STRUCTURAL STUDIES AIMED AT IMPROVING THE ANTIGENICITY OF CONGOPAIN

HLUMANI HUMPHREY NDLOVU

BSc (Hons) Biochemistry (*cum laude*)

Submitted in fulfilment of the academic requirement for the degree of
Master of Science in the Discipline of Biochemistry,
School of Biochemistry, Genetics and Microbiology,
University of KwaZulu-Natal,

Pietermaritzburg
May 2009

PREFACE

The experimental work described in this dissertation was carried out at the School of Biochemistry, Genetics and Microbiology, University of KwaZulu-Natal, Pietermaritzburg, from January 2007 to December 2008, under the supervision of Prof. Theresa Coetzer and co-supervision of Dr Alain Boulangé.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma at any University. Where use has been made of the work of others, it is duly acknowledged in the text.

_____

Hlumani H. Ndlovu (candidate)

_____

Prof. Therese H. T. Coetzer (supervisor)

_____

Dr Alain F. V. Boulangé (co-supervisor)

DEDICATION

I dedicate this work to my late father, Mfanozi Solomon Ndlovu and sister, Nombuso Patience Ndlovu. May your souls rest in eternal peace.

DECLARATION - PLAGIARISM

I, Hlumani Humphrey Ndlovu declare that

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

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

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

4. This thesis 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 thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed: ……

ABSTRACT

African animal trypanosomosis or nagana is a tsetse fly-transmitted disease, caused by *Trypanosoma congolense*, *T. vivax* and to a lesser extent *T. brucei brucei*. The disease causes major losses in revenue in many livestock-producing African countries. The available control methods, including chemotherapeutic drugs and insecticidal spraying, have become environmentally unacceptable. Antigenic variation displayed by the parasites has hindered vaccine development efforts. In this context, rather than focusing solely on the parasite itself, efforts in vaccine development have shifted towards targeting pathogenic factors released by the parasites during infection.

Congopain, the major cysteine protease of *T. congolense*, has been shown to act as a pathogenic factor in the disease process. Analysis of the immune response of trypano-tolerant cattle revealed that these animals have the ability to control congopain activity *in vivo*. Therefore, congopain is an attractive vaccine candidate. To test the protective potential of congopain, immunisation studies had been conducted in cattle using the baculovirus-expressed catalytic domain of congopain (C2) in RWL, a saponin-based proprietary adjuvant from SmithKline-Beecham. Immunised animals were partially protected against a disease caused by an infection with *T. congolense*. Unfortunately, subsequent attempts to reproduce these results were disappointing. It was hypothesised that this failure could be due to the different expression system (*P. pastoris*) used to produce the antigen (C2), or the different adjuvant, ISA206 (Seppic), used, thus hinting towards an epitope presentation problem. Congopain had been shown to dimerise at physiological pH *in vitro*. Sera from trypano-tolerant cattle preferentially recognised the dimer conformation, advocating for protective epitopes to be dimer associated. For that reason, the present study aimed at improving the antigenicity of congopain through firstly, the elucidation of the protective epitopes associated with the dimer, secondly, the determination of the 3-D structure of the protease in order to map protective epitopes to later design mimotopes, and thirdly improve the delivery of congopain to the immune cells while maintaining the conformation of the protease by using a molecular adjuvant, BiP.

A dimerisation model was proposed, identifying the amino acid residues forming the dimerisation motif of congopain. In the present study, particular amino acid residues located in

the dimerisation motif were mutated by PCR-based site-directed mutagenesis to generate mutants with different dimerisation capabilities. The congopain mutants were expressed in yeast and their dimerisation capability was assessed by PhastGel® SDS-PAGE. The mutations altered both the electrophoretic mobility of the mutants and their enzymatic characteristics compared to wild-type congopain. This advocated for the involvement of these amino acid residues in the dimerisation process, although they seem not to be the only partakers.

Wild-type C2 and mutant forms of C2 were heterologously expressed in *P. pastoris* and purified to crystallisation purity levels. Crystallisation of these proteins is currently underway, but the results are still unknown. While awaiting the crystallisation results, *in silico* homology modelling was employed to gain insight into the 3-D structure, using cruzipain crystal structure as a template. The modelled 3-D structure of congopain followed the common framework of cathepsin L-like cysteine proteases. Due to time constraints and awaiting the crystal-derived 3-D structure, the 3-D model of congopain was not exploited to design mimotopes with the potential to provide protection against the disease.

As it was shown that protective epitopes are likely to be dimer-specific, maintaining the native conformation of congopain is essential for stimulating a protective immune response in animals. Chemically formulated adjuvants usually contain high salt concentration, at acidic or basic pH, thus might change the conformation of the protease. Adjuvants capable of efficiently delivering the antigen to immune cells while maintaining the conformation of the protease were sought. Proteins belonging to the HSP70 family are natural adjuvants in higher eukaryotes. A protein belonging to the HSP70 family was previously identified in *T. congolense* lysates and is homologous to mammalian BiP. Congopain was genetically fused with *T. congolense* BiP in order to improve antigen delivery and production of congopain activity-inhibiting antibodies. The chimeric proteins were successfully expressed in both bacteria and yeasts. The low yields of recombinantly expressed chimeras in yeast and problems associated with renaturation and purification of bacteria-expressed chimeras prevented immunisation studies in mice. However, the groundwork was laid for producing BiP-congopain chimeras for use in an anti-disease vaccine for African trypanosomosis.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following organisations and people:

The National Research Foundation and the French Embassy in South Africa for financial assistance.

My supervisor Professor Theresa Coetzer for her expert advice, guidance, patience and assistance with the preparation of this dissertation.

Dr. Alain Boulangé for co-supervising this study and critical reading of this dissertation.

Charmaine Ahrens and Robyn Hillebrand for their assistance with administrative issues.

My fellow labmates Ike Achilonu, Cara-Lesley Bartlett, Lorelle Bizaaré, Titos Cau, Dave Choveaux, Bridgette Cumming, Sabelo Hadebe, Ramona Hurdayal, Richard Kangethe, Hermógenes Mucache, Davita Pillay, Perina Vather, Jackie Viljoen and Phillia Vukea for their friendship and tolerance.

My colleagues at Student Housing, Cebisile Gumede, Nhlanhla Mbudu, Lindelani Mnguni, Mjabuliseni Ngidi, Sibonelo Shozi and Nompumelelo Xingwana for their friendship and advice.

To my boys, Smiso Bhengu, Bonginkosi Mbatha, Monde Mchunu, Mlondi Ngcobo, Sthembiso Ngubo and Zibonele Nxele for grooming me to become a man.

My mother and all my siblings for their unconditional love, understanding and support.

Silindile for being my friend, lover and mother of our son.

