide substrate. The selection of this substrate was based on the P-site preferences determined for leishmanolysin (Bouvier *et al.*, 1990). This activity was shown to decrease in the presence of increasing concentrations of metalloprotease inhibitors, EDTA and 1,10 phenanthroline. These activity assays were preliminary and further enzymes kinetics will be done. This includes the hydrolysis of a range of peptide substrates (H-Tyr-AMC, H-Leu-AMC, H-Ala-AMC, H-Phe-AMC, H-Arg-AMC, H-Met-AMC) to determine the P1 site preference for each enzyme and inhibition studies using a range of inhibitors (E64, AEBSF, pepstatin A, leupeptin, phosphoromidon, EDTA, 1,10 phenanthroline) specific for each protease family. The observed activity of r*Tb*MSP-C and r*Tco*MSP-C was in the presence of the proregion and is expected to be latent activity, as observed for leishmanolysin (Macdonald *et al.*, 1995) and t*Tb*MSP-C. Further work on enzyme activity of each protease without the proregion, is required. The employment of the cysteine switch mechanism of inhibition by MSPs can be disrupted by chemicals such as APMA and organomercurials (Stricklin *et al.*, 1983; Van Wart and Birkedal-Hansen, 1990; Macdonald *et al.*, 1995; Shapiro *et al.*, 1995; Galazka *et al.*, 1996).

The pH optimum for r*Tb*MSP-C was found to be pH 5, while, that of r*Tco*MSP-C was pH 4.5. This favouring of acidic pH was seen for leishmanolysin (Chaudhuri and Chang, 1988;

Chaudhuri *et al.*, 1989), however, was found to be substrate dependent (Tzinia and Soteriadou, 1991). This enzyme is also active at neutral to basic pH (Etges *et al.*, 1986; Ip *et al.*, 1990). It is suspected that these differences in pH optima are a regulatory mechanism of leishmanolysin which allows it to select for certain substrates in certain environments (Tzinia and Soteriadou, 1991; Ilg *et al.*, 1993). Assessment of the ability of r*Tb*MSP-C and r*Tco*MSP-C to cleave different substrates at different pH-values still needs to be done. However, activity at physiological pH is expected considering that MSP-C might be secreted into the host blood stream. The cleavage of substrates that may be present within the host, such as extracellular matrix components, will be assessed. This may help deduce possible functions of MSP-C, since other proteases within *Trypanosoma* also function to cleave extracellular matrix components (Bastos *et al.*, 2010; de Sousa *et al.*, 2010).

The MSP-C proteases from both *T. brucei* and *T. congolense* were shown to have similar characteristics. The multimerisation of the proteases was unique in that this has only been reported for MSP once before (Soteriadou *et al.*, 1988). Furthermore, the pH optimum at acidic pH suggests activity in a non-physiological environment although, the pH optimum of these proteases has been shown to be uniquely substrate dependent (Tzinia and Soteriadou, 1991). Further analyses on the activity of these proteases from *T. brucei* and *T. congolense* still need to be done, however, the detection of the smaller active form of native MSP-C, in parasite lysate in the present study, has provided evidence that these proteases are active within the parasite.

Additional studies that could be done on these proteases are described below. These include determining the cellular localisation of *Trypanosoma* MSP-B using confocal microscopy as was done previously with leishmanolysin from *L. braziliensis* (Cuervo *et al.*, 2008). Furthermore, the identification of MSP-C in the parasite culture medium would give evidence of a secreted protein as was shown for other secreted matrix metalloproteases within mouse and human cell lines (Lewis *et al.*, 2000; Duellman *et al.*, 2015). The identification of a secreted MSP-C protease has significant benefits since the diagnostic capability of these MSPs would benefit from a secreted antigen and the function of this protease within the host blood would be of interest.

To assess whether these proteases could be possible drug targets and, therefore, treatment options, they need to be essential genes required for parasite survival and life cycle progression. RNA interference (RNAi) studies have the ability to diminish mRNA levels of a specific targeted gene and hence protein levels within a cell.

Therefore, this method would be useful to assess whether MSP-B and -C are essential for the parasite as was demonstrated for *Tb*MSP-B (Grandgenett *et al.*, 2007). This study

showed that *Tb*MSP-B, as well as GPI-phospholipase C, are both essential for parasite differentiation from blood stage to procyclic forms within the host (Grandgenett *et al.*, 2007). The absence of these proteins within parasites also showed diminished parasite proliferation. Furthermore, immunisation of mice with rMSP-B or -C before parasite infection could give insight into the necessity of these proteases as was done before for *Tco*MSP-D (Marcoux *et al.*, 2010). Unexpectedly, the study showed increased susceptibility to parasitaemia of previously immunised mice. A better outcome to such a study would be increased mouse survival after parasite infection which would indicate the possible use of MSPs for treatment.

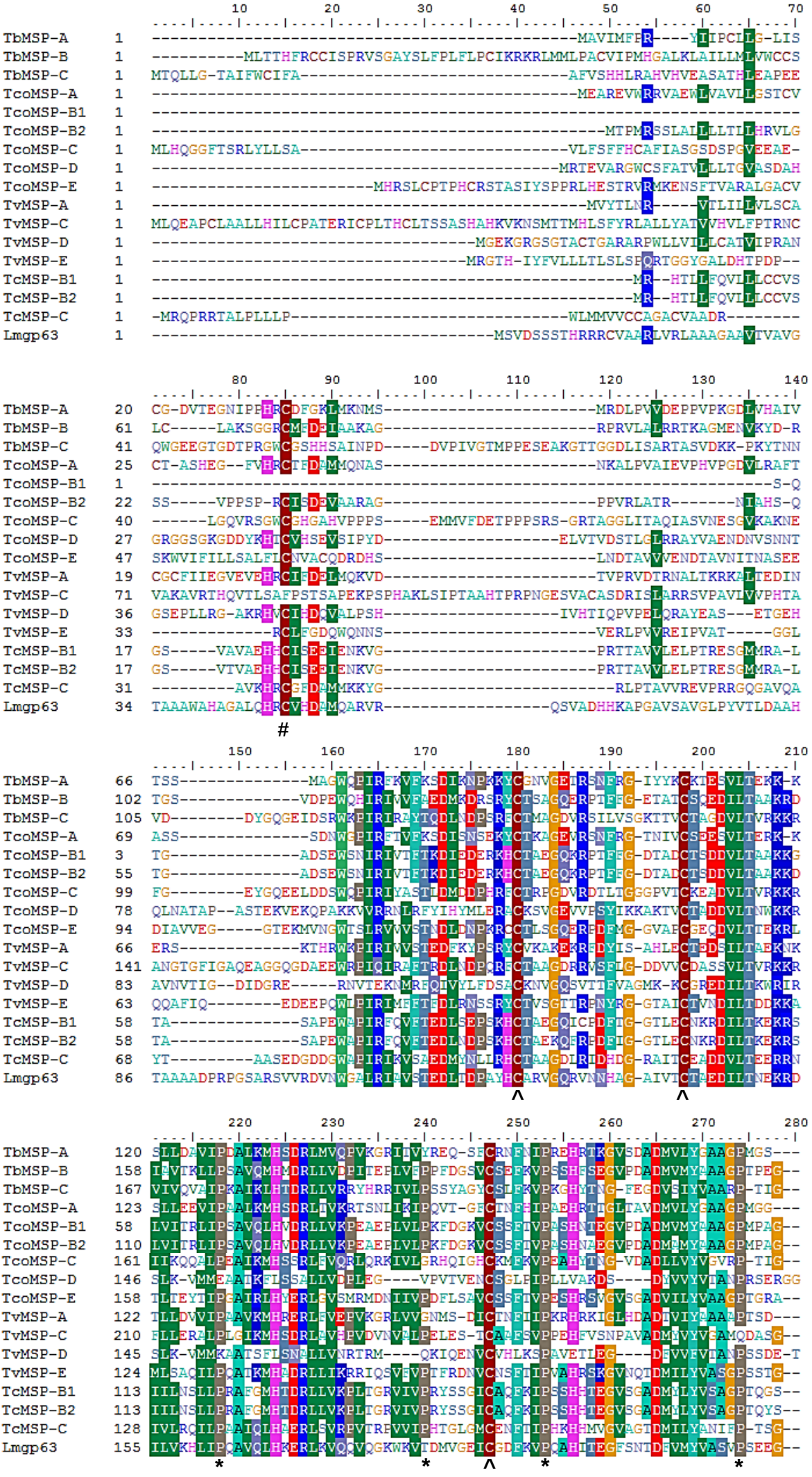
In conclusion, the MSPs from African trypanosomes are an interesting group of proteases which have been shown to possess unique characteristics as a group of proteases and across species and genera. They were shown to be novel diagnostic targets. However, the high level of identity between species MSPS resulted in cross-reactivity in their detection by anti-peptide antibodies and further work to produce species-specific detection molecules is required. Additionally, detection of the native antigen was shown in parasite lysate. The native antigen must be detected in infected sera or blood as well, to be useful for the development of a diagnostic test. The detection of active native MSP-C, from both *T. brucei and T. congolense*, was reported for the first time. The function of MSP-C remains significant due to their expected secretion by these parasites, directly into the mammalian host. We have shown that r*Tb*MSP-C and r*Tco*MSP-C form high molecular weight multimers and have acidic pH optima. Further assessment of their activity is required to determine possible functions within African trypanosomes.

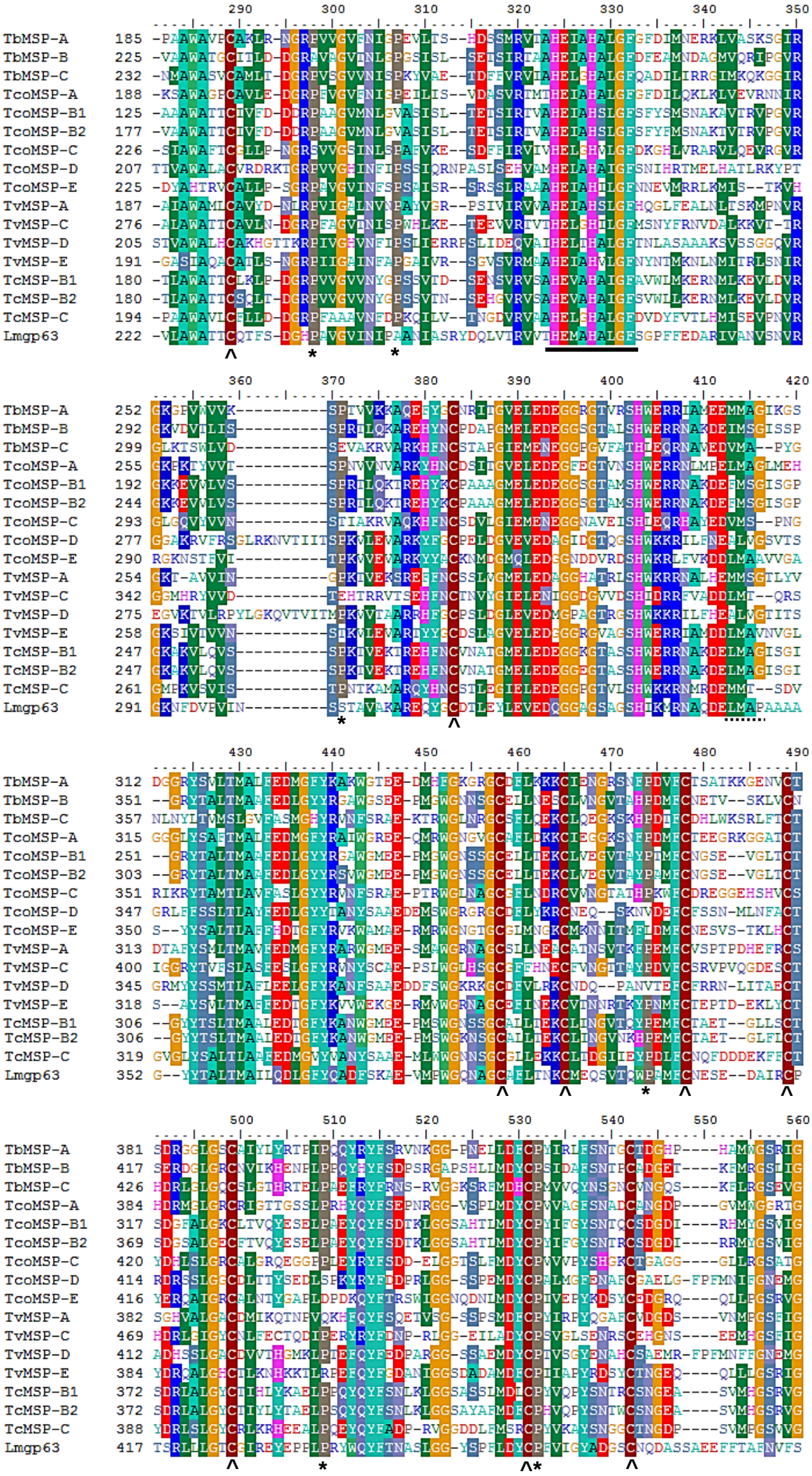
Appendix 1
Sequence Identities and Alignments

Table A1: Sequence identities of all Major Surface Proteases from *Trypanosoma brucei, T. congolense, T. vivax*, *T. cruzi* and leishmanolysin.

Sequence		TbMSP			TcoMSP						TvMSP				TcMSP (gp63)			
	Family	**A**	**B**	**C**	**A**	**B1**	**B2**	**C**	**D**	**E**	**A**	**C**	**D**	**E**	**B1 (Ia)**	**B2 (Ib)**	**C (II)**	**Lmgp63**
TbMSP	**A**	ID	31%	26%	48%	31%	32%	24%	22%	28%	36%	25%	19%	32%	36%	36%	29%	24%
	B	31%	ID	27%	32%	53%	56%	29%	19%	32%	31%	23%	18%	34%	44%	44%	34%	27%
	C	26%	27%	ID	29%	27%	28%	48%	19%	25%	25%	30%	17%	27%	28%	30%	32%	22%
TcoMSP	**A**	48%	32%	29%	ID	33%	33%	27%	23%	29%	32%	25%	20%	33%	36%	36%	33%	27%
	B1	31%	53%	27%	33%	ID	87%	27%	19%	31%	30%	24%	18%	35%	47%	46%	33%	28%
	B2	32%	56%	28%	33%	87%	ID	27%	20%	32%	32%	24%	19%	35%	49%	48%	33%	30%
	C	24%	29%	48%	27%	27%	27%	ID	19%	27%	24%	31%	17%	27%	28%	28%	33%	22%
	D	22%	19%	19%	23%	19%	20%	19%	ID	18%	21%	17%	42%	20%	21%	20%	20%	20%
	E	28%	32%	25%	29%	31%	32%	27%	18%	ID	25%	23%	18%	43%	35%	34%	29%	24%
TvMSP	**A**	36%	31%	25%	32%	30%	32%	24%	21%	25%	ID	22%	19%	28%	32%	33%	29%	24%
	C	25%	23%	30%	25%	24%	24%	31%	17%	23%	22%	ID	15%	21%	25%	25%	26%	20%
	D	19%	18%	17%	20%	18%	19%	17%	42%	18%	19%	15%	ID	18%	19%	19%	18%	20%
	E	32%	34%	27%	33%	35%	35%	27%	20%	43%	28%	21%	18%	ID	37%	37%	32%	24%
TcMSP (gp63)	**B1 (Ia)**	36%	44%	28%	36%	47%	49%	28%	21%	35%	32%	25%	19%	37%	ID	92%	35%	28%
	B2 (Ib)	36%	44%	30%	36%	46%	48%	28%	20%	34%	33%	25%	19%	37%	92%	ID	35%	28%
	C (II)	29%	34%	32%	33%	33%	33%	33%	20%	29%	29%	26%	18%	32%	35%	35%	ID	25%
	Lmgp63	24%	27%	22%	27%	28%	30%	22%	20%	24%	24%	20%	20%	24%	28%	28%	25%	ID

Yellow indicates highest identity for each protein. Blue indicates those that do not apply because these are within one class.