Lastly, to God for giving me the gift of life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

2×YT	2 × yeast tryptone medium
3-D	three-dimensional
AMC	7-amino-4-methylcoumarin
BCA	bicinchoninic acid
BiP	immunoglobulin heavy-chain binding protein
Bis-Tris	2-bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol
BMGY	buffered media glycerol yeast extract
BMM	buffered minimal media
BSA	bovine serum albumin
C2	catalytic domain of congopain
CP	cysteine protease
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
E-64	*trans*-epoxysuccinyl-L-leucyl-amido(4-guanidino)butane
g	relative centrifugal force
HRPO	horse radish peroxidase
IgG	immunoglobulin G
IPTG	isopropyl-beta-D-thiogalactopyranoside
kDa	kilodalton
MEC	molecular exclusion chromatography
M_r	relative molecular mass
OD_{600}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
RT	room temperature
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulfate

TBS	tris buffered saline
TPP	three phase partitioning
VAT	variable antigen type
VSG	variable surface glycoprotein
YNB	yeast nitrogen base
YPD	yeast peptone dextrose media
Z	benzyloxycarbonyl

CHAPTER 1

INTRODUCTION

1.1 African animal trypanosomosis

African animal trypanosomosis is a wasting disease that severely limits livestock production in the developing world. It is caused by protozoan parasites of the genus *Trypanosoma*, generally transmitted by tsetse flies. The disease is estimated to cost US$ 1.3-5 billion per annum in Africa alone (Kristjanson *et al*., 1999; Shaw, 2004). Moreover, a third of the African continent is infested with tsetse flies, preventing the use of a large area of land for livestock farming (Kristjanson *et al*., 1999). Distribution of the trypanosomes coincides with the presence of the tsetse fly vector in Africa while the disease also occurs in Asia and South America, transmitted by biting flies (Stevens and Brisse, 2004).

Current methods used to control trypanosomosis are inadequate. The available trypanocidal drugs (isometamidium, diaminazene aceturate and homidium) have been on the market for over five decades. It is thus not surprising that drug-resistant parasites have been reported in Africa (Anene *et al*., 2001; Geerts *et al*., 2001). Although control of the tsetse fly vector using insecticidal spraying and bush clearing was effective, their negative impact on the environment is unacceptable. Hence, there is an urgent need to develop new methods to control African animal trypanosomosis.

1.1.1 Pathogenesis of African animal trypanosomosis

Nagana (caused by *T. congolense*, *T. vivax* and to a lesser extent *T. b. brucei*), surra (caused by *T. evansi* in equines, water buffaloes and camels) and dourine (caused by *T. equiperdum* in horses) represent the three disease states caused by pathogenic trypanosomes. The geographical location of a trypanosome isolate seems to contribute to varied pathogenicity between parasite stocks of the same species. Some *T. vivax* isolates from East Africa cause an acute haemorrhagic disease in cattle, while the West African isolates result in milder non-haemorrhagic disease (Taylor and Authié, 2004). Furthermore, infection with haemorrhagic strains of *T. vivax* seems to be more pathogenic to cattle than *T. congolense* and *T. b. brucei* as indicated by high parasitemia (10^8 parasites/ml of blood) in *T. vivax* infected cattle (Taylor and Authié, 2004).

Generally, bovine trypanosomosis is characterised by three successive stages: acute, asymptomatic/stabilisation and chronic stages that are greatly influenced by the host, trypanosome species and strain of infecting trypanosome (Taylor and Authié, 2004). The acute phase of the disease is characterised by a severe drop in haematocrit value (packed cell volume, PCV), haemoglobin concentration and red blood cells numbers. Clinical signs of acute bovine trypanosomosis are anaemia, enlargement of the spleen and the lymph nodes, weakness, lethargy, loss of condition, abortion and reduced milk production. Anaemia is the major cause of death in animals infected with *T. congolense* and *T. vivax* (Taylor, 1998). The animals that survive the acute phase stabilise the infection after 6-8 weeks and begin a slow recovery process. These animals enter into a chronic phase of infection that is characterised by cachexia, intermittent parasitemia, stunting, wasting and infertility. Finally, the animals die of the chronic infection due to congestive heart failure, prolonged anaemia, myocardial damage and increased vascular permeability (Taylor and Authié, 2004).

1.1.2 Trypanotolerance

Certain breeds of cattle have long been recognised to possess a genetic ability to survive and remain productive in tsetse-infested areas without chemotherapeutic treatment. These animals are termed trypano-tolerant and include the *Bos taurus* N'Dama, Baoulé, Muturu and Dahomey breeds, found mainly in West Africa (Authié, 1994; d'Ieteren *et al*., 1998; Taylor, 1998; Murray *et al*., 2004). The *B. indicus* (Zebu) species of cattle is susceptible to infection with trypanosomes as indicated by their failure to survive in tsetse-infested areas without treatment. Trypano-tolerant cattle are able to control anaemia and parasitemia (d'Ieteren *et al*., 1998). Many species of wild herbivores such as the Cape buffalo are resistant to an infection with trypanosomes as indicated by their ability to survive in tsetse infested areas (Black *et al*., 2001). These animals serve as reservoirs of parasites that infect domestic animals (and humans).

1.2 Diagnosis of African animal trypanosomosis

Survival of infected cattle relies on rapid and specific diagnosis of the disease in the field. The efficacy of trypanocidal drugs depends greatly on the stage of infection the animal is at during their administration. Treatment of cattle in the early acute phase of infection results in a rapid

response to trypanocidal drugs, while cattle treated in the late chronic phase of infection take much longer to respond to the treatment or die of it (Taylor and Authié, 2004).

Diagnosis of African trypanosomosis relies mainly on clinical and parasitological methods, although serological and molecular tools are becoming available. The clinical methods are most relevant in field conditions, where physical examination is often the sole means of diagnosis available (Eisler *et al*., 2004). The symptoms displayed by trypanosome infected cattle are not characteristic, thus, may often be mistaken for one of many other tropical diseases.

Parasitological methods of diagnosis involve direct visualisation of the parasites in the blood of infected animals using microscopic techniques. Besides the low sensitivity of this technique, especially in detecting parasites in the chronic phase of infection, the high level of morphological similarity exhibited by different trypanosome species makes differential diagnosis difficult. Therefore, serological methods for detection of circulating parasite antigens or specific antibodies against parasite antigens, and molecular tools, have been developed. Serological methods include indirect immunofluorescent antibody (IFA) tests, direct and indirect enzyme-linked immunosorbent assays (ELISAs), and complement fixation tests. Species-specific diagnosis was not achieved with the IFA test since cross-reactivity was detected between *T. congolense*, *T. vivax* and *T. brucei* (Eisler *et al*., 2004). However, direct and indirect ELISA allowed for species-specific diagnosis depending on the antigen used in the ELISA setup (Eisler *et al*., 1998). Molecular biology tools used to diagnose bovine trypanosomosis include the use of radioactively labelled DNA probes and polymerase chain reaction (PCR). Between the two methods, PCR is the best since it allows for species-specific diagnosis of the parasites and is more sensitive than radioactively labelled DNA probes (Desquesnes and Davila, 2002).