TbMSP-A
TbMSP-B
TbMSP-C
TcoMSP-A
TcoMSP-B1
TcoMSP-B2
TcoMSP-C
TcoMSP-D
TcoMSP-E
TvMSP-A
TvMSP-C
TvMSP-D
TvMSP-E
TcMSP-B1
TcMSP-B2
TcMSP-C
Lmgp63
#
^
^
*
*
^
*
*


TbMSP-A
TbMSP-B
TbMSP-C
TcoMSP-A
TcoMSP-B1
TcoMSP-B2
TcoMSP-C
TcoMSP-D
TcoMSP-E
TvMSP-A
TvMSP-C
TvMSP-D
TvMSP-E
TcMSP-B1
TcMSP-B2
TcMSP-C
Lmgp63

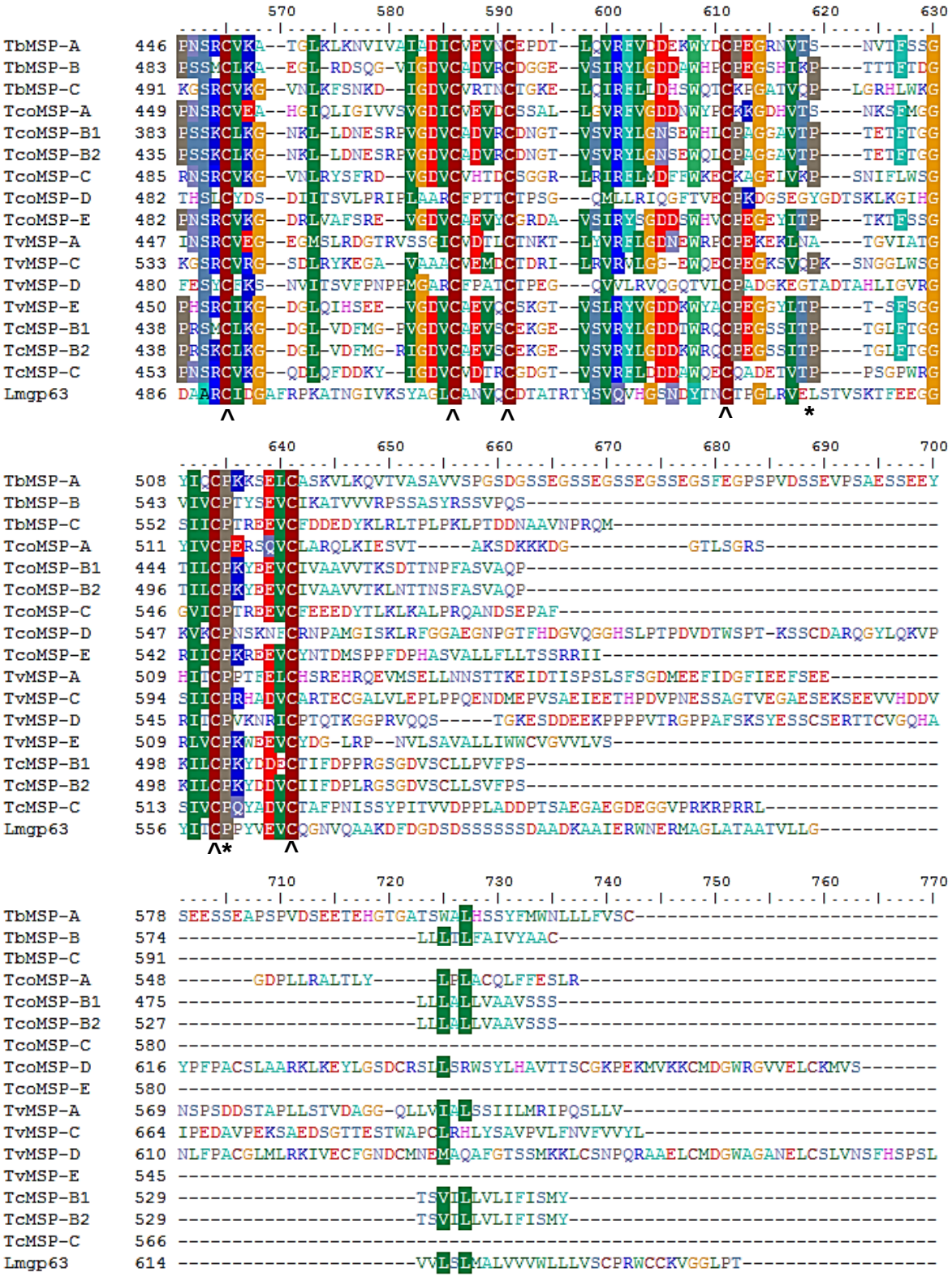


Figure A1: Amino acid sequence alignment of all Major Surface Proteases from *T. brucei, T. congolense, T. vivax, T. cruzi* and leishmanolysin. # denotes almost conserved cysteine with in the propeptide. ^ denotes 18 fully conserved cysteines. * denotes 12 semi-conserved prolines (11 or more sequences contain a proline at that position). Black line indicates truncated zinc binding motif HEXXHXXGF. Dotted line represents the Met-turn. Sequences obtained from NCBI and accession numbers are as follows; *Tb*MSP-A1 (AAB61263.1), *Tb*MSP-B (AAX79769.1), *Tb*MSP-C (AAO72980.1), *Tco*MSP-A (CCC95375.1), *Tco*MSP-B1 (CCD15541.1), *Tco*MSP-B2 (CCD15540.1), *Tco*MSP-C (CCC93446.1), *Tco*MSP-D (CCC95354.1), *Tco*MSP-E (CCC95375.1), *Tc*MSP-B1 (AAP22090.1), *Tc*MSP-B2 (AAP22091.1), *Tc*MSP-C (AAP22092.1) and *Lm*gp63 (AAC39120.1). *T. vivax* sequences included but discussed in chapter 2 *Tv*MSP-A (CCC52269.1), *Tv*MSP-C (CCD18177.1), *Tv*MSP-D (CCC53317.1), *Tv*MSP-E (CCC53338.1).

Appendix 2
Primary Structural Analyses

Table A2: Total number of amino acid residues and size of various Major Surface Proteases from *T. brucei, T. congolense, T. vivax*, *T. cruzi* and leishmanolysin.

Protein	Total number of amino acid residues	Size (kDa)
TbMSP-A1	622	68
TbMSP-B	587	64
TbMSP-C	591	66
TcoMSP-A	572	63
TcoMSP-B1	558	60
TcoMSP-B2	540	58
TcoMSP-C	580	64
TcoMSP-D	678	75
TcoMSP-E	580	63
TvMSP-A	608	68
TvMSP-C	706	77
TvMSP-D	695	75
TvMSP-E	545	60
TcMSP-B1	543	59
TcMSP-B2	543	60
TcMSP-C	566	62
Lmgp63	602	64

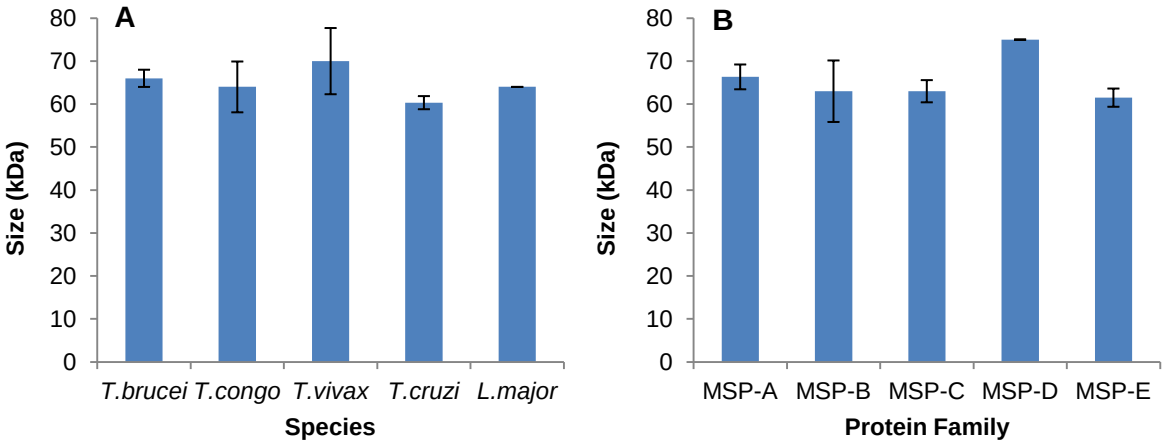


Figure A2: Comparison of sizes (kDa) of various MSP protein, between different species (A) or between different families (B).

Appendix 3
Examples of Quick 2D programme analyses

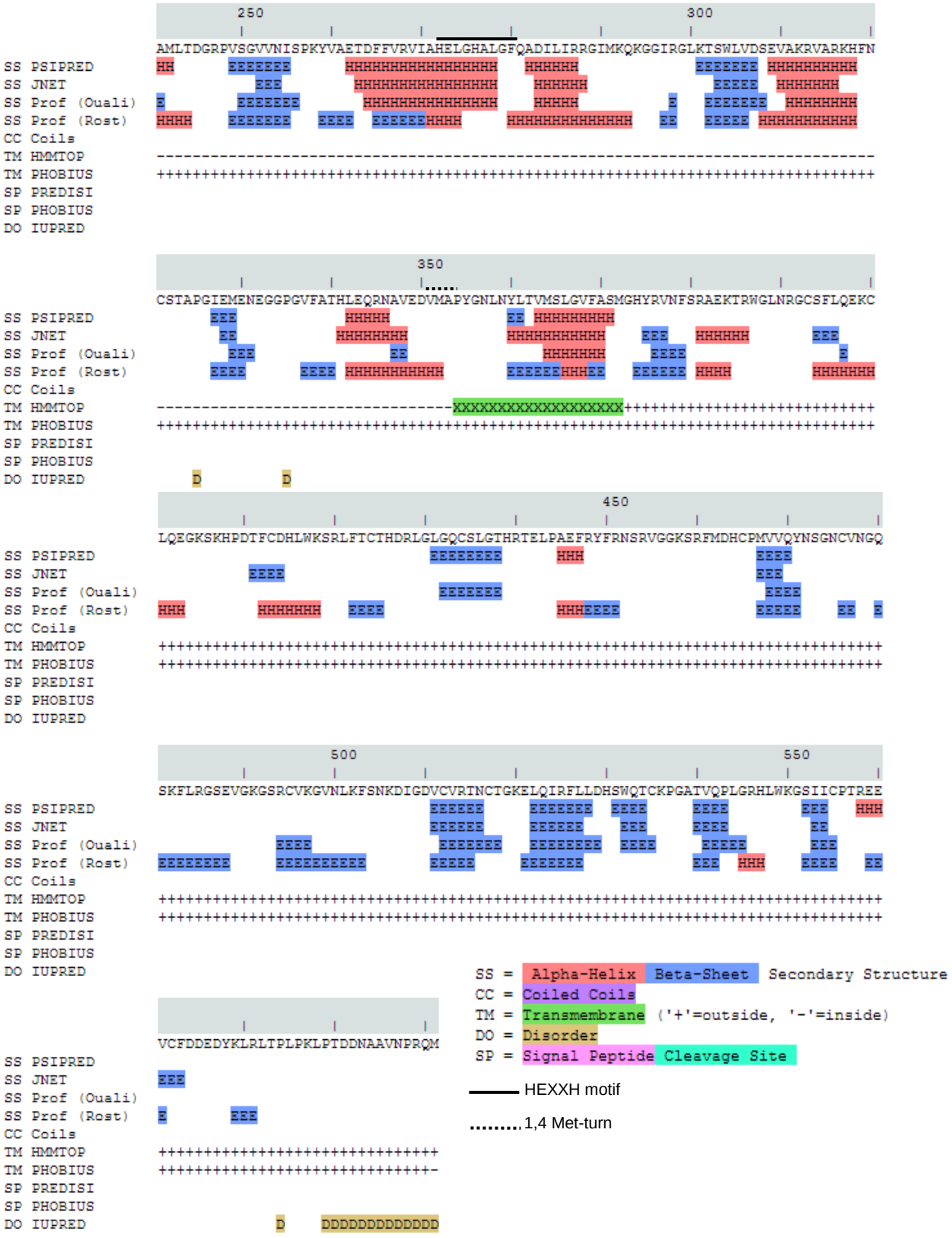


Figure A3: Secondary structure prediction of *Tb*MSP-C using Quick 2D software. (http://toolkit.tuebingen.mpg.de/quick2_d)

Appendix 4
PredGPI and DAS-TMfilter programmes analyses

Table A4: MSPs predicted to be GPI-anchored or contain a transmembrane domain.

Familiy of MSP	Accession Number	Probability of GPI-Anchor*	Trans-membrane Domain**
TbMSP-A1	AAB61263.1	Highly Probable	Unlikely
TbMSP-B	AAX79769.1	Probable	Likely
TbMSP-C	AAO72980.1	Not GPI-Anchored	Unlikely
TcoMSP-A	CCC95375.1	Probable	Likely
TcoMSP-B1	CCD15541.1	Highly Probable	Likely
TcoMSP-B2	CCD15540.1	Highly Probable	Likely
TcoMSP-C	CCC93446.1	Not GPI-Anchored	Likely
TcoMSP-D	CCC95354.1	Not GPI-Anchored	Unlikely
TcoMSP-E	CCC95375.1	Probable	Likely
TcMSP-B1	AAP22090.1	Highly Probable	Likely
TcMSP-B2	AAP22091.1	Highly Probable	Likely
TcMSP-C	AAP22092.1	Not GPI-anchored	Likely

*Using PredGPI: http://gpcr2.biocomp.unibo.it/gpipe/pred.htm

** Using DAS-TMfilter: http://mendel.imp.ac.at/sat/DAS/DAS.html

Appendix 5
*Tv*MSP Bioinfomatical Analyses: raw data

Table A5.1 Secondary structural analyses on identified TvMSP classes: comparison between four bioinformatical databases.

		Quick 2D	PredictProtein	PSIPred	Scratch Protein Predictor
***Tv*MSP-A**	Helix A	~ 115 - 140	118 - 140	117 - 139	~ 109 - 131
	Helix B	~ 220 - 235	220 - 233	224 - 233	~ 210 - 225
	HEXXH motif	Present		Present	Present
	Third His ligand	Present ~ 66 residues after 2nd His Ligand	-- Residues not show	Present ~ 64 residues After 2nd His ligand	Present ~ 64 residues after 2nd His ligand
	1,4 met-turn ~ 305 – 308	End of small helix	After helix: predicted to be buried	No secondary structure present	Forms four residue helix
	Helix C	~ 315 - 327	319 - 327	318 - 326	~ 309 - 318
	β-sheet predominant C-terminal domain	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures (domain 3 and 4)*	β-sheets predominate C-terminus of protein with few helical structures
		10[1]	11	10	9

	Total Number of Dominant Helices				
	Total Number of Dominant Sheets	15[1] (10 in C-terminal Domain)	31 (17 in C-terminal domain)[2]	24 (13 in C-terminal domain)	21 (12 in C-terminal domain)
	Domain Prediction	--	--	Domain 1: 1 - 126 Domain 2: 127 - 260 Domain 3: 261 - 349* Domain 4: 350 - 520*	Domain 1: 1 - 370 Domain 2: 371 - 470 Domain 3: 471 - 608
	TM protein	--	--	Helix: 585 - 605	56 % chance Non- 43 % chance α-helix- 0.5% chance β barrel-
	Extracellular Part	--	--	1 - 584	--
***Tv*MSP-C**	Helix A	~ 205 - 225 Note an extra helix is present at ~ 50	206 - 228 Note an extra helix is present at ~ 50	205 - 225 Note 2 extra helices at ~ 35 and ~ 45	~ 197 - 219 Note an extra helix is present at ~ 50
	Helix B	~ 305 - 322	310 - 324	308 - 323	~ 298 - 313

HEXXH motif	Present	Present	Present	Present
Third His ligand	Present ~ 64 residues after 2nd His ligand	-- Residues not show	~ 64 residues after 2nd His ligand	Present ~ 63 residues after 2nd His ligand
1,4 met-turn ~ 393 – 397	End of helix	Partially Buried	No secondary structure	Helix structure
Helix C	~ 403 - 414	409 - 414	352 - 359	~ 390 - 400
β-sheet predominant C-terminal domain	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures
Total Number of Dominant Helices	10[1]	11	10	15
Total Number of	21[1] (8 in C-terminal	16 (7 in C-terminal	21 (in C-terminal	21 (8 in C-terminal

	Dominant Sheets	domain)	domain)[2]	domain)	domain)
	Domain Prediction	--	--	Domain 1:1 - 118 Domain 2: 119 - 206 Domain 3: 207 - 611	Domain 1: 1 - 456 Domain 2: 457 - 618 Domain 3: 619 – 706
	TM Protein	--	--	Helix: 49 - 70 688 - 704	76 % chance Non- 22 % chance α-helix- 1.4 % chance β barrel-
	Extracellular part	--	--	1 - 48; 705 – 706	--
***Tv*MSP-D**	Helix A	~ 140 - 160	140 - 159	140 - 159	~ 139 - 159
	Helix B	~ 340 - 354	242 - 251	242 - 251	~ 234 - 248
	HEXXH motif	Present	Present	Present	Present

Third His ligand	Present 75 residues after 2nd His Ligand	--	Present 70 residues after 2nd His Ligand	Present 75 residues after 2nd His Ligand
1,4 met-turn	Not Present	Not Present	Not Present	Not Present
Helix C	~ 351 - 358	295 - 305	295 - 359	343 - 353
β-sheet predominant C-terminal domain	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures	β-sheets predominate C-terminus of protein with few helical structures
Total Number of Dominant Helices	10[1]	8	8	13
Total Number of Dominant Sheets	18[1] (7 in C-terminal Domain)	25 (13 in C-terminal Domain)[2]	21 (10 in C-terminal Domain)	21 (6 in C-terminal Domain)
Domain Prediction			Domain 1: 1 - 365 Domain 2: 366 - 565 Domain 3: 566 – 695	Domain 1: 1 - 51 Domain 2: 52 - 562 Domain 3: 563 – 695
TM Protein			244 - 259 Pore-lining	\ 76 % chance Non- 22 % chance α-helix-

				helix	1.5 % chance β barrel
	Extracellular			1 - 243	
***Tv*MSP-E**	Helix A	~ 120 – 139	123 - 139	119 - 138	~ 124 – 143
	Helix B	~ 225 – 235	268 – 278	224 – 236	~ 229 – 242
	HEXXH motif	Present	Present	Present	Present
	Third His ligand	Present ~ 65 residues after 2nd His Ligand	--	Present ~ 65 residues after 2nd His Ligand	Present ~ 65 residues after 2nd His Ligand
	1,4 met-turn ~ 310 – 313	Beginning of helical structure	No secondary structure-exposed	No secondary structure	End of helical structure
	Helix C	~ 320 - 330	327 - 330	320 - 329	313 - 324
	β-sheet	β-sheets predominate	β-sheets predominate	β-sheets predominate	β-sheets predominate

predominant C-terminal domain	C-terminus of protein with few helical structures	C-terminus of protein with few helical structures	C-terminus of protein with few helical structures	C-terminus of protein with few helical structures
Total Number of Dominant Helices	7[1]	6	7	8
Total Number of Dominant Sheets	19[1] (9 in C-terminal Domain)	28 (11 in C-terminal Domain)[2]	20 (10 in C-terminal Domain)	19 (9 in C-terminal Domain)
Domain Prediction			Domain 1: 1 - 110 Domain 2: 111 - 208 Domain 3: 209 - 292 Domain 4:293 - 545	Domain 1: 1 - 372 Domain 2: 373 - 504 Domain 3: 505 - 545
TM Helix			543 - 547 Helix	75 % chance Non- 23 % chance α- helix- 0.98 % chance β barrel-
Extracellular			1 - 526	

[1]at least 3 of the 4 programs used by quick 2D indicate a helix [2]Predicted sheets of 2 residues or less not included

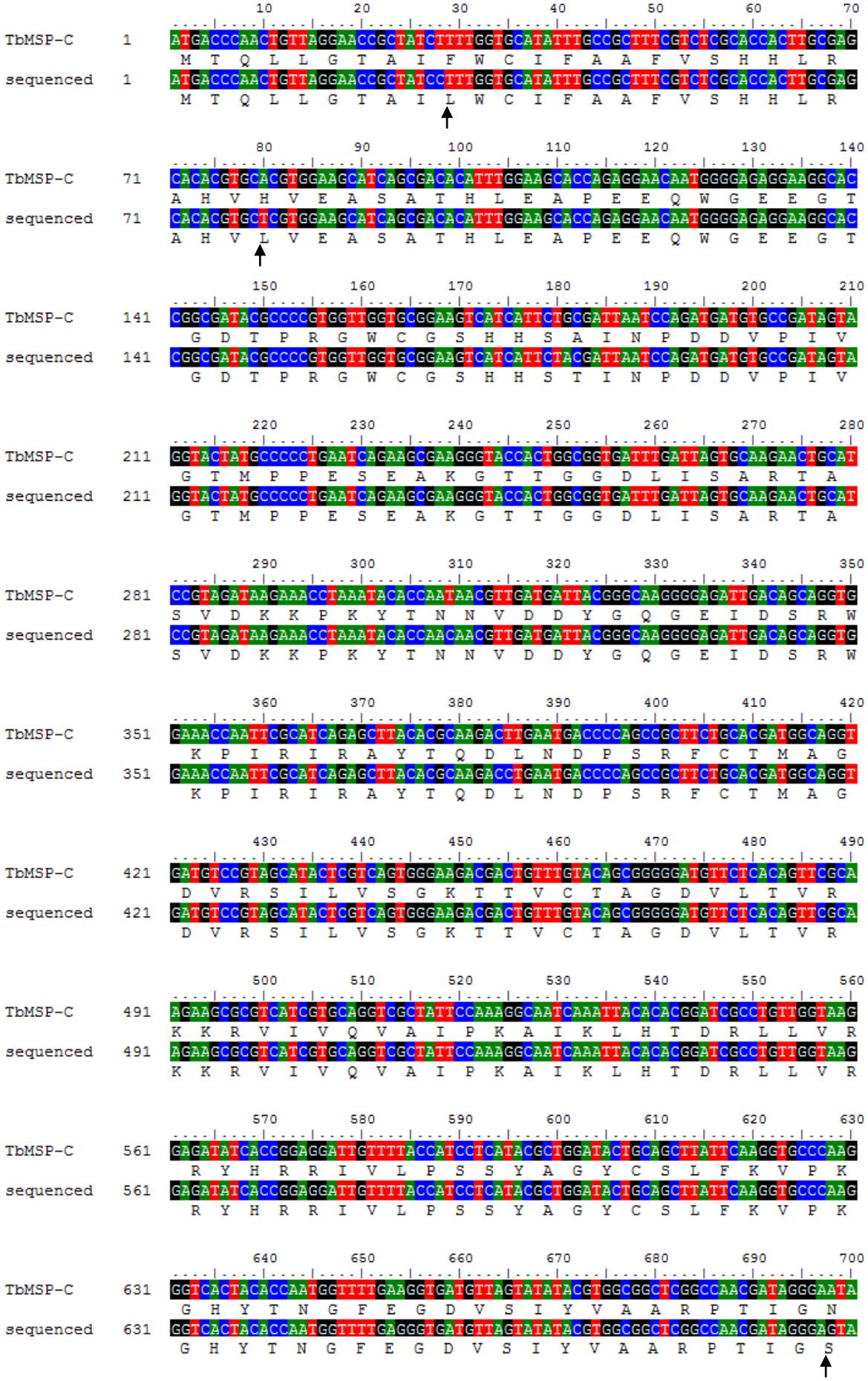
Table A5.2: Secondary Structure Analysis: PredGPI and DAS-TMfilter programs.

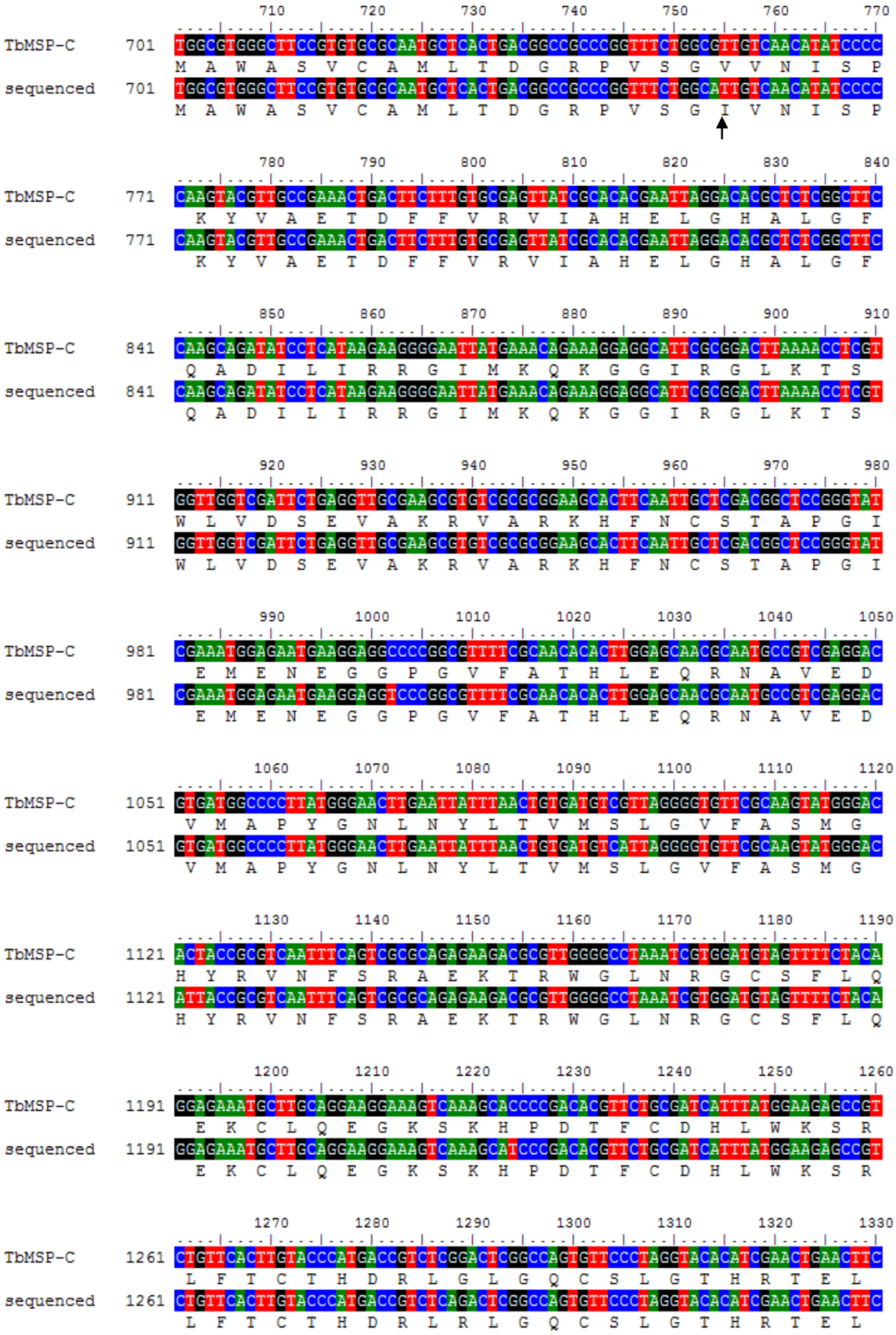
Familiy of MSP	Accession Number	Probability of GPI-Anchor	Trans-membrane Domain
TvMSP-A	CCC52269.1	Highly Probable	Likely
TvMSP-C	CCD18177.1	Highly Probable	Likely
TvMSP-D	CCC53317.1	Not GPI-anchored	Unlikely
TvMSP-E	CCC53338.1	Probable	Likely

Table A5.3: Disulfide bond prediction by Scratch Protein Predictor.