1.3 Control of trypanosomosis

1.3.1 Tsetse control

Control of tsetse flies using conventional methods, including aerial and ground spraying with insecticides, insecticide-treated cattle, odour-baited traps and bush-clearing, is very effective (Allsopp and Hursey, 2004; Vale and Torr, 2004, 2005). In Nigeria, insecticides were used to eradicate tsetse flies in an area of approximately 200 000 km^2 (Schofield and Maudlin, 2001).

Although insecticides such as DTT and benzene hexachloride are very effective against the vector, their application was regarded as unsound due to their negative impact on the environment (Allsopp and Hursey, 2004). Less conventional, the Pan-African Tsetse and Trypanosomosis Eradication Campaign (PATTEC) launched a large scale tsetse control project funded by the African Development Bank. The project was aimed at using the sterile insect technique as was done in the USA to eradicate the screwworm fly, and in Zanzibar for *Glossina austeni* to eradicate the tsetse fly populations (Torr *et al*., 2005). However, there are serious doubts about the feasibility of the project over the large area of land infested with tsetse flies (Rogers and Randolph, 2002).

1.3.2 Chemotherapy

The administration of trypanocidal drugs still remains the principal method for controlling African animal trypanosomosis. There are approximately 35 million doses of trypanocidal drugs administered each year by farmers (Holmes *et al*., 2004). Chemotherapeutic treatment of cattle relies on three trypanocidal drugs: Samorin® (isometamidium chloride), Novidium® (homidium chloride) and Berenil® (diaminazene aceturate). The exact mechanism of action of these drugs has not been fully elucidated; they are thought to target kinetoplast DNA (Holmes *et al*., 2004). Trypanocidal drugs such as pentamidine, diminazene and suramin have been shown to act by inhibiting oligopeptidase B, a trypanosomal serine protease and a pathogenic factor (Morty *et al*., 1998). These trypanocidal drugs are supplied by European pharmaceutical companies at high cost. However, generic forms of the drugs have been available in the African market for a long time, but doubts surrounding their quality have often hindered their application by farmers (Holmes *et al*., 2004).

Emergence of drug-resistant parasites has been reported in over thirteen countries in Africa (Geerts *et al*., 2001; Holmes *et al*., 2004). There is thus an urgent need to develop new trypanocidal drugs for the treatment of African animal trypanosomosis. However, pharmaceutical companies are not interested in pursuing the development of new trypanocidal drugs due to the unprofitability of the market. Studies have revealed that the total market value of trypanocidal drugs for farmers is approximately US$ 30 million per year, while the cost of developing and licensing a single drug is estimated to cost US$ 250 million (Holmes *et al*., 2004; Delespaux *et*

al., 2008). Therefore, African farmers are left with the only option of using the old trypanocidal drugs.

1.3.3 Vaccine

Emergence of drug-resistant parasites has been the impetus for research towards the development of an effective vaccine for the disease. Antigenic variation displayed by trypanosome species precludes vaccination with trypanosomes against bovine trypanosomosis (Donelson, 2003). Therefore, research has shifted towards targeting accessible invariant antigens of the parasites. Flagellar pocket antigens derived from *T. b. rhodesiense* protected cattle against infection with *T. congolense* and *T. vivax* (Mkunza *et al*., 1995). Immunisation studies conducted in mice using native tubulin isolated from *T. brucei*, provided complete protection against an infection with *T. brucei* and partial protection against *T. congolense* and *T. b. rhodesiense* (Lubega *et al*., 2002). Similar results were obtained using recombinant tubulin from *T. evansi* (Li *et al*., 2007).

The concept of an anti-disease vaccine was first developed for malaria (Playfair *et al*., 1990). This concept is based on the assumption that pathogenic factors that are either secreted or released by dead parasites rather than the parasite itself cause the disease. Thus, antibodies inhibiting the activity of pathogenic factors are expected to provide protection against the disease (Playfair *et al*., 1990; Playfair *et al*., 1991; Authié, 1994; Schofield, 2007). A cysteine protease, named congopain, has been implicated in the pathology of trypanosomosis caused by infection with *T. congolense*. Immunisation experiments conducted with congopain in cattle provided partial protection against *T. congolense* infection (Authié *et al*., 2001). This suggested possible involvement of further pathogenic factors in bovine trypanosomosis. Identification of such factors and their inclusion in a multi-component vaccine should provide better protection against an infection with *T. congolense*.

1.4 Trypanosomes

Trypanosomes are protozoan parasites of medical and veterinary importance. Sir David Bruce demonstrated their involvement in an outbreak of nagana in the Zululand region of South Africa in 1894, that was responsible for the death of many cattle (Duggan, 1977). The tsetse transmitted *Trypanosoma congolense*, *T. vivax* and *T. b. brucei* are the main causative agents of bovine

trypanosomosis in sub-Saharan Africa (Taylor, 1998; Stevens and Brisse, 2004). However, *T. vivax* is endemic in Central and South America, where the parasite is mechanically transmitted by biting flies (Otte *et al*., 1994). The non-tsetse transmitted *T. evansi* is widespread in sub-Saharan Africa, South America, Asia and has been reported in places as remote from the tropics as France and Australia (Payne *et al*., 1991; Franke *et al*., 1994). The infectivity of trypanosomes is not limited to cattle; they infect a wide range of domestic animals including horses, camels, sheep, goats, dogs and pigs (Stevens and Brisse, 2004). *Trypanosoma b. gambiense* and *T. b. rhodesiense* are the two sub-species of trypanosomes infective to humans, causing sleeping sickness or human African trypanosomosis (Stevens and Brisse, 2004). *Trypanosoma cruzi* occurs in South America and causes Chagas' disease in humans (Schmidt and Roberts, 1989; Stevens and Brisse, 2004).

1.4.1 Classification of trypanosomes

The classification of trypanosomes of medical and veterinary importance is provided in Fig. 1.1. African trypanosomes belong to the order *Kinetoplastida*, characterised by the presence of a kinetoplast at the base of the flagellum that contains 20% of the total parasite DNA, and a single flagellum extending from the flagellar pocket, as an attached undulating membrane (El-Sayed *et al*., 2000; Stevens and Brisse, 2004). The family *Trypanosomatidae* is subdivided into two genera, *Trypanosoma* and *Leishmania*. Genus classification is based on morphological features and host range of the parasites (Momen, 2001).

The *Trypanosoma* genus is divided into two sections that develop in different parts of the vector. In the section *Stercoraria*, the trypanosomes develop in the hindgut of the vector and are transmitted in faeces by the vector. In section *Salivaria*, the trypanosomes are inoculated with saliva into the host via the mouth parts of the vector. In addition, salivarian trypanosomes can change their surface coat, a phenomenon known as antigenic variation since they contain and express variable surface glycoprotein (VSG) genes (Donelson, 2003; Stevens and Brisse, 2004). At present, the *Salivaria* section contains four subgenera, *viz. Duttonela*, *Nannomonas*, *Trypanozoon*, and *Pycnomonas*. The *Nannomonas* subgenus contains three species: *T. congolense*, *T. godfreyi* and *T. simiae*. *T. congolense* is the most common pathogen of ruminants in Africa, while *T. godfreyi* and *T. simiae* are pathogenic to pigs (Stevens and Brisse, 2004). These three species are difficult to distinguish due to morphological similarities and identical

developmental cycle in the tsetse vector. *T. congolense* has been further divided into three groups based on the group of vector that carries the parasites. The three groups are savannah, riverine/forest and Kenya (Kilifi) coast. Unlike the other species belonging to the *Nannomonas* subgenus, *T. congolense* can be easily cultured in laboratory rodents (Stevens and Brisse, 2004).

1.4.2 Biology of African trypanosomes

1.4.2.1 Morphology and genomic organisation

T. congolense (Fig. 1.2), is a unicellular, small and short trypanosome of between 8 and 24 μm in length. The parasite possesses a single flagellum that extends from the flagellar pocket to the posterior end of a parasite (Stevens and Brisse, 2004). *T. congolense* lacks a free flagellum, thus can be easily distinguished from *T. brucei* that possesses a free flagellum (Stevens and Brisse, 2004). The parasite contains a Golgi apparatus, sac of secretion and a secretory reticulum (Stevens and Brisse, 2004).

The kinetoplast and the nucleus contain the genomic DNA of the parasite. The highly unusual kinetoplast consists of a network of interlinked maxicircles and minicircles (El-Sayed and Donelson, 1997; Ersfeld *et al*., 1998; El-Sayed *et al*., 2000). The genome of *T. brucei* has been completely sequenced (www.sanger.ac.uk/Project/T.brucei). Although the genomes of *T. congolense* and *T. vivax* have not yet been completely sequenced, extrapolations from the genome of *T. brucei* are informative (Melville *et al*., 2004). The nuclear genome of *T. brucei* is approximately 35 Mb in size, roughly the same size as the genomes of *Plasmodium falciparum*, *Leishmania major* and *T. cruzi* (Melville *et al*., 2004). By comparison, the nuclear genome of *Saccharomyces cerevisae* is 12 Mb and contains 6 000 genes, while the *Caenorhabditis elegans* genome is 97 Mb with 19 000 genes (Mewes *et al*., 1997). Hence, the *T. brucei* genome is expected to harbour 12 000 genes with 50% of the genes coding for proteins involved in housekeeping (El-Sayed and Donelson, 1997).

1.4.2.2 Antigenic Variation

The membrane of bloodstream form trypanosomes is covered by a variant surface glycoprotein (VSG) coat (Barry and McCulloch, 2001; Donelson, 2003; Barry *et al*., 2005). The VSG coat is

the most exposed, immunodominant antigen that is recognised by the host's immune system. However, the parasite possesses a genetic capacity to alter the VSG coat by a process known as antigenic variation (Barry and McCulloch, 2001; Vanhamme *et al*., 2001). Each distinct VSG coat is termed variant antigen type (VAT). Although anti-VSG antibodies targeting a particular VAT are able to modulate parasitemia, trypanosomes expressing a different VAT evade the host's immune response, giving rise to a new wave of parasitemia. Therefore, trypanosome infection is characterised by fluctuations (waves) of parasitemia indicating a different VAT on the surfaces of the parasites. Expression of VSG genes occurs by duplication of the usually non-transcribed regions of a minichromosome into transcriptionally active subtelomeric regions (Barry *et al*., 2005).

Immunisation of cattle with irradiated trypanosomes or purified VSG provides protection against the same VAT only (Morrison *et al*., 1982; Wells *et al*., 1982). Antigenic variation exhibited by trypanosomes hence prevents the development of a VSG-based vaccine for bovine trypanosomosis.

1.4.3 Life cycle of African trypanosomes

African trypanosomes are generally transmitted by the bite of a tsetse fly (*Glossina* spp). Both male and female tsetse flies transmit the parasites (Aksoy, 2003). At present, there are 23 species and eight subspecies of the *Glossina* genus which are divided into three groups: *fusca* (forest), *palpalis* (riverine) and *morsitans* (savannah) (Uilenberg, 1998).

T. congolense, *T. vivax* and *T. brucei* spend part of their developmental cycle in the tsetse fly vector and part in the mammalian host (Fig. 1.3). The life cycle begins when a tsetse fly transmits non-dividing metacyclic forms of the parasites with saliva during a bloodmeal to its mammalian host. This causes swelling of the skin at the bite site, known as the chancre (Naessens, 2006). The parasites migrate via the lymphatic vessels to the local lymph nodes and from there into the bloodstream of the host, where they differentiate into rapidly dividing short bloodstream forms. The bloodstream form parasites are taken up by the tsetse fly vector as it feeds on the infected animal (Authié, 1994).