	Bond	Cys 1 Position	Cys 2 Position		Bond	Cys 1 Position	Cys 2 Position
TvMSP-A	1	92	109	TvMSP-D	1	115	132
	2	157	195		2	12	27
	3	380	390		3	410	420
	4	349	359		4	176	213
	5	421	432		5	657	668
	6	451	470		6	451	462
	7	512	519		7	525	548
	8	475	493		8	598	604
	9	21	33		9	382	400
TvMSP-C	1	245	284		10	615	630
	2	180	197		11	484	508
	3	456	477	TvMSP-E	1	160	199
	4	537	554		2	382	392
	5	7	16		3	519	538
	6	559	576		4	94	111
	7	23	28		5	476	494
	8	507	518		6	435	471
	9	424	436		7	352	359
	10	597	609		8	424	454
	11	70	118				

Appendix 6
Sequence of Cloned t*Tb*MSP-C Gene




710 720 730 740 750 760 770
TbMSP-C 701 TGGCGTGGGCTTCCGTGTGCGCAATGCTCACTGACGGCCGCCCGGTTTCTGGCGTTGTCAACATATCCCC
M A W A S V C A M L T D G R P V S G V V N I S P
sequenced 701 TGGCGTGGGCTTCCGTGTGCGCAATGCTCACTGACGGCCGCCCGGTTTCTGGCATTGTCAACATATCCCC
M A W A S V C A M L T D G R P V S G I V N I S P
780 790 800 810 820 830 840
TbMSP-C 771 CAAGTACGTTGCCGAAACTGACTTCTTTGTGCGAGTTATCGCACACGAATTAGGACACGCTCTCGGCTTC
K Y V A E T D F F V R V I A H E L G H A L G F
sequenced 771 CAAGTACGTTGCCGAAACTGACTTCTTTGTGCGAGTTATCGCACACGAATTAGGACACGCTCTCGGCTTC
K Y V A E T D F F V R V I A H E L G H A L G F
850 860 870 880 890 900 910
TbMSP-C 841 CAAGCAGATATCCTCATAAGAAGGGGAATTATGAAACAGAAAGGAGGCATTCGCGGACTTAAAACCTCGT
Q A D I L I R R G I M K Q K G G I R G L K T S
sequenced 841 CAAGCAGATATCCTCATAAGAAGGGGAATTATGAAACAGAAAGGAGGCATTCGCGGACTTAAAACCTCGT
Q A D I L I R R G I M K Q K G G I R G L K T S
920 930 940 950 960 970 980
TbMSP-C 911 GGTTGGTCGATTCTGAGGTTGCGAAGCGTGTCGCGCGGAAGCACTTCAATTGCTCGACGGCTCCGGGTAT
W L V D S E V A K R V A R K H F N C S T A P G I
sequenced 911 GGTTGGTCGATTCTGAGGTTGCGAAGCGTGTCGCGCGGAAGCACTTCAATTGCTCGACGGCTCCGGGTAT
W L V D S E V A K R V A R K H F N C S T A P G I
990 1000 1010 1020 1030 1040 1050
TbMSP-C 981 CGAAATGGAGAATGAAGGAGGCCCCGGCGTTTTCGGCAACACACTTGGAGCAACGCAATGCCGTCGAGGAC
E M E N E G G P G V F A T H L E Q R N A V E D
sequenced 981 CGAAATGGAGAATGAAGGAGGTCCCGGCGTTTTCGGCAACACACTTGGAGCAACGCAATGCCGTCGAGGAC
E M E N E G G P G V F A T H L E Q R N A V E D
1060 1070 1080 1090 1100 1110 1120
TbMSP-C 1051 GTGATGGCCCCTTATGGGAACTTGAATTATTTAACTGTGATGTCGTTAGGGGTGTTCGCAAGTATGGGAC
V M A P Y G N L N Y L T V M S L G V F A S M G
sequenced 1051 GTGATGGCCCCTTATGGGAACTTGAATTATTTAACTGTGATGTCATTAGGGGTGTTCGCAAGTATGGGAC
V M A P Y G N L N Y L T V M S L G V F A S M G
1130 1140 1150 1160 1170 1180 1190
TbMSP-C 1121 ACTACCGCGTCAATTTCAGTCGCGCAGAGAAGACGCGTTGGGGCCTAAATCGTGGATGTAGTTTTCTACA
H Y R V N F S R A E K T R W G L N R G C S F L Q
sequenced 1121 ATTACCGCGTCAATTTCAGTCGCGCAGAGAAGACGCGTTGGGGCCTAAATCGTGGATGTAGTTTTCTACA
H Y R V N F S R A E K T R W G L N R G C S F L Q
1200 1210 1220 1230 1240 1250 1260
TbMSP-C 1191 GGAGAAATGCTTGCAGGAAGGAAAGTCAAAGCACCCCGACACGTTCTGCGATCATTTATGGAAGAGCCGT
E K C L Q E G K S K H P D T F C D H L W K S R
sequenced 1191 GGAGAAATGCTTGCAGGAAGGAAAGTCAAAGCATCCCGACACGTTCTGCGATCATTTATGGAAGAGCCGT
E K C L Q E G K S K H P D T F C D H L W K S R
1270 1280 1290 1300 1310 1320 1330
TbMSP-C 1261 CTGTTCACTTGTACCCATGACCGTCTCGGACTCGGCCAGTGTTCCCTAGGTACACATCGAACTGAACTTC
L F T C T H D R L G L G Q C S L G T H R T E L
sequenced 1261 CTGTTCACTTGTACCCATGACCGTCTCAGACTCGGCCAGTGTTCCCTAGGTACACATCGAACTGAACTTC
L F T C T H D R L R L G Q C S L G T H R T E L

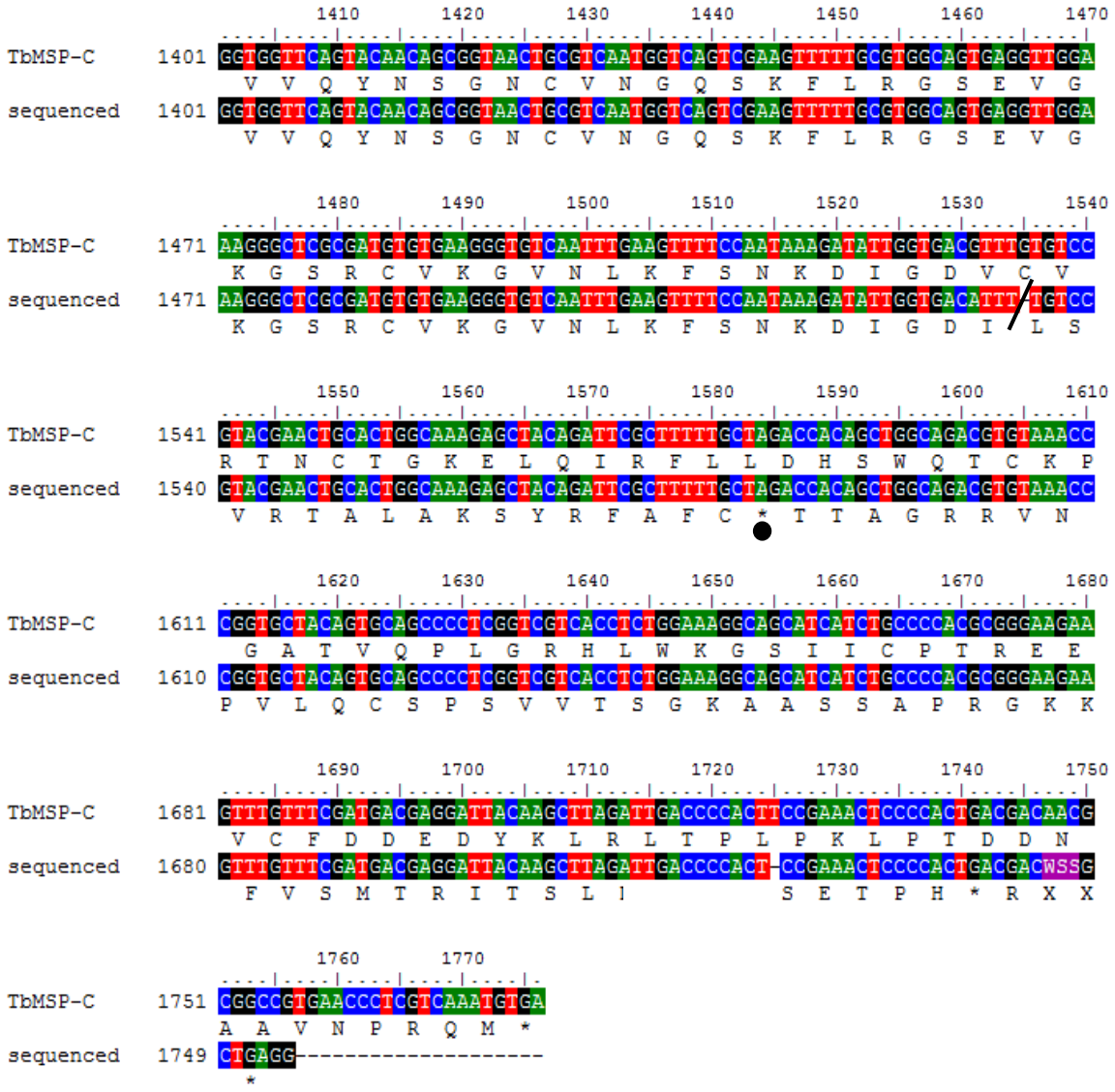


Figure A6: Cloned *TbMSP-C* gene sequencing compared to actual sequence including translated protein sequence. Sequencing done using specific gene primers and edited using the BioEdit. Black arrows indicate mutation in protein sequence and black line indicated deletion and frameshift. Stars indicate stop codons. Black circle indicates where expressed protein would prematurely stop.

Appendix 7
Plasmid Vector Maps

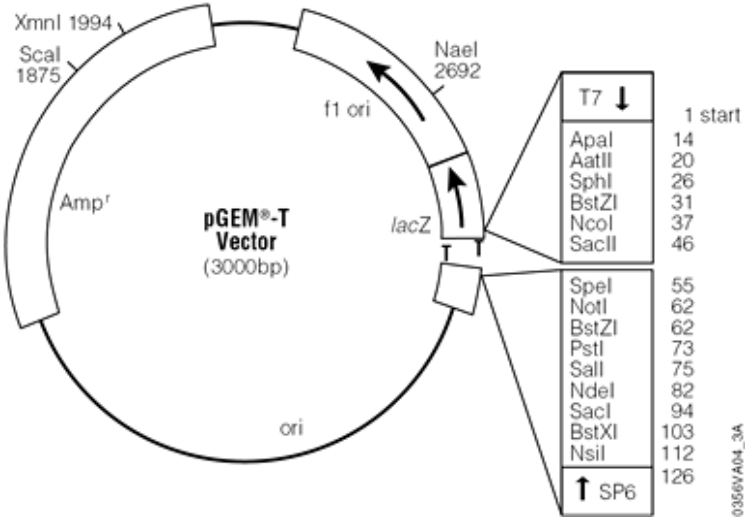


Figure A7.1: pGEM®-T vector map. Going anticlockwise around the plasmid, Amp^r, ampicillin resistance; ori, origin of replication of double stranded DNA; thymidine overhang; multiple cloning site (MCS); lacZ, gene that codes for β-galactosidase which cleaves X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) into blue chromogenic product; f1 ori, origin of replication of single stranded DNA. The vector also contains T7 and SP6 primer binding sites.

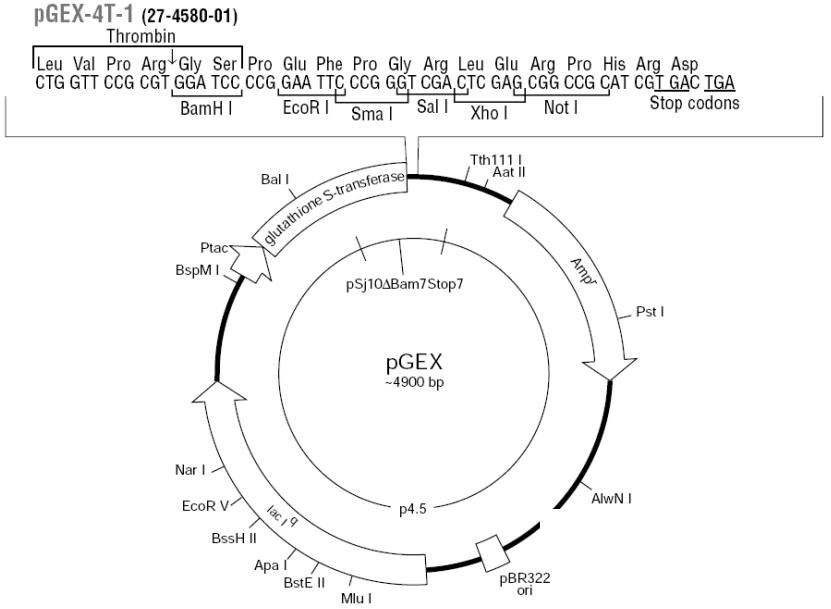


Figure A7.2: pGEX 4T-1 vector map. Going clockwise around the plasmid, Amp^r, ampicillin resistance; ori, origin of replication of double stranded DNA; lac I^q, lac repression; Ptac, Taq promoter which induces expression in the presence of IPTG; glutathione S-transferase (GST) gene, GST and results in the expression of fusion protein; multiple cloning site (MCS) containing the thrombin cleavage site.

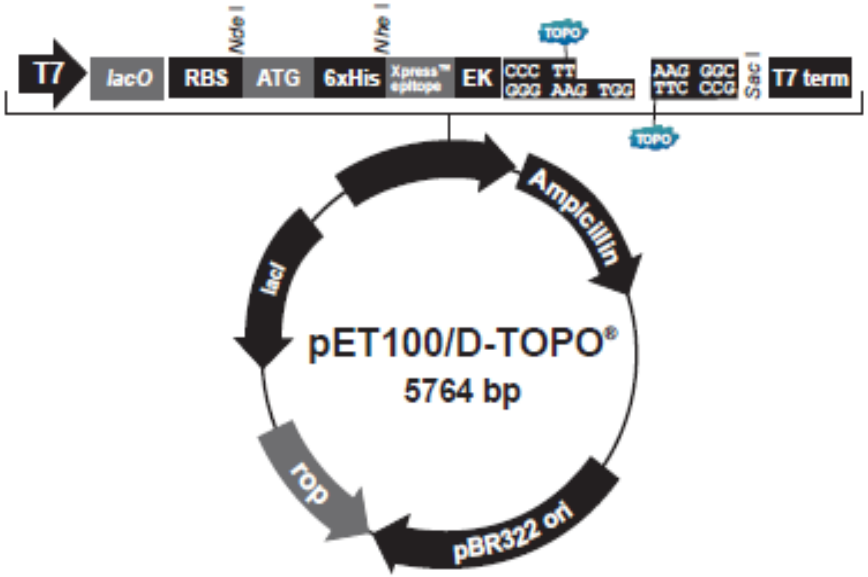


Figure A7.3: pET100/D-TOPO vector map. Going clockwise around the plasmid; lac I^q, lac repression; ampicillin resistance; ori, origin of replication of double stranded DNA; rop, repressor of primer protein.

References

Aalberse, R.; Akkerdaas, J., and Van Ree, R. 2001. Cross-reactivity of IgE antibodies to allergens. *Allergy*, 56 (6): 478-490.

Abdoulmoumini, M.; Jean, E. N.; Suh, P. F., and Youssouf, M. M. 2015. Prevalence and impact of bovine trypanosomiasis in Mayo Rey division, a Soudano-Sahelian zone of Cameroon. *Journal of Parasitology and Vector Biology*, 7 (5): 80-88.

Allsopp, R., and Hursey, B. 2004. Insecticidal control of tsetse, CAB International, England. pp 491-507

Alsford, S.; Kelly, J. M.; Baker, N., and Horn, D. 2013. Genetic dissection of drug resistance in trypanosomes. *Parasitology*, 140 (12): 1478-1491.

Appolaire, A.; Durá, M. A.; Ferruit, M.; Andrieu, J. P.; Godfroy, A.; Gribaldo, S., and Franzetti, B. 2014. The TET2 and TET3 aminopeptidases from Pyrococcus horikoshii form a hetero-subunit peptidasome with enhanced peptide destruction properties. *Molecular Microbiology*, 94 (4): 803-814.

Armah, D. A., and Mensa-Wilmot, K. 2000. Tetramerization of glycosylphosphatidylinositol-specific phospholipase C from *Trypanosoma brucei*. *Journal of Biological Chemistry*, 275 (25): 19334-19342.