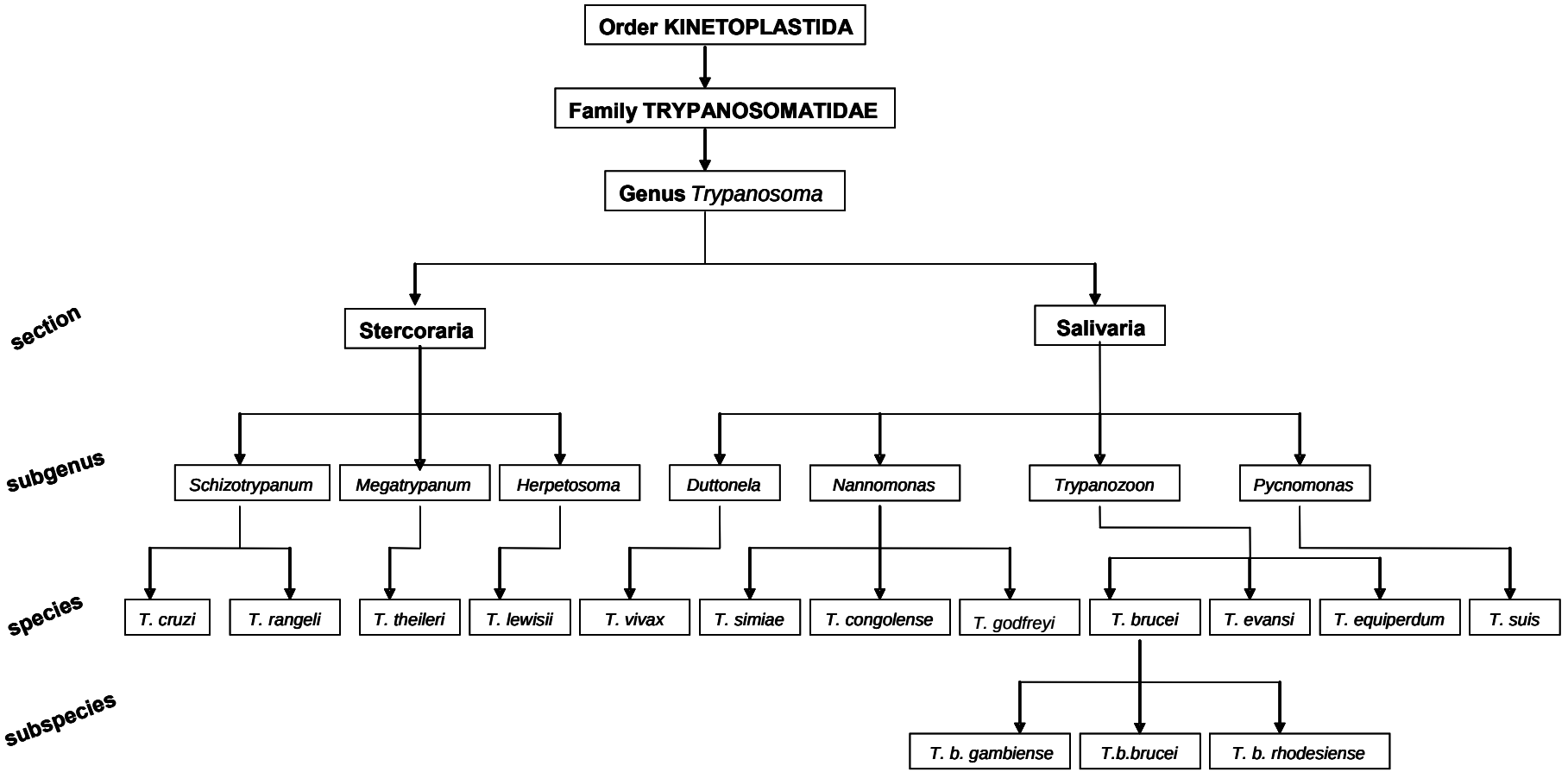

Figure 1.1. Classification of protozoan parasites belonging to the *Kinetoplastida* order (adapted from Stevens and Brisse, 2004).

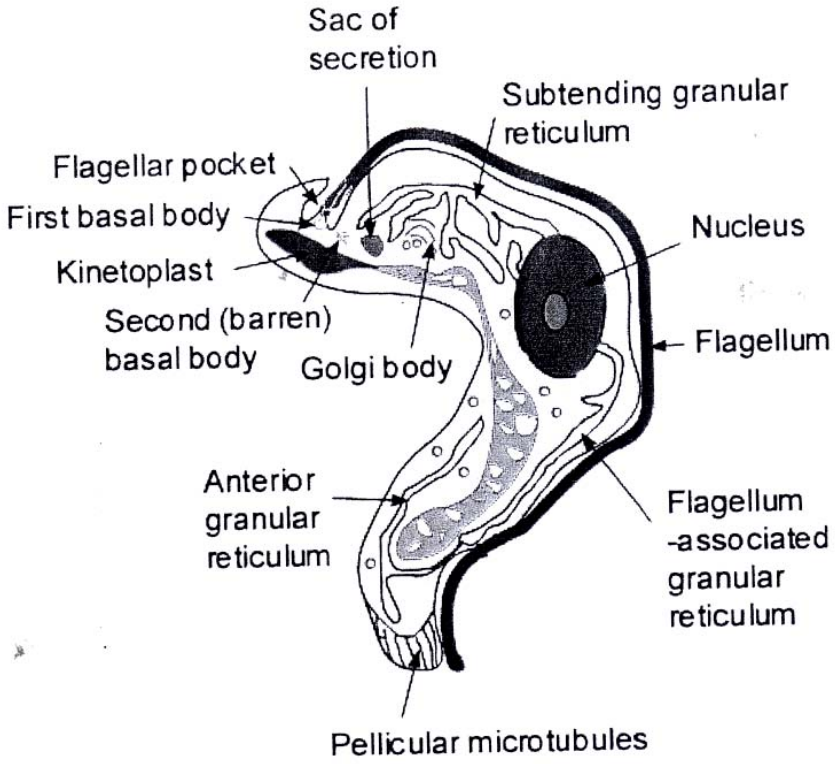

Figure 1.2. The general morphological features of *Trypanosoma congolense* in its bloodstream form stage (Vickerman, 1969)

Upon ingestion by the tsetse fly, the parasites enter the midgut that contains proteases (Sbicego *et al*., 1999; Roditi and Lehane, 2008). Concurrent with the transformation of bloodstream form parasites to procyclic forms, is a change in energy metabolism, crucial for the survival of the parasites in the vector (Aksoy, 2003). The procyclic forms transform into the epimastigote forms by removing their variable surface coat in the proboscis of the vector. The epimastigote forms migrate to the salivary glands where they become the infective metacyclic forms. In contrast, the entire life cycle of *T. vivax* occurs in the proboscis of the vector (Matthews *et al*., 2004).

1.5 Proteases

Proteases are a group of enzymes that possess the ability to catalyse the hydrolysis of proteins into peptides and free amino acids (North *et al*., 1990). Hydrolysis of protein substrates occurs in the active site that is usually located in a groove on the surface of the enzyme. Amino acid residues lining the active site groove dictate the specificity of the enzyme for a particular substrate (North *et al*., 1990). Amino acid residues N-terminal to the scissile bond of the

substrate are designated P_1, followed by P_2, P_3 etc., while P_1', P_2' etc. describe amino acid residues C-terminal of the scissile bond (Fig. 1.4). The corresponding sites on the protease accommodating the substrate are designated S_1,S_2,S_3 and S_1', S_2', S_3' , etc (Schechter and Berger, 1967).