Ashford, R. W. 2001. Encyclopedia of arthropod-transmitted infections of man and domesticated animals, CABI, New York, USA. pp 33-46

Azzazy, H. M., and Highsmith, W. E. 2002. Phage display technology: clinical applications and recent innovations. *Clinical Biochemistry*, 35 (6): 425-445.

Bailey, J., and Smith, D. 1994. The quantitative buffy coat for the diagnosis of trypanosomes. *Tropical Doctor*, 24 (2): 54-56.

Baker, C.; Hood, G.; Hunt, E. L.; Lecesne, G. I.; de Boissière, R. F.; Davies, A. H.; da Cunha, E.; Lunn, J.; Belilios, R., and Fletcher, W. 1902. The discovery of the human *Trypanosoma*. *The British Medical Journal*, 2 (2187): 1741.

Bangs, J. D.; Ransom, D. M.; McDowell, M. A., and Brouch, E. M. 1997. Expression of bloodstream variant surface glycoproteins in procyclic stage *Trypanosoma brucei:* role of GPI anchors in secretion. *The EMBO Journal*, 16 (14): 4285-4294.

Bangs, J. D.; Ransom, D. A.; Nimick, M.; Christie, G., and Hooper, N. M. 2001. In vitro cytocidal effects on *Trypanosoma brucei* and inhibition of *Leishmania major* GP63 by peptidomimetic metalloprotease inhibitors. *Molecular and Biochemical Parasitology*, 114 (1): 111-117.

Baral, T. N. 2010. Immunobiology of African trypanosomes: need of alternative interventions. *BioMed Research International*, 2010.

Barrett, A. J.; Woessner, J. F., and Rawlings, N. D. 2004. Handbook of proteolytic enzymes, vol. 1,Elsevier Academic Press, England. pp 231-1047

Barrett, M.; Boykin, D.; Brun, R., and Tidwell, R. 2007. Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. *The British Journal of Pharmacology*, 152 (8): 1155-1171.

Barry, D., and Carrington, M. 2004. Antigenic Variation, CAB International, England. pp

Bastin, P.; Sherwin, T., and Gull, K. 1998. Paraflagellar rod is vital for trypanosome motility. *Nature*, 391 (6667): 548-548.

Bastos, I.; Motta, F.; Grellier, P., and Santana, J. 2013. Parasite prolyl oligopeptidases and the challenge of designing chemotherapeuticals for Chagas disease, leishmaniasis and African trypanosomiasis. *Current Medicinal Chemistry*, 20 (25): 3103.

Bastos, I. M.; Motta, F. N.; Charneau, S.; Santana, J. M.; Dubost, L.; Augustyns, K., and Grellier, P. 2010. Prolyl oligopeptidase of *Trypanosoma brucei* hydrolyzes native collagen, peptide hormones and is active in the plasma of infected mice. *Microbes and Infection*, 12 (6): 457-466.

Benjamin, D. C., and Perdue, S. S. 1996. Site-directed mutagenesis in epitope mapping. *Methods*, 9 (3): 508-515.

Berg, J. M.; Tymoczko, J. L., and Stryer, L. 2002. Biochemistry, WH Freeman, USA. pp 81-95

Berriman, M.; Ghedin, E.; Hertz-Fowler, C.; Blandin, G.; Renauld, H.; Bartholomeu, D. C.; Lennard, N. J.; Caler, E.; Hamlin, N. E., and Haas, B. 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science*, 309 (5733): 416-422.

Bertenshaw, G. P.; Norcum, M. T., and Bond, J. S. 2003. Structure of Homo-and Hetero-oligomeric Meprin Metalloproteases. *Journal of Biological Chemistry*, 278 (4): 2522-2532.

Bett, B.; Randolph, T.; Irungu, P.; Nyamwaro, S.; Kitala, P.; Gathuma, J.; Grace, D.; Vale, G.; Hargrove, J., and McDermott, J. 2010. Field trial of a synthetic tsetse-repellent technology

developed for the control of bovine trypanosomosis in Kenya. *Preventive Veterinary Medicine*, 97 (3): 220-227.

Birhanu, H.; Rogé, S.; Simon, T.; Baelmans, R.; Gebrehiwot, T.; Goddeeris, B. M., and Büscher, P. 2015. Surra Sero K-SeT, a new immunochromatographic test for serodiagnosis of *Trypanosoma evansi* infection in domestic animals. *Veterinary Parasitology*, 211 (3-4): 153-157.

Bjarnason, J. B., and Fox, J. W. 1995. Snake venom metalloendopeptidases: Reprolysins. *Methods in Enzymology*, 248: 345-368.

Black, S. J., and Mansfield, J. M. 2016. Prospects for vaccination against pathogenic African trypanosomes. *Parasite Immunology*.

Blum, H.; Beier, H., and Gross, H. J. 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, 8 (2): 93-99.

Bode, W.; Gomis-Rüth, F.-X., and Stöckler, W. 1993. Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Letters*, 331 (1): 134-140.

Bond, J. S., and Beynon, R. J. 1995. The astacin family of metalloendopeptidases. *Protein Science*, 4 (7): 1247-1261.

Bouvier, J.; Schneider, P., and Malcolm, B. 1993. A fluorescent peptide substrate for the surface metalloprotease of *Leishmania*. *Experimental Parasitology*, 76 (2): 146-155.

Bouvier, J.; Schneider, P.; Etges, R., and Bordier, C. 1990. Peptide substrate specificity of the membrane-bound metalloprotease of *Leishmania*. *Biochemistry*, 29 (43): 10113-10119.

Bouyer, J.; Stachurski, F.; Gouro, A., and Lancelot, R. 2009. Control of bovine trypanosomosis by restricted application of insecticides to cattle using footbaths. *Veterinary Parasitology*, 161 (3): 187-193.

Bowman, D. D. 2006. Successful and currently ongoing parasite eradication programs. *Veterinary Parasitology*, 139 (4): 293-307.

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 (1): 248-254.

Brittingham, A.; Morrison, C. J.; McMaster, W. R.; McGwire, B. S.; Chang, K.-P., and Mosser, D. M. 1995. Role of the *Leishmania* surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *The Journal of Immunology*, 155 (6): 3102-3111.

Brockmann, E.-C.; Cooper, M.; Strömsten, N.; Vehniäinen, M., and Saviranta, P. 2005. Selecting for antibody scFv fragments with improved stability using phage display with denaturation under reducing conditions. *Journal of Immunological Methods*, 296 (1): 159-170.

Brockmann, R.; Beyer, A.; Heinisch, J. J., and Wilhelm, T. 2007. Posttranscriptional expression regulation: what determines translation rates. *PLoS Computational Biology*, 3 (3): e57.

Brown, D., and Waneck, G. L. 1992. Glycosyl-phosphatidylinositol-anchored membrane proteins. *Journal of the American Society of Nephrology*, 3 (4): 895-906.

Bruce, D. 1895. Preliminary Report on the Tsetse Fly Disease or Nagana, in Zululand, Bennett & Davis, South Africa. pp 4-25

Brun, R.; Blum, J.; Chappuis, F., and Burri, C. 2010. Human african trypanosomiasis. *The Lancet*, 375 (9709): 148-159.

Buchan, J. R., and Stansfield, I. 2007. Halting a cellular production line: responses to ribosomal pausing during translation. *Biology of the Cell*, 99 (9): 475-487.

Bülow, R.; Nonnengässer, C., and Overath, P. 1989. Release of the variant surface glycoprotein during differentiation of bloodstream to procyclic forms of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 32 (1): 85-92.

Büscher, P., and Lejon, V. 2004. Diagnosis of human African trypanosomiasis, CAB International, England. pp 203-218

Büscher, P.; Gilleman, Q., and Lejon, V. 2013. Rapid diagnostic test for sleeping sickness. *New England Journal of Medicine*, 368 (11): 1069-1070.

Büscher, P.; Mertens, P.; Leclipteux, T.; Gilleman, Q.; Jacquet, D.; Mumba-Ngoyi, D.; Pyana, P. P.; Boelaert, M., and Lejon, V. 2014. Sensitivity and specificity of HAT Sero-K-SeT, a rapid diagnostic test for serodiagnosis of sleeping sickness caused by *Trypanosoma brucei gambiense*: a case-control study. *The Lancet* 2(6): e359-e363.

Button, L. L., and McMaster, W. 1988. Molecular cloning of the major surface antigen of leishmania. *The Journal of Experimental Medicine*, 167 (2): 724-729.

Cármenes, R.; Freije, J.; Molina, M., and Martin, J. 1989. Predict7, a program for protein structure prediction. *Biochemical and Biophysical Research Communications*, 159 (2): 687-693.

Casgrain, P.-A.; Martel, C.; McMaster, W. R.; Mottram, J. C.; Olivier, M., and Descoteaux, A. 2016. Cysteine Peptidase B Regulates *Leishmania mexicana* Virulence through the Modulation of GP63 Expression. *PLoS Pathogens*, 12 (5): e1005658.

Chappuis, F.; Loutan, L.; Simarro, P.; Lejon, V., and Büscher, P. 2005. Options for field diagnosis of human African trypanosomiasis. *Clinical Microbiology Reviews*, 18 (1): 133-146.

Chaudhuri, G., and Chang, K.-P. 1988. Acid protease activity of a major surface membrane glycoprotein (gp63) from *Leishmania mexicana* promastigotes. *Molecular and Biochemical Parasitology*, 27 (1): 43-52.

Chaudhuri, G.; Chaudhuri, M.; Pan, A., and Chang, K. 1989. Surface acid proteinase (gp63) of *Leishmania mexicana*. A metalloenzyme capable of protecting liposome-encapsulated proteins from phagolysosomal degradation by macrophages. *Journal of Biological Chemistry*, 264 (13): 7483-7489.

Chothia, C., and Janin, J. 1975. Principles of protein-protein recognition. *Nature*, 256: 705-708.

Cirilli, M.; Gallina, C.; Gavuzzo, E.; Giordano, C.; Gomis-Rüth, F.; Gorini, B.; Kress, L.; Mazza, F.; Paglialunga Paradisi, M., and Pochetti, G. 1997. 2 Å X-ray structure of adamalysin II complexed with a peptide phosphonate inhibitor adopting a retro-binding mode. *FEBS Letters*, 418 (3): 319-322.

Cserzo, M.; Eisenhaber, F.; Eisenhaber, B., and Simon, I. 2002. On filtering false positive transmembrane protein predictions. *Protein Engineering*, 15 (9): 745-752.

Cserzo, M.; Eisenhaber, F.; Eisenhaber, B., and Simon, I. 2004. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. *Bioinformatics*, 20 (1): 136-137.

Cuervo, P.; Santos, A. L.; Alves, C. R.; Menezes, G. C.; Silva, B. A.; Britto, C.; Fernandes, O.; Cupolillo, E., and De Jesus, J. B. 2008. Cellular localization and expression of gp63 homologous metalloproteases in *Leishmania (Viannia) braziliensis* strains. *Acta Tropica*, 106 (3): 143-148.

Cuevas, I. C.; Cazzulo, J. J., and Sánchez, D. O. 2003. gp63 homologues in *Trypanosoma cruzi:* surface antigens with metalloprotease activity and a possible role in host cell infection. *Infection and Immunity*, 71 (10): 5739-5749.

d'Ieteren, G.; Authie, E.; Wissocq, N., and Murray, M. 1998. Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis. *International Office of Epizootics*, 17 (1): 154-175.

Dahler, G.; Barras, F., and Keen, N. 1990. Cloning of genes encoding extracellular metalloproteases from Erwinia chrysanthemi EC16. *Journal of Bacteriology*, 172 (10): 5803-5815.

de Haard, H. J.; Kazemier, B.; Koolen, M. J.; Nijholt, L. J.; Meloen, R. H.; van Gemen, B.; Hoogenboom, H. R., and Arends, J.-W. 1998. Selection of recombinant, library-derived antibody fragments against p24 for human immunodeficiency virus type 1 diagnostics. *Clinical and Diagnostic Laboratory Immunology*, 5 (5): 636-644.

De Marco, A.; Vigh, L.; Diamant, S., and Goloubinoff, P. 2005. Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid-or benzyl alcohol–overexpressed molecular chaperones. *Cell Stress and Chaperones*, 10 (4): 329-339.

de Sousa, K. P.; Atouguia, J., and Silva, M. S. 2010. Partial biochemical characterization of a metalloproteinase from the bloodstream forms of *Trypanosoma brucei brucei* parasites. *The Protein Journal*, 29 (4): 283-289.

Deborggraeve, S.; Claes, F.; Laurent, T.; Mertens, P.; Leclipteux, T.; Dujardin, J.; Herdewijn, P., and Büscher, P. 2006. Molecular dipstick test for diagnosis of sleeping sickness. *Journal of Clinical Microbiology*, 44 (8): 2884-2889.

Deborggraeve, S.; Lejon, V.; Ekangu, R. A.; Ngoyi, D. M.; Pyana, P. P.; Ilunga, M.; Mulunda, J. P., and Büscher, P. 2011. Diagnostic accuracy of PCR in gambiense sleeping sickness diagnosis, staging and post-treatment follow-up: a 2-year longitudinal study. *PLoS Neglected Tropical Diseases*, 5 (2): e972.

Dedet, J.-P. 2002. Current status of epidemiology of leishmaniases, Kluwer Academic Press, Netherlands pp 1-10

DeMartino, G. N., and Slaughter, C. A. 1999. The proteasome, a novel protease regulated by multiple mechanisms. *Journal of Biological Chemistry*, 274 (32): 22123-22126.

Desquesnes, M.; McLaughlin, G.; Zoungrana, A., and Dávila, A. M. 2001. Detection and identification of *Trypanosoma* of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA. *International Journal for Parasitology*, 31 (5): 610-614.

Desquesnes, M.; Holzmuller, P.; Lai, D.-H.; Dargantes, A.; Lun, Z.-R., and 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.

Dougan, D. A.; Malby, R. L.; Gruen, L. C.; Kortt, A. A., and Hudson, P. J. 1998. Effects of substitutions in the binding surface of an antibody on antigen affinity. *Protein Engineering*, 11 (1): 65-74.

Duellman, T.; Burnett, J., and Yang, J. 2015. Quantitation of secreted proteins using mCherry fusion constructs and a fluorescent microplate reader. *Analytical Biochemistry*, 473: 34-40.

Duong, F.; Lazdunski, A.; Carni, B., and Murgier, M. 1992. Sequence of a cluster of genes controlling synthesis and secretion of alkaline protease in *Pseudomonas aeruginosa*: relationships to other secretory pathways. *Gene*, 121 (1): 47-54.

Dutton, J. E. 1902. Note on a *Trypanosoma* occurring in the blood of man. *The British Medical Journal*: 881-884.

Eisler, M.; Dwinger, R.; Majiwa, P., and Picozzi, K. 2004. Diagnosis and epidemiology of African animal trypanosomiasis, CAB International, Wallingford. pp 253-267

El-Sayed, N. M., and Donelson, J. E. 1997a. African trypanosomes have differentially expressed genes encoding homologues of the *Leishmania* GP63 surface protease. *Journal of Biological Chemistry*, 272 (42): 26742-26748.