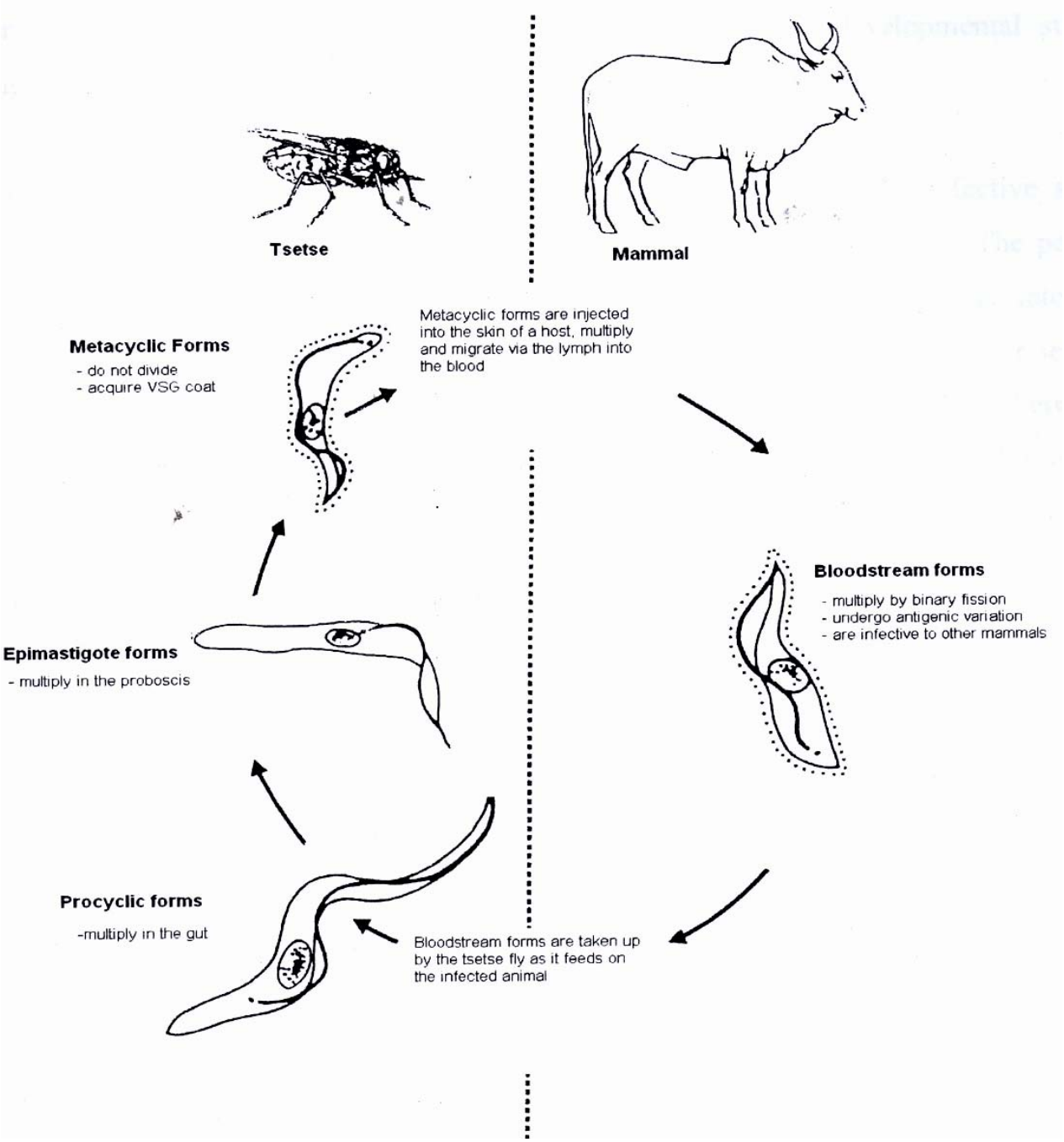

Figure 1.3. Life cycle of *T. congolense* showing the different morphological stages present in both the vector and the mammalian host (Authié, 1994).

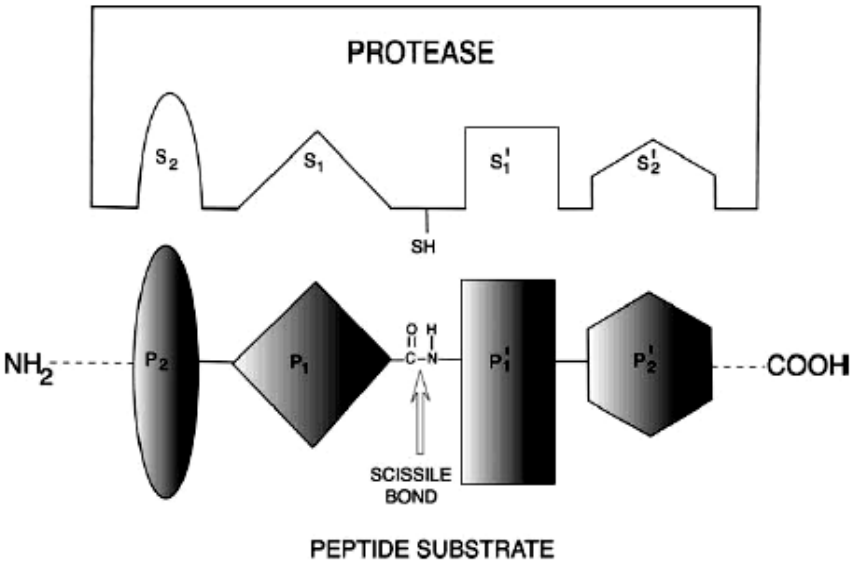

Figure 1.4. Schematic diagram showing the specificity of the protease for the substrate (Sajid and McKerrow, 2002).

Proteases have been grouped into five major classes based on the mechanism of catalysis; *viz.* cysteine, aspartic, serine, metallo and threonine proteases (Rawlings and Barrett, 1999). There are a few proteases for which the catalytic mechanism has not been elucidated. Proteases play different roles within cells or organisms. Some proteases are very specific for their substrate, thus performing highly specialised roles while others, such as the lysosomal cysteine proteases, are active against a wide range of protein substrates.

1.5.1 Cysteine proteases

Cysteine proteases form a group of enzymes that is widely distributed amongst organisms. The cysteine proteases are classified into clans, which are further divided into families based on sequence similarity, presence of inserted peptide loops and substrate specificity. Clan CA contains two families, i.e. family C1 that contains the papain superfamily of cysteine proteases and family C2 that contains calpain-like cysteine proteases (Sajid and McKerrow, 2002). Family C1 enzymes are the most abundant in living organisms (Turk *et al*., 2000).

Clan CA enzymes are characterised by a catalytic triad consisting of C^{25}, H^{159} and N^{175} (papain numbering). This group of enzymes is completely inhibited by the irreversible inhibitor L-*trans*-epoxysuccinyl-4-guanidinobutane (E-64). Substrate specificity of these enzymes is determined

by the S_2 pocket where a hydrophobic residue is preferred. Most enzymes belonging to this group are located in the lysosomal compartment such as mammalian cathepsins L and B or are secreted (Turk *et al*., 2000; Turk *et al*., 2001). Thus, lysosomal cysteine proteases are expected to be optimally active in the acidic pH range. However, some mammalian cathepsin L have also been shown to be highly active in the extracellular matrix where the pH is neutral (Dehrmann *et al*., 1995).

1.5.2 Parasite cysteine proteases

Cysteine protease activity has been demonstrated in many protozoan parasites including *Fasciola hepatica* (Prowse *et al*., 2002), *Entamoeba histolytica* (Quintas-Granados *et al*., 2009), *L. mexicana* (Sanderson *et al*., 2000), *T. cruzi* (Martinez *et al*., 1991), *T. congolense* (Mbawa *et al*., 1992), *Toxoplasma gondii* (Huang *et al*., 2009) and *Trichobilharzia regenti* (Dolečková *et al*., 2009). These cysteine proteases are closely related to papain and mammalian cathepsins L and B as indicated by their enzymatic characteristics and amino acid sequence identity (Sajid and McKerrow, 2002). Lysosomal cysteine proteases of protozoan parasites play an important role in intracellular protein degradation, invasion of host cells and stimulation of the host immune system (Klemba and Goldberg, 2002; McKerrow *et al*., 2006). Cysteine proteases of protozoan parasites are the major virulence factors implicated in the disease process (Mottram *et al*., 1996; Alexander *et al*., 1998; McKerrow *et al*., 1999; Mottram *et al*., 2004).