El-Sayed, N. M., and Donelson, J. E. 1997b. A survey of the *Trypanosoma brucei rhodesiense* genome using shotgun sequencing. *Molecular and Biochemical Parasitology*, 84 (2): 167-178.

El-Sayed, N. M.; Hegde, P.; Quackenbush, J.; Melville, S. E., and Donelson, J. E. 2000. The African trypanosome genome. *International Journal for Parasitology*, 30 (4): 329-345.

El-Sayed, N. M.; Myler, P. J.; Blandin, G.; Berriman, M.; Crabtree, J.; Aggarwal, G.; Caler, E.; Renauld, H.; Worthey, E. A., and Hertz-Fowler, C. 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science*, 309 (5733): 404-409.

Endemann, H., and Model, P. 1995. Location of filamentous phage minor coat proteins in phage and in infected cells. *Journal of Molecular Biology*, 250 (4): 496-506.

Enyaru, J. C.; Ouma, J. O.; Malele, I. I.; Matovu, E., and Masiga, D. K. 2010. Landmarks in the evolution of technologies for identifying trypanosomes in tsetse flies. *Trends in Parasitology*, 26 (8): 388-394.

Erez, E.; Fass, D., and Bibi, E. 2009. How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature*, 459 (7245): 371-378.

Ersfeld, K., and Gull, K. 1997. Partitioning of large and minichromosomes in *Trypanosoma brucei*. *Science*, 276 (5312): 611-614.

Etges, R. 1992. Identification of a surface metalloproteinase on 13 species of *Leishmania* isolated from humans, *Crithidia fasciculata*, and *Herpetomonas samuelpessoai*. *Acta Tropica*, 50 (3): 205-217.

Etges, R.; Bouvier, J., and Bordier, C. 1986. The major surface protein of *Leishmania* promastigotes is a protease. *Journal of Biological Chemistry*, 261 (20): 9098-9101.

Fairlamb, A. H.; Henderson, G. B., and Cerami, A. 1989. Trypanothione is the primary target for arsenical drugs against African trypanosomes. *Proceedings of the National Academy of Sciences of the USA*, 86 (8): 2607-2611.

Farrell, J. 2002. *Leishmania*, vol. 4,Springer Science, USA. pp 58-61

Fenn, K., and Matthews, K. R. 2007. The cell biology of *Trypanosoma brucei* differentiation. *Current Opinion in Microbiology*, 10 (6): 539-546.

Ferguson, M. A., and Hart, G. W. 2009. Glycosylphosphatidylinositol Anchors, 2 ed.Cold Spring Harbor Laboratory Press, USA. pp 143-162

Fleming, J. R.; Sastry, L.; Wall, S. J.; Sullivan, L., and Ferguson, M. A. 2016. Proteomic Identification of Immunodiagnostic Antigens for *Trypanosoma vivax* Infections in Cattle and Generation of a Proof-of-Concept Lateral Flow Test Diagnostic Device. *PLOS Neglected Tropical Diseases*, 10 (9): e0004977.

Foil, L. D. 1999. Mechanical transmission of agents of livestock diseases by tabanids. , Inter-American Institute for Cooperation on Agriculture (IIAC), Georgetown, Guyana. pp 61-65

Frommel, T. O.; Button, L. L.; Fujikura, Y., and McMaster, W. R. 1990. The major surface glycoprotein (GP63) is present in both life stages of *Leishmania*. *Molecular and Biochemical Parasitology*, 38 (1): 25-32.

Galazka, G.; Windsor, L. J.; Birkedal-Hansen, H., and Engler, J. A. 1996. APMA (4-aminophenylmercuric acetate) activation of stromelysin-1 involves protein interactions in addition to those with cysteine-75 in the propeptide. *Biochemistry*, 35 (34): 11221-11227.

Garcia-Carreon, F. L., and Del Toro, M. 1997. Classification of proteases without tears. *Biochemical Education*, 25 (3): 161-167.

Garcia, A.; Jamonneau, V.; Magnus, E.; Laveissière, C.; Lejon, V.; N'guessan, P.; N'dri, L.; Meirvenne, N., and Büscher, P. 2000. Follow up of Card Agglutination Trypanosomiasis Test (CATT) positive but apparently aparasitaemic individuals in Côte d'Ivoire: evidence for a complex and heterogeneous population. *Tropical Medicine & International Health*, 5 (11): 786-793.

Gikonyo, N. K.; Hassanali, A.; Njagi, P. G., and Saini, R. K. 2003. Responses of *Glossina morsitans* morsitans to blends of electroantennographically active compounds in the odors of its preferred (buffalo and ox) and nonpreferred (waterbuck) hosts. *Journal of Chemical Ecology*, 29 (10): 2331-2345.

Goldring, J., and Coetzer, T. H. 2003. Isolation of chicken immunoglobulins (IgY) from egg yolk. *Biochemistry and Molecular Biology Education*, 31 (3): 185-187.

Gómara, M. J., and Haro, I. 2007. Synthetic peptides for the immunodiagnosis of human diseases. *Current Medicinal Chemistry*, 14 (5): 531-546.

Gomis-Ruth, F.-X. 1997. Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature*, 389 (6646): 77-81.

Gomis-Rüth, F.; Stöcker, W.; Huber, R.; Zwilling, R., and Bode, W. 1993. Refined 1· 8 Å X-ray Crystal Structure of Astacin, a Zinc-endopeptidase from the Crayfish, *Astacus astacus*: Structure Determination, Refinement, Molecular Structure and Comparison with Thermolysin. *Journal of Molecular Biology*, 229 (4): 945-968.

Gomis-Rüth, F.; Kress, L.; Kellermann, J.; Mayr, I.; Lee, X.; Huber, R., and Bode, W. 1994. Refined 2· 0 Å X-ray Crystal Structure of the Snake Venom Zinc-endopeptidase Adamalysin II: Primary and Tertiary Structure Determination, Refinement, Molecular Structure and Comparison with Astacin, Collagenase and Thermolysin. *Journal of Molecular Biology*, 239 (4): 513-544.

Gomis-Rüth, F. X. 2003. Structural aspects of the metzincin clan of metalloendopeptidases. *Molecular Biotechnology*, 24 (2): 157-202.

Gomis-Rüth, F. X. 2009. Catalytic domain architecture of metzincin metalloproteases. *Journal of Biological Chemistry*, 284 (23): 15353-15357.

Grandgenett, P.; Coughlin, B.; Kirchhoff, L., and Donelson, J. 2000. Differential expression of GP63 genes in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 110 (2): 409-415.

Grandgenett, P. M.; Otsu, K.; Wilson, H. R.; Wilson, M. E., and Donelson, J. E. 2007. A function for a specific zinc metalloprotease of African trypanosomes. *PLoS Pathogens*, 3 (10): e150.

Grellier, P.; Vendeville, S.; Joyeau, R.; Bastos, I. M.; Drobecq, H.; Frappier, F.; Teixeira, A. R.; Schrével, J.; Davioud-Charvet, E., and Sergheraert, C. 2001. *Trypanosoma cruzi* prolyl oligopeptidase Tc80 is involved in nonphagocytic mammalian cell invasion by trypomastigotes. *Journal of Biological Chemistry*, 276 (50): 47078-47086.

Griep, R. A.; Prins, M.; van Twisk, C.; Keller, H. J.; Kerschbaumer, R. J.; Kormelink, R.; Goldbach, R. W., and Schots, A. 2000. Application of phage display in selecting Tomato spotted wilt virus-specific single-chain antibodies (scFvs) for sensitive diagnosis in ELISA. *Phytopathology*, 90 (2): 183-190.

Gruby, D. 1843. Recherches et observations sur une nouvelle espèce d'hématozoaire, *Trypanosoma sanguinis*. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 7 1134-1136.

Gruszynski, A. E.; van Deursen, F. J.; Albareda, M. C.; Best, A.; Chaudhary, K.; Cliffe, L. J.; del Rio, L.; Dunn, J. D.; Ellis, L., and Evans, K. J. 2006. Regulation of surface coat exchange by differentiating African trypanosomes. *Molecular and Biochemical Parasitology*, 147 (2): 211-223.

Guhl, F., and Vallejo, G. A. 2003. *Trypanosoma (Herpetosoma) rangeli tejera*, 1920: an updated review. *Memórias do Instituto Oswaldo Cruz*, 98 (4): 435-442.

Hanau, S.; Rippa, M.; Bertelli, M.; Dallocchio, F., and Barrett, M. P. 1996. 6-Phosphogluconate Dehydrogenase from *Trypanosoma brucei*. *European Journal of Biochemistry*, 240 (3): 592-599.

Hassani, K.; Shio, M. T.; Martel, C.; Faubert, D., and Olivier, M. 2014. Absence of metalloprotease GP63 alters the protein content of Leishmania exosomes. *PLoS One*, 9 (4): e95007.

Helm, J. R.; Wilson, M. E., and Donelson, J. E. 2009. Differential expression of a protease gene family in African trypanosomes. *Molecular and Biochemical Parasitology*, 163 (1): 8-18.

Herrero, M.; Grace, D.; Njuki, J.; Johnson, N.; Enahoro, D.; Silvestri, S., and Rufino, M. 2013. The roles of livestock in developing countries. *Animal*, 7 (s1): 3-18.

Heussen, C., and Dowdle, E. B. 1980. Electrophoretic analysis of plasminogen activators in polyacrylamide gels containing sodium dodecyl sulfate and copolymerized substrates. *Analytical Biochemistry*, 102 (1): 196-202.

Hill, E. W.; O'Gorman, G. M.; Agaba, M.; Gibson, J. P.; Hanotte, O.; Kemp, S. J.; Naessens, J.; Coussens, P. M., and MacHugh, D. E. 2005. Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics. *Veterinary Immunology and Immunopathology*, 105 (3): 247-258.

Hoare, C. A. 1964. Morphological and taxonomic studies on mammalian trypanosomes. X. Revision of the systematics. *The Journal of Protozoology*, 11 (2): 200-207.

Holliger, P.; Prospero, T., and Winter, G. 1993. " Diabodies": small bivalent and bispecific antibody fragments. *Proceedings of the National Academy of Sciences*, 90 (14): 6444-6448.

Holmes, P.; Eisler, M., and Geerts, S. 2004. Current chemotherapy of animal trypanosomiasis, CAB International, England. pp 431-444

Holmquist, B., and Vallee, B. L. 1974. Metal substitutions and inhibition of thermolysin: spectra of the cobalt enzyme. *Journal of Biological Chemistry*, 249 (14): 4601-4607.

Holt, H.; Selby, R.; Mumba, C.; Napier, G., and Guitian, J. 2016. Assessment of animal African trypanosomiasis (AAT) vulnerability in cattle-owning communities of sub-Saharan Africa. *Parasites & Vectors*, 9 (1): 1.

Hoogenboom, H. R.; Griffiths, A. D.; Johnson, K. S.; Chiswell, D. J.; Hudson, P., and Winter, G. 1991. Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Research*, 19 (15): 4133-4137.

Hooper, N. M. 1994. Families of zinc metalloproteases. *FEBS Letters*, 354 (1): 1-6.

Hopsu-Havu, V. K.; Järvinen, M., and Kirschke, H. 1997. Proteolysis in cell functions, vol. 13,IOS Press, Netherlands. pp 3-21

Hsiao, C.-H. C.; Yao, C.; Storlie, P.; Donelson, J. E., and Wilson, M. E. 2008. The major surface protease (MSP or GP63) in the intracellular amastigote stage of *Leishmania chagasi*. *Molecular and Biochemical Parasitology*, 157 (2): 148-159.

Ilboudo, H.; Bras-Gonçalves, R.; Camara, M.; Flori, L.; Camara, O.; Sakande, H.; Leno, M.; Petitdidier, E.; Jamonneau, V., and Bucheton, B. 2014. Unravelling Human Trypanotolerance: IL8 is Associated with Infection Control whereas IL10 and TNFα Are Associated with Subsequent Disease Development. *PLoS Pathogens*, 10 (11): e1004469.

Ilg, T.; Harbecke, D., and Overath, P. 1993. The lysosomal gp63 related protein in *Leishmania mexicana amastigotes* is a soluble metalloproteinase with an acidic pH optimum. *FEBS Letters*, 327 (1): 103-107.

Ip, H. S.; Orn, A.; Russell, D. G., and Cross, G. A. 1990. *Leishmania mexicana mexicana* gp63 is a site-specific neutral endopeptidase. *Molecular and Biochemical Parasitology*, 40 (2): 163-172.

Ivens, A. C.; Peacock, C. S.; Worthey, E. A.; Murphy, L.; Aggarwal, G.; Berriman, M.; Sisk, E.; Rajandream, M.-A.; Adlem, E., and Aert, R. 2005. The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, 309 (5733): 436-442.

Jackson, A. P.; Berry, A.; Aslett, M.; Allison, H. C.; Burton, P.; Vavrova-Anderson, J.; Brown, R.; Browne, H.; Corton, N., and Hauser, H. 2012. Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proceedings of the National Academy of Sciences*, 109 (9): 3416-3421.

Jaffe, C. L., and Dwyer, D. M. 2003. Extracellular release of the surface metalloprotease, gp63, from *Leishmania* and insect trypanosomatids. *Parasitology Research*, 91 (3): 229-237.

Jamonneau, V.; Ilboudo, H.; Kaboré, J.; Kaba, D.; Koffi, M.; Solano, P.; Garcia, A.; Courtin, D.; Laveissière, C., and Lingue, K. 2012. Untreated human infections by *Trypanosoma brucei gambiense* are not 100% fatal. *PLoS Neglected Tropical Diseases*, 6 (6): e1691.

Jamonneau, V.; Camara, O.; Ilboudo, H.; Peylhard, M.; Koffi, M.; Sakande, H.; N'Dri, L.; Sanou, D.; Dama, E., and Camara, M. 2015. Accuracy of Individual Rapid Tests for Serodiagnosis of Gambiense Sleeping Sickness in West Africa. *PLoS Neglected Tropical Diseases*, 9 (2): e0003480.

Jensen, K.; Østergaard, P. R.; Wilting, R., and Lassen, S. F. 2010. Identification and characterization of a bacterial glutamic peptidase. *BMC Biochemistry*, 11 (1): 47.

Jesudhasan, P. R.; Tan, C.-W., and Woo, P. T. 2007. A metalloproteinase gene from the pathogenic piscine hemoflagellate, *Cryptobia salmositica*. *Parasitology Research*, 100 (4): 899-904.