Expression of cysteine proteases by parasites is developmentally regulated. Mbawa *et al*. (1991a) showed that levels of expression of a major cysteine protease by *T. congolense*, *T. brucei* and *T. vivax* changes from one life cycle stage to another. Importantly, cysteine protease activity was observed to be highest in mammalian-infective stages of the parasites (metacyclics and bloodstream forms), thus, it was presumed that cysteine proteases may play a crucial role in host-parasite interactions (Mbawa *et al*., 1991a; Mbawa *et al*., 1992). For these reasons, parasite cysteine proteases have become targets for the development of new inhibitors with potential chemotherapeutic use (McKerrow *et al*., 1999).

Cysteine protease activity of *T. brucei*, *T. congolense* and *T. cruzi* is located in lysosomal organelles (Mbawa *et al*., 1991b; Cazzulo *et al*., 2001). Cysteine protease activity of *T. congolense* and *T. cruzi* has also been shown to be present in the flagellar pocket (Mbawa *et al*.,

1991b; Cazzulo *et al*., 2001). This extracellular location suggests a role for these proteases in the disease process. It is important to elucidate the biochemical characteristics of these enzymes in order to design and develop effective inhibitors or vaccines against the pathogenic effects of trypanosome infection.

1.5.3 Cysteine proteases of *T. congolense*

1.5.3.1 Congopain

A cysteine protease of *T. congolense* was purified from bloodstream form parasite lysates using thiopropyl-Sepharose and was named trypanopain-Tc (Mbawa *et al*., 1992). Trypanopain-Tc was shown to cleave the synthetic substrate Z-Phe-Arg-AMC. In a similar independent study, a 33 kDa antigen was isolated from *T. congolense* bloodstream form lysates using cystatin affinity chromatography and monoclonal antibody immunoaffinity chromatography (Authié *et al*., 1992). Analysis of enzymatic characteristics of the 33 kDa antigen revealed features similar to trypanopain-Tc. This led to the conclusion that trypanopain-Tc and the 33 kDa antigen are one and the same molecule, later named congopain (Authié, 1994).

Congopain (Fig. 1.5) consists of a pre-region, a pro-region, a catalytic domain that contains the active site residues and an unusual C-terminal extension that is linked to the catalytic domain via a polyproline hinge region (polythreonine in cruzipain) (Fish *et al*., 1995; Chagas *et al*., 1997). The pre-region or signal peptide of congopain is hydrophobic and targets the proenzyme to the endoplasmic reticulum. The pro-region of the protease inhibits activity of the parent enzyme and also plays a crucial role in the proper folding of the enzyme (Vernet *et al*., 1995; Lalmanach *et al*., 1998). The inhibitory effect of the proenzyme was narrowed down to a pentapeptide "YHNGA" that was shown to specifically inhibit cruzipain and congopain but not mammalian cathepsins L and B and could form the basis of the development of new inhibitory drugs (Lalmanach *et al*., 1998).

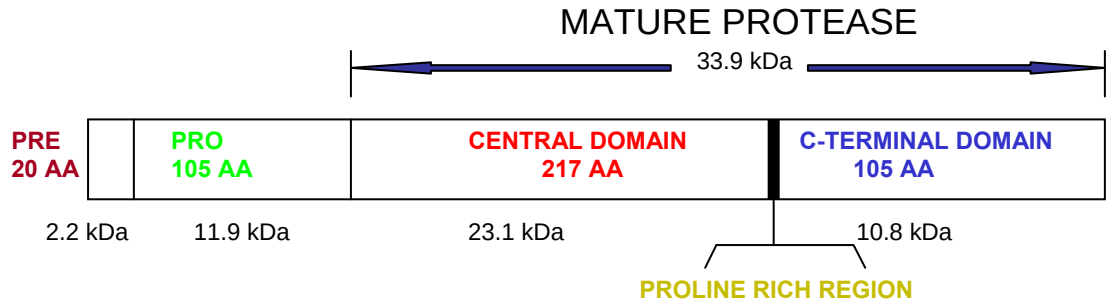

Figure 1.5. Schematic representation of the full-length congopain structure (Fish *et al*., 1995). The protease is composed of a signal peptide, a propeptide, a catalytic domain, polyproline hinge region, and a unique C-terminal domain.

The zymogen form of congopain is converted into a mature, active enzyme through autocatalytic cleavage of the proregion at acidic pH (Boulangé *et al*., 2001; Serveau *et al*., 2003). Processing of the proregion is greatly influenced by ionic strength and the presence of polyamines (Serveau *et al*., 2003). Furthermore, processing of procongopain was shown to occur at pH 8.0 in the presence of dextran sulfate, an unusual feature among papain-like enzymes (Serveau *et al*., 2003). The presence of procongopain in the plasma of infected cattle and the activity of the protease at neutral pH suggest extracellular processing of procongopain to the active form in the presence of blood vessel glycosaminoglycans (Serveau *et al*., 2003). This has serious implications for host-parasite relationships since congopain has been identified as a potential pathogenic factor in the disease process.

Congopain, like mammalian cysteine proteases, is activated by reducing agents such as dithiothreitol (DTT), L-cysteine and 2-mercaptoethanol (Mbawa *et al*., 1992). The protease is inhibited by E-64, cystatin, T-kininogens and peptide diazomethane (Mbawa *et al*., 1992; Chagas *et al*., 1997; Lalmanach *et al*., 2002). Congopain has specificity for bulky, hydrophobic residues such as phenylalanine or the artificial analogue cyclohexylalanine at the P_2 position (Lecaille *et al*., 2001). This specificity is conferred by the presence of leucine at position 205 at the bottom of the S_2 pocket rather than glutamate in cruzipain (Lecaille *et al*., 2001). Thus, congopain has a higher specificity for Z-Phe-Arg-AMC than Z-Arg-Arg-AMC (Mbawa *et al*., 1992; Chagas *et al*., 1997). Native congopain has higher affinity for Z-Phe-Arg-AMC than the catalytic domain (C2) that was recombinantly expressed in a baculovirus expression system (Table 1.1).

Table 1.1. Kinetic characterisation of congopain and C2 (Boulangé *et al*., 2001).

Enzyme	K_m (μM)	V_{max} pmol.s^{-1}.μg^{-1}	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu M^{-1}.s^{-1}$)
Congopain	4.8	229	21.8	4.5
C2	13.5	1.96	4.9	0.36

Analysis of the immune response of trypano-tolerant N'Dama cattle revealed that these animals possess an improved ability to mount a prominent IgG response against congopain when compared to trypano-susceptible Boran cattle (Authié *et al*., 1993b). This was further supported by data obtained from tests conducted using F1 crosses between N'Dama and Boran cattle (Taylor, 1998). The study revealed that a correlation exists between a degree of resistance to the disease and production of anti-congopain IgG by an infected animal (Taylor, 1998).