Kane, J. F. 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology*, 6 (5): 494-500.
Kapp, K.; Schrempf, S.; Lemberg, M. K., and Dobberstein, B. 2009. Post-targeting functions of signal peptides. pp 1-16
Karplus, P. A., and Schulz, G. E. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften*, 72: 212-213.
Katiyar, S.; Kufareva, I.; Behera, R.; Thomas, S., and Ogata, Y. 2013. Lapatinib-Binding Protein Kinases in the African Trypanosome: Identification of Cellular Targets for Kinase-Directed Chemical Scaffolds. *PloS one*, 8 (2): e56150.
Kennedy, P. G. 2004. Human African trypanosomiasis of the CNS: current issues and challenges. *Journal of Clinical Investigation*, 113 (4): 496.
Klemba, M., and Goldberg, D. E. 2002. Biological roles of proteases in parasitic protozoa. *Annual Review of Biochemistry*, 71 (1): 275-305.
Klimpel, S., and Mehlhorn, H. 2014. Bats (Chiroptera) as Vectors of Diseases and Parasites: Facts and Myths, Springer-Verlag, Germany. pp 63-83
Kudo, R. 1922. On the protozoa parasitic in frogs. *Transactions of the American Microscopical Society*: 59-76.
Kulkarni, M. M.; Olson, C. L.; Engman, D. M., and McGwire, B. S. 2009. *Trypanosoma cruzi* GP63 proteins undergo stage-specific differential posttranslational modification and are important for host cell infection. *Infection and Immunity*, 77 (5): 2193-2200.
Kulkarni, M. M.; Karafova, A.; Kamysz, W., and McGwire, B. S. 2014. Design of protease-resistant pexiganan enhances antileishmanial activity. *Parasitology Research*, 113 (5): 1971-1976.
Kulkarni, M. M.; McMaster, W. R.; Kamysz, E.; Kamysz, W.; Engman, D. M., and McGwire, B. S. 2006. The major surface-metalloprotease of the parasitic protozoan, *Leishmania*, protects against antimicrobial peptide-induced apoptotic killing. *Molecular Microbiology*, 62 (5): 1484-1497.
Kyte, J., and Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157: 105-132.
LaCount, D. J.; Gruszynski, A. E.; Grandgenett, P. M.; Bangs, J. D., and Donelson, J. E. 2003. Expression and function of the *Trypanosoma brucei* major surface protease (GP63) genes. *Journal of Biological Chemistry*, 278 (27): 24658-24664.
Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227 (5259): 680-685.
Lalmanach, G.; Boulangé, A.; Serveau, C.; Lecaille, F.; Scharfstein, J.; Gauthier, F., and Authié, E. 2002. Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. *Biological Chemistry*, 383 (5): 739-749.
Lanham, S. M., and Godfrey, D. 1970. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. *Experimental Parasitology*, 28 (3): 521-534.
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., and Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23 (21): 2947-2948.
Lee, B.-S.; Huang, J.-S.; Jayathilaka, G. L. P.; Lateef, S. S., and Gupta, S. 2010. Production of Antipeptide Antibodies, Humana Press, New Jersey, USA. pp 93-108
Lewis, M.; Tippett, H.; Sinanan, A.; Morgan, M., and Hunt, N. 2000. Gelatinase-B (matrix metalloproteinase-9; MMP-9) secretion is involved in the migratory phase of human and murine muscle cell cultures. *Journal of Muscle Research and Cell Motility*, 21 (3): 223-233.
Lindquist, D.; Abusowa, M., and Hall, M. 1992. The New World screwworm fly in Libya: a review of its introduction and eradication. *Medical and Veterinary Entomology*, 6 (1): 2-8.
Liu, B.; Liu, Y.; Motyka, S. A.; Agbo, E. E., and Englund, P. T. 2005. Fellowship of the rings: the replication of kinetoplast DNA. *Trends in Parasitology*, 21 (8): 363-369.
Lloyd, L. L., and Johnson, W. B. 1924. The trypanosome infections of tsetse-flies in Northern Nigeria and a new method of estimation. *Bulletin of Entomological Research*, 14 (03): 265-288.
Lodge, J.; Lund, P., and Minchin, S. 2007. Gene cloning, Garland Science, England. pp 60-65
Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D., and Darnell, J. 2000. Molecular Cell Biology: Hierarchical structure of proteins, Fourth ed.W. H. Freeman and Company, USA. pp 64-72
Løset, G.; Roos, N.; Bogen, B., and Sandlie, I. 2011. Expanding the versatility of phage display II: improved affinity selection of folded domains on protein VII and IX of the filamentous phage. *PLoS One*, 6 (2): e17433.

Macdonald, M. H.; Morrison, C. J., and McMaster, W. R. 1995. Analysis of the active site and activation mechanism of the Leishmania surface metalloproteinase GP63. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1253 (2): 199-207.

Madison-Antenucci, S.; Grams, J., and Hajduk, S. L. 2002. Editing machines: the complexities of trypanosome RNA editing. *Cell*, 108 (4): 435-438.

Magnus, E.; Vervoort, T., and Van Meirvenne, N. 1978. A card-agglutination test with stained trypanosomes (C.A.T.T.) for the serological diagnosis of *T. b. gambiense* trypanosomiasis. *Annales de la Societe Belge de Medecine Tropicale*, 58 (3): 169-179.

Maitima, J. M. 2007. Guidelines for assessing environmental and socio-economic impacts of tsetse and trypanosomiasis interventions, ILRI (aka ILCA and ILRAD). pp 11

Majekodunmi, A. O.; Fajinmi, A.; Dongkum, C.; Picozzi, K.; Thrusfield, M. V., and Welburn, S. C. 2013. A longitudinal survey of African animal trypanosomiasis in domestic cattle on the Jos Plateau, Nigeria: prevalence, distribution and risk factors. *Parasites & Vectors*, 6: 239.

Manson, P.; Christy, C.; Dutton, J. E.; Todd, J.; Castellani, A.; Sambon, L. W.; Low, G., and Rogers, L. 1903. Discussion on trypanosomiasis. *The British Medical Journal*: 645-654.

Marcoux, V.; Wei, G.; Tabel, H., and Bull, H. J. 2010. Characterization of major surface protease homologues of *Trypanosoma congolense*. *BioMed Research International*, 2010.

Matte, C.; Casgrain, P.-A.; Séguin, O.; Moradin, N.; Hong, W. J., and Descoteaux, A. 2016. *Leishmania major* Promastigotes Evade LC3-Associated Phagocytosis through the Action of GP63. *PLoS Pathogens*, 12 (6): e1005690.

Matthews, B.; Colman, P.; Jansonius, J.; Titani, K.; Walsh, K., and Neurath, H. 1972. Structure of thermolysin. *Nature*, 238 (80): 41-43.

Matthews, K. 1999. Developments in the Differentiation of *Trypanosoma brucei*. *Parasitology Today*, 15 (2): 76-80.

Matthews, K. R. 2005. The developmental cell biology of *Trypanosoma brucei*. *Journal of Cell Science*, 118 (2): 283-290.

Matthews, K. R.; Ellis, J. R., and Paterou, A. 2004. Molecular regulation of the life cycle of African trypanosomes. *Trends in Parasitology*, 20 (1): 40-47.

McGwire, B. S., and Chang, K.-P. 1996. Posttranslational Regulation of a *Leishmania* HEXXH Metalloprotease (gp63) The Effects Of Site-Specific Mutagenesis Of Catalytic, Zinc Binding, N-Glycosylation, And Glycosyl Phosphatidylinositol Addition Sites On N-Terminal End Cleavage, Intracellular Stability, And Extracellular Exit. *Journal of Biological Chemistry*, 271 (14): 7903-7909.

McGwire, B. S.; Chang, K.-P., and Engman, D. M. 2003a. Migration through the extracellular matrix by the parasitic protozoan *Leishmania* is enhanced by surface metalloprotease gp63. *Infection and Immunity*, 71 (2): 1008-1010.

McGwire, B. S.; Olson, C. L.; Tack, B. F., and Engman, D. M. 2003b. Killing of African trypanosomes by antimicrobial peptides. *Journal of Infectious Diseases*, 188 (1): 146-152.

McMaster, W.; Morrison, C.; MacDonald, M., and Joshi, P. 1994. Mutational and functional analysis of the *Leishmania* surface metalloproteinase GP63: similarities to matrix metalloproteinases. *Parasitology*, 108 (S1): S29-S36.

Mendoza-Palomares, C.; Biteau, N.; Giroud, C.; Coustou, V.; Coetzer, T.; Authié, E.; Boulangé, A., and Baltz, T. 2008. Molecular and biochemical characterization of a cathepsin B-like protease family unique to Trypanosoma congolense. *Eukaryotic cell*, 7 (4): 684-697.

Miller, B. R.; Demarest, S. J.; Lugovskoy, A.; Huang, F.; Wu, X.; Snyder, W. B.; Croner, L. J.; Wang, N.; Amatucci, A., and Michaelson, J. S. 2010. Stability engineering of scFvs for the development of bispecific and multivalent antibodies. *Protein Engineering Design and Selection*, 23 (7): 549-557.

Mo, Z.; Suganuma, K.; Ruttayaporn, N.; Nguyen, T. T.; Yamasaki, S.; Igarashi, I.; Kawazu, S.; Suzuki, Y., and Inoue, N. 2014. Identification and Characterization of a *Trypanosoma congolense* 46 kDa Protein as a Candidate Serodiagnostic Antigen. *The Journal of Veterinary Medical Science*, 76 (6): 799.

Mochabo, K. M.; Zhou, M.; Suganuma, K.; Kawazu, S.-i.; Suzuki, Y., and Inoue, N. 2013. Expression, immunolocalization and serodiagnostic value of Tc38630 protein from *Trypanosoma congolense*. *Parasitology Research*, 112 (9): 3357-3363.

Mock, W. L., and Yao, J. 1997. Kinetic characterization of the serralysins: a divergent catalytic mechanism pertaining to astacin-type metalloproteases. *Biochemistry*, 36 (16): 4949-4958.

Mohler, J. R., and Schoening, H. W. 1920. Dourine of horses, US Department of Agriculture, USA. pp

Morty, R. E.; Troeberg, L.; Pike, R. N.; Jones, R.; Nickel, P.; Lonsdale-Eccles, J. D., and Coetzer, T. H. 1998. A trypanosome oligopeptidase as a target for the trypanocidal agents pentamidine, diminazene and suramin. *FEBS Letters*, 433 (3): 251-256.

Morty, R. E.; Lonsdale-Eccles, J. D.; Morehead, J.; Caler, E. V.; Mentele, R.; Auerswald, E. A.; Coetzer, T. H.; Andrews, N. W., and Burleigh, B. A. 1999. Oligopeptidase B from *Trypanosoma brucei,* a new member of an emerging subgroup of serine oligopeptidases. *Journal of Biological Chemistry*, 274 (37): 26149-26156.

Mott, F. W. 1908. Histological observations on sleeping sickness and other trypanosome infections. *The Journal of Nervous and Mental Disease*, 35 (2): 126.

Mugnier, M. R.; Stebbins, C. E., and Papavasiliou, F. N. 2016. Masters of Disguise: Antigenic Variation and the VSG Coat in *Trypanosoma brucei*. *PLoS Pathogens*, 12 (9): e1005784.

Munday, J. C.; Settimo, L., and De Koning, H. P. 2015. Transport proteins determine drug sensitivity and resistance in a protozoan parasite, *Trypanosoma brucei*. *Frontiers in Pharmacology*, 6: 32.

Murray, M.; d'Leteren, G., and Teale, A. 2004. Trypanotolerance, CAB International England. pp 461-477

Nagase, H.; Visse, R., and Murphy, G. 2006. Structure and function of matrix metalloproteinases and TIMPs. *Cardiovascular Research*, 69 (3): 562-573.

Nakahama, K.; Yoshimura, K.; Marumoto, R.; Kikuchi, M.; Lee, I. S.; Hase, T., and Matsubara, H. 1986. Cloning and sequencing of Serratia protease gene. *Nucleic Acids Research*, 14 (14): 5843-5855.

Namangala, B., and Odongo, S. 2014. Animal African Trypanosomiasis in Sub-Saharan Africa and beyond African Borders, Springer, USA. pp 240-252

Nantulya, V. 1990. Trypanosomiasis in domestic animals: the problems of diagnosis. *Scientific and Technical Review* 9(2): 357-367.

Nelson, A. L. 2010. Antibody fragments: hope and hype. *MAbs*, 2 (1): 77-83.

Nguewa, P. A.; Fuertes, M. A.; Cepeda, V.; Iborra, S.; Carrión, J.; Valladares, B.; Alonso, C., and Pérez, J. M. 2005. Pentamidine is an antiparasitic and apoptotic drug that selectively modifies ubiquitin. *Chemistry and Biodiversity*, 2 (10): 1387-1400.

Njiokou, F.; Laveissière, C.; Simo, G.; Nkinin, S.; Grébaut, P.; Cuny, G., and Herder, S. 2006. Wild fauna as a probable animal reservoir for *Trypanosoma brucei gambiense* in Cameroon. *Infection, Genetics and Evolution*, 6 (2): 147-153.

Odeyemi, S.; Iliyasu, D.; Omotosho, K.; Omolade, L.; Ajani, J., and Auwal, U. 2015. Hitches of Trypanosomosis in Nigeria from Livestock Production Perspective. *Methods in Molecular Biology*, 657: 93-105.

Odongo, S.; Sterckx, Y. G. J.; Stijlemans, B.; Pillay, D.; Baltz, T.; Muyldermans, S., and Magez, S. 2016. An Anti-proteome Nanobody Library Approach Yields a Specific Immunoassay for*Trypanosoma congolense* Diagnosis Targeting Glycosomal Aldolase. *PLoS Neglected Tropical Diseases*, 10 (2): e0004420.

Ohno, S. 2013. Evolution by gene duplication, Springer Science & Business Media, New York, USA. pp 59-65

Oladiran, A., and Belosevic, M. 2012. Recombinant glycoprotein 63 (Gp63) of *Trypanosoma carassii* suppresses antimicrobial responses of goldfish *(Carassius auratus l.)* monocytes and macrophages. *International Journal for Parasitology*, 42 (7): 621-633.

Oliver, M.; Atayde, V. D.; Isnard, A.; Hassani, K., and Shio, M. 2012. Leishmania virulence factors: focus on metalloprotease GP63. *Microbes and Infection*, 14: 1377-1389.

Overath, P.; Stierhof, Y.-D., and Wiese, M. 1997. Endocytosis and secretion in trypanosomatid parasites—tumultuous traffic in a pocket. *Trends in Cell Biology*, 7 (1): 27-33.

Park, A.; Matrisian, L.; Kells, A.; Pearson, R.; Yuan, Z., and Navre, M. 1991. Mutational analysis of the transin (rat stromelysin) autoinhibitor region demonstrates a role for residues surrounding the "cysteine switch". *Journal of Biological Chemistry*, 266 (3): 1584-1590.

Pays, E.; Vanhamme, L., and Perez-Morga, D. 2004. Antigenic variation in *Trypanosoma brucei:* facts, challenges and mysteries. *Current Opinion in Microbiology*, 7 (4): 369-374.

Peeling, R. W.; Holmes, K. K.; Mabey, D., and Ronald, A. 2006. Rapid tests for sexually transmitted infections (STIs): the way forward. *Sexually Transmitted Infections*, 82 (suppl 5): v1-v6.

Pelmenschikov, V.; Blomberg, M. R., and Siegbahn, P. E. 2002. A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *Journal of Biological Inorganic Chemistry*, 7 (3): 284-298.

Penchenier, L.; Grebaut, P.; Njokou, F.; Eyenga, V. E., and Büscher, P. 2003. Evaluation of LATEX/Tb gambiense for mass screening of *Trypanosoma brucei gambiense* sleeping sickness in Central Africa. *Acta Tropica*, 85 (1): 31-37.