Congopain has been shown to be circulating in the bloodstream of *T. congolense*-infected animals (Serveau *et al*., 2003). Since the protease was shown to be processed at neutral pH, it was suggested that cysteine proteases play a crucial role in the disease process. An immunisation challenge experiment was conducted in cattle using two recombinant isoforms of congopain (C1 and C2) (Authié *et al*., 2001). The immunised cattle were challenged with *T. congolense* infection and the effects of immunisation on anaemia and immunosuppression were examined. Animals were partially protected from the disease even though the cattle developed acute anaemia during early stages of infection (Authié *et al*., 2001). Moreover, animals produced a prominent IgG response to native congopain and to VSG following challenge (Authié *et al*., 2001). Therefore, inhibition of congopain activity by antibodies induced by immunisation might contribute to resistance to the disease.

1.5.3.2 Dimerisation of congopain

The catalytic domain of congopain (C2) has been expressed in *Pichia pastoris* (Boulangé *et al*., submitted). Analysis of the recombinantly expressed protein by PhastGel® SDS-PAGE revealed dimerisation of congopain at physiological pH, an unusual feature amongst cathepsin L-like cysteine proteases. Dimerisation of congopain is a pH dependant phenomenon (Boulangé *et al*., submitted). Moreover, antibodies from trypano-tolerant cattle better recognised the dimeric form

of the protease as revealed by western blotting (Boulangé *et al*., submitted). It was hypothesised that protective epitopes are associated with the dimeric conformation of the protease.

A dimerisation model was proposed based on the 3-D structure of cruzipain, that is always a monomer (McGrath *et al*., 1995) and cathepsin W that is a dimer (Brinkworth *et al*., 2000). The model identified a stretch of charged and uncharged amino acid residues that form salt bridges depending on pH to be involved in the dimerisation mechanism (Prof J. Hoebeke, IBMC, Strasbourg, France, personal communication). In congopain, K^{39} and E^{44} were identified to be crucial for the dimerisation of the protease, while R^{39} and D^{43} are essential for dimerisation of cathepsin W (Fig. 1.6). However, cruzipain, the monomeric form of the model, contains F^{39} and P^{44} (Fig. 1.6).

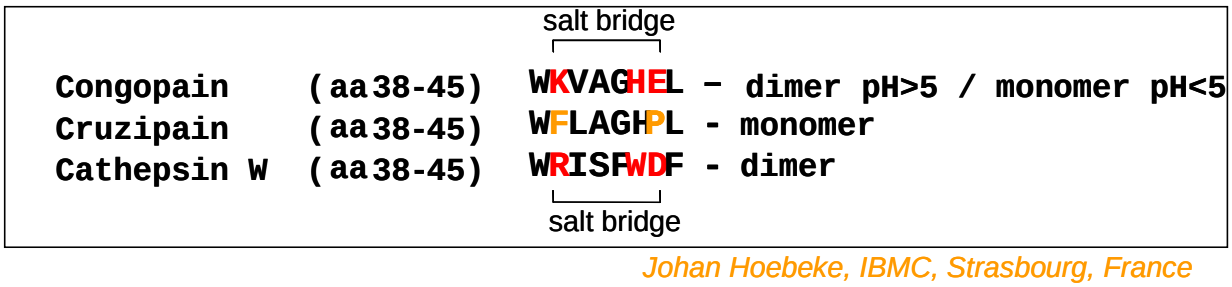

Figure 1.6. Schematic representation of the region thought to be responsible for the dimerisation of congopain. The congopain dimerisation model was proposed by Prof. J. Hoebeke (IBMC, Strasbourg, France, personal communication).

At acidic pH, K^{39} and H^{43} are positively charged, while E^{44} is negatively charged. Thus, a salt bridge is formed between H^{43} and E^{44} resulting in formation of an α-helix that is poorly immunogenic (Prof. J. Hoebeke, IBMC, Strasbourg, France, personal communication). Under these conditions, the protease exhibits a monomeric conformation. However, at neutral/physiological pH, H^{43} is deprotonated while K^{39} is positively charged and E^{44} is negatively charged. Therefore, these residues form a salt bridge resulting in a β-sheet formation (Prof. J. Hoebeke, IBMC, Strasbourg, France, personal communication). In this state, the protease displays a dimeric conformation. Table 1.2 describes in detail the side chain charges that result in monomer and dimer conformation in congopain.

Table 1.2. Amino acid side chain charges determining the proposed monomer or dimer conformation in congopain (Prof. J. Hoebeke, IBMC, Strasbourg, France, personal communication).

Protease	Residue	pKa side chain	Charge at pH 5	Monomer/ dimer	Charge at pH 7	Monomer/ dimer
Congopain	K	10	+		+	
	H	6.0	+	monomer	No charge	dimer
	E	3.8	-		-	
Cruzipain	F	-	None	monomer	None	monomer
	P	-	None		None	
Cathepsin W	R	12.5	+		+	
	W	-	None	dimer	None	dimer
	D	3.9	-		-	

1.5.3.3 Crystallisation of congopain.

The crystal structure of congopain has not yet been elucidated. However, the catalytic domain of cruzipain, the congopain homologue in *T. cruzi* has been crystallised and its structure determined (McGrath *et al*., 1995). Congopain shares 68% sequence identity with cruzipain in its catalytic domain (Lecaille *et al*., 2001; Serveau *et al*., 2003). Thus, a 3-D structural model of congopain (Fig. 1.7) was constructed based on the crystal structure of cruzipain (Lecaille *et al*., 2001). Similarly to the structure of papain-like cysteine proteases, the model revealed that congopain consists of two domains: an α-helical L domain and an R domain comprised mainly of β-sheets (Lecaille *et al*., 2001). The catalytic triad consisting of C^{25}, H^{159} and N^{175} (papain numbering) is buried in a cleft between the L and the R domains (Lecaille *et al*., 2001).

Crystallisation of full-length cruzipain that includes its C-terminal domain has not been achieved. This is mainly due to the difficulty associated with recombinantly expressing a full-length enzyme to the quantities required for crystallisation (Alvarez *et al*., 2002). Heterologous expression of functionally active full-length congopain in *P. pastoris* revealed intensive self-degradation of the protease in expression supernatants (Boulangé *et al*., submitted). However,

high level expression of truncated congopain devoid of the C-terminal domain was achieved in *P. pastoris* (Boulangé *et al*., submitted).

Figure 1.7. Structure of the congopain catalytic domain based on sequence identity with cruzipain (Lecaille *et al*., 2001). The X-Ray crystallographic structure of cruzipain is represented in black and that of congopain is represented in grey. The catalytic domain (C^{25}, H^{159} and N^{175}, papain numbering) is located in the cleft between the L and the R domain.

Immunisation of animals with recombinant congopain provided incomplete protection against a *T. congolense* infection (Authié *et al*., 2001). The inability of congopain to stimulate a protective immune response seems to be associated with its limited antigenicity. Protective epitopes recognised by sera from trypano-tolerant cattle are associated with the dimeric conformation of the protease (Boulangé *et al*., submitted). Therefore, it was important to study the effects of adjuvants in maintaining the appropriate conformation of the protease during immunisation and thereby stimulating high levels