Pentreath, V. W., and Kennedy, P. G. 2004. Pathogenesis of Human African trypanosomiasis, CAB International, England. pp 283-301

Pillai, B.; Cherney, M. M.; Hiraga, K.; Takada, K.; Oda, K., and James, M. N. 2007. Crystal structure of scytalidoglutamic peptidase with its first potent inhibitor provides insights into substrate specificity and catalysis. *Journal of Molecular Biology*, 365 (2): 343-361.

Pillay, D.; Izotte, J.; Fikru, R.; Büscher, P.; Mucache, H.; Neves, L.; Boulangé, A.; Seck, M. T.; Bouyer, J., and Napier, G. B. 2013. *Trypanosoma vivax* GM6 antigen: a candidate antigen for diagnosis of African animal trypanosomosis in cattle. *PLoS one*, 8 (10): e78565.

Polson, A.; Coetzer, T.; Kruger, J.; Von Maltzahn, E., and Van der Merwe, K. 1985. Improvements in the isolation of IgY from the yolks of eggs laid by immunized hens. *Immunological Investigations*, 14 (4): 323-327.

Posnett, D. N.; McGrath, H., and Tam, J. P. 1988. A novel method for producing anti-peptide antibodies: Production of site-specific antibodies to the T cell antigen receptor beta-chain. *Journal of Biological Chemistry*, 263 (4): 1719-1725.

Priotto, G.; Kasparian, S.; Mutombo, W.; Ngouama, D.; Ghorashian, S.; Arnold, U.; Ghabri, S.; Baudin, E.; Buard, V., and Kazadi-Kyanza, S. 2009. Nifurtimox-eflornithine combination therapy for second-stage African *Trypanosoma brucei gambiense* trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. *The Lancet*, 374 (9683): 56-64.

Qi, H.; Lu, H.; Qiu, H.-J.; Petrenko, V., and Liu, A. 2012. Phagemid vectors for phage display: properties, characteristics and construction. *Journal of Molecular Biology*, 417 (3): 129-143.

Ramamoorthy, R.; Donelson, J.; Paetz, K.; Maybodi, M.; Roberts, S., and Wilson, M. 1992. Three distinct RNAs for the surface protease gp63 are differentially expressed during development of *Leishmania donovani chagasi* promastigotes to an infectious form. *Journal of Biological Chemistry*, 267 (3): 1888-1895.

Rand, K. N. 1996. Crystal violet can be used to visualize DNA bands during gel electrophoresis and to improve cloning efficiency. *Technical Tips Online*, 1 (1): 23-24.

Rawlings, N. D., and Barrett, A. J. 1993. Evolutionary families of peptidases. *Biochemical Journal*, 290: 205-218.

Rawlings, N. D.; Barrett, A. J., and Bateman, A. 2011. Asparagine Peptide Lyases A Sebenth Catalytic Type of Proteolytic Enzymes. *Journal of Biological Chemistry*, 286 (44): 38321-38328.

Rawlings, N. D.; Waller, M.; Barrett, A. J., and Bateman, A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 42: D503-D509.

Rayaisse, J.; Tirados, I.; Kaba, D.; Dewhirst, S.; Logan, J.; Diarrassouba, A.; Salou, E.; Omolo, M.; Solano, P., and Lehane, M. 2010. Prospects for the development of odour baits to control the tsetse flies *Glossina tachinoides* and *G. palpalis s.l. PLoS Neglected Tropical Diseases*, 4 (3): e632.

Roberts, R. J.; Belfort, M.; Bestor, T.; Bhagwat, A. S.; Bickle, T. A.; Bitinaite, J.; Blumenthal, R. M.; Degtyarev, S. K.; Dryden, D. T., and Dybvig, K. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, 31 (7): 1805-1812.

Roberts, S. C.; Swihart, K. G.; Agey, M. W.; Ramamoorthy, R.; Wilson, M. E., and Donelson, J. E. 1993. Sequence diversity and organization of the *msp* gene family encoding gp63 of *Leishmania chagasi Molecular and Biochemical Parasitology*, 62 (2): 157-171.

Robinson, D. R.; Sherwin, T.; Ploubidou, A.; Byard, E. H., and Gull, K. 1995. Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. *The Journal of Cell Biology*, 128 (6): 1163-1172.

Robinson, N. P.; Burman, N.; Melville, S. E., and Barry, J. D. 1999. Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. *Molecular and Cellular Biology*, 19 (9): 5839-5846.

Rosano, G. L., and Ceccarelli, E. A. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5: 172.

Roy, A.; Kucukural, A., and Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5 (4): 725-738.

Russell, D., and Wilhelm, H. 1986. The involvement of the major surface glycoprotein (gp63) of Leishmania promastigotes in attachment to macrophages. *The Journal of Immunology*, 136 (7): 2613-2620.

Russell, D. G. 1987. The macrophage-attachment glycoprotein gp63 is the predominant C3-acceptor site on *Leishmania mexicana* promastigotes. *European Journal of Biochemistry*, 164 (1): 213-221.

Sapats, S.; Heine, H.; Trinidad, L.; Gould, G.; Foord, A.; Doolan, S.; Prowse, S., and Ignjatovic, J. 2003. Generation of chicken single chain antibody variable fragments (scFv) that differentiate and neutralize infectious bursal disease virus (IBDV). *Archives of virology*, 148 (3): 497-515.

Schechter, I., and Berger, A. 1967. Protease subsite nomenclature. *Biochemical and Biophysical Research Communications*, 27: 157-162.

Schechter, I., and Berger, A. 1968. On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochemical and Biophysical Research Communications*, 32 (5): 898-902.

Schlagenhauf, E.; Etges, R., and Metcalf, P. 1998. The crystal structure of the *Leishmania major* surface proteinase leishmanolysin (gp63). *Structure*, 6 (8): 1035-1046.

Schomburg, D., and Stephan, D. 2012. Enzyme Handbook 16: First Supplement Part 2 Class 3: Hydrolases, vol. 1,Spinger-Verlag, Germany. pp 1-3

Shapiro, S. D.; Fliszar, C. J.; Broekelmann, T. J.; Mecham, R. P.; Senior, R. M., and Welgus, H. G. 1995. Activation of the 92-kDa Gelatinase by Stromelysin and 4-Aminophenylmercuric Acetate. *Journal of Biological Chemistry*, 270 (11): 6351-6356.

Shaw, A.; Cecchi, G.; Wint, G.; Mattioli, R., and Robinson, T. 2014. Mapping the economic benefits to livestock keepers from intervening against bovine trypanosomosis in Eastern Africa. *Preventive Veterinary Medicine*, 113 (2): 197-210.

Shimogawa, M. M.; Saada, E. A.; Vashisht, A. A.; Barshop, W. D.; Wohlschlegel, J. A., and Hill, K. L. 2015. Cell surface proteomics provides insight into stage-specific remodeling of the host-parasite interface in *Trypanosoma brucei*. *Molecular & Cellular Proteomics*, 14 (7): 1977-1988.

Simarro, P. P.; Diarra, A.; Postigo, J. A. R.; Franco, J. R., and Jannin, J. G. 2011. The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000–2009: the way forward. *PLoS Neglected Tropical Diseases*, 5 (2): e1007.

Simpson, S., and Casas, J. 2009. Advances in Insect Physiology: Physiology of Human and Animal Disease Vectors, Academic Press, England. pp 121-122

Sims, A. H.; Dunn-Coleman, N. S.; Robson, G. D., and Oliver, S. G. 2004. Glutamic protease distribution is limited to filamentous fungi. *FEMS Microbiology Letters*, 239 (1): 95-101.

Skottrup, P. D. 2010. Small biomolecular scaffolds for improved biosensor performance. *Analytical Biochemistry*, 406 (1): 1-7.

Smith, G. P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228 (4705): 1315-1317.

Smyth, J. D. 1994. Introduction to animal parasitology, Cambridge University Press, USA. pp 58-69

Solano, P.; Jamonneau, V.; N'guessan, P.; N'Dri, L.; Dje, N.; Miezan, T.; Lejon, V.; Büscher, P., and Garcia, A. 2002. Comparison of different DNA preparation protocols for PCR diagnosis of Human African Trypanosomosis in Cote d'Ivoire. *Acta Tropica*, 82 (3): 349-356.

Soteriadou, K. P.; Tzinia, A. K.; Hadziantoniou, M. G., and Tzartos, S. J. 1988. Identification of monomeric and oligomeric forms of a major *Leishmania infantum* antigen by using monoclonal antibodies. *Infection and Immunity*, 56 (5): 1180-1186.

Stevens, J.; Brisse, S.; Maudlin, I.; Holmes, P., and Miles, M. 2004. Systematics of trypanosomes of medical and veterinary importance, CAB International, England. pp 1-23

Stevens, J. R., and Gibson, W. 1999. The molecular evolution of trypanosomes. *Parasitology Today*, 15 (11): 432-437.

Steverding, D. 2008. The history of African trypanosomiasis. *Parasite Vectors*, 1 (3): 411.

Stocker, W., and Bode, W. 1995. Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Current Opinion in Structural Biology*, 5 (3): 383-390.

Stocker, W.; Gomis-Ruth, F.-X.; Bode, W., and Zwilling, R. 1993. Implications of the three-dimensional structure of astacin for the structure and function of the astacin family of zinc endopeptidases. *European Journal of Biochemistry*, 214 (1): 215-231.

Stocker, W.; Grams, F.; Reinemer, P.; Bode, W.; Baumann, U.; Gomis-Rüth, F.-X., and Mckay, D. B. 1995. The metzincins—Topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a super family of zinc-peptidases. *Protein Science*, 4 (5): 823-840.

Stricklin, G. P.; Jeffrey, J. J.; Roswit, W. T., and Eisen, A. Z. 1983. Human skin fibroblast procollagenase: mechanisms of activation by organomercurials and trypsin. *Biochemistry*, 22 (1): 61-68.

Tallant, C.; García-Castellanos, R.; Baumann, U., and Gomis-Rüth, F. X. 2010. On the relevance of the Met-turn methionine in metzincins. *Journal of Biological Chemistry*, 285 (18): 13951-13957.

Taylor, J. E., and Rudenko, G. 2006. Switching trypanosome coats: what's in the wardrobe? *Trends in Genetics*, 22 (11): 614-620.

Taylor, K., and Authié, E. M.-L. 2004. Pathogenesis of Animal Trypanosomiasis, CAB International, England. pp 331-353

Telleria, E. L.; Benoit, J. B.; Zhao, X.; Savage, A. F.; Regmi, S.; Alves e Silva, T. L.; O'Neill, M., and Aksoy, S. 2014. Insights into the trypanosome-host interactions revealed through transcriptomic analysis of parasitized tsetse fly salivary glands. *PLoS Neglected Tropical Diseases*, 8 (4): e2649.

Telleria, J., and Tibayrenc, M. 2010. American trypanosomiasis: Chagas disease one hundred years of research, Elsevier, USA. pp 25-44

Thomas, P. T., and Woo, P. T. 1991. In vitro and in vivo effects of antimicrobial agents on viability of *Cryptobia salmositica* (*Sarcomastigophora: Kinetoplastida*). *Diseases of Aquatic Organisms*, 10 (1): 7-11.

Tiberti, N.; Hainard, A., and Sanchez, J.-C. 2013. Translation of human African trypanosomiasis biomarkers towards field application. *Translational Proteomics*, 1 (1): 12-24.

Tzinia, A. K., and Soteriadou, K. P. 1991. Substrate-dependent pH optima of gp63 purified from seven strains of *Leishmania*. *Molecular and Biochemical Parasitology*, 47 (1): 83-89.

Uilenberg, G. 1998. A field guide for the diagnosis, treatment and prevention of African animal trypanosomosis, Food & Agriculture Organisation Rome. pp 43-135

Valentin, G. 1841. Ueber ein Entozoon im Blute von *Salmo fario*. *Archiv für Anatomie, Physiologie und Wissenschaftliche Medicin.*: 435–436.

Vallejo, G.; Guhl, F., and Schaub, G. 2009. Triatominae–*Trypanosoma cruzi* and *T. rangeli*: Vector–parasite interactions. *Acta Tropica*, 110 (2): 137-147.

Van Bogaert, l., and Haemers, A. 1989. Eflornithine. *Pharmaceutisch Weekblad*, 11 (3): 69-75.

Van Wart, H. E., and Birkedal-Hansen, H. 1990. The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proceedings of the National Academy of Sciences of the USA*, 87 (14): 5578-5582.

Van Wyngaardt, W.; Malatji, T.; Mashau, C.; Fehrsen, J.; Jordaan, F.; Miltiadou, D., and Du Plessis, D. H. 2004. A large semi-synthetic single-chain Fv phage display library based on chicken immunoglobulin genes. *BMC Biotechnology*, 4 (1): 6.

Vickerman, K. 1985. Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin*, 41 (2): 105-114.

Vickerman, K.; Tetley, L.; Hendry, K. A., and Turner, C. M. R. 1988. Biology of African trypanosomes in the tsetse fly. *Biology of the Cell*, 64 (2): 109-119.

Visse, R., and Nagase, H. 2003. Matrix metalloproteinases and tissue inhibitors of metalloproteinases structure, function, and biochemistry. *Circulation Research*, 92 (8): 827-839.

Voth, B. R.; Kelly, B. L.; Joshi, P. B.; Ivens, A. C., and McMaster, W. R. 1998. Differentially expressed *Leishmania major gp63* genes encode cell surface leishmanolysin with distinct signals for glycosylphosphatidylinositol attachment. *Molecular and Biochemical Parasitology*, 93 (1): 31-41.

Vreysen, M. 2000. Principles of area-wide integrated tsetse fly control using the sterile insect technique. *Medecine tropicale: revue du Corps de sante colonial*, 61 (4-5): 397-411.

Vreysen, M. J.; Saleh, K. M.; Ali, M. Y.; Abdulla, A. M.; Zhu, Z.-R.; Juma, K. G.; Dyck, V. A.; Msangi, A. R.; Mkonyi, P. A., and Feldmann, H. U. 2000. *Glossina austeni* (*Diptera: Glossinidae*) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *Journal of Economic Entomology*, 93 (1): 123-135.

Wang, Y. A.; Yu, X.; Overman, S.; Tsuboi, M.; Thomas, G. J., and Egelman, E. H. 2006. The structure of a filamentous bacteriophage. *Journal of Molecular Biology*, 361 (2): 209-215.

Welling, G.; Weijer, W.; Van Der Zee, R., and Welling-Wester, S. 1985. Prediction of sequential antigenic regions in proteins. *FEBS Letters*, 188: 215-218.

Wheeler, R. J. 2010. The trypanolytic factor–mechanism, impacts and applications. *Trends in Parasitology*, 26 (9): 457-464.

Wilson, M. E., and Hardin, K. K. 1988. The major concanavalin A-binding surface glycoprotein of *Leishmania donovani chagasi* promastigotes is involved in attachment to human macrophages. *The Journal of Immunology*, 141 (1): 265-272.

Woo, P. 1970. The haematocrit centrifuge technique for the diagnosis of African trypanosomiasis. *Acta Tropica*, 27 (4): 384-386.

Zhang, Y. 2007. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins: Structure, Function, and Bioinformatics*, 69 (S8): 108-117.

Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9 (1): 40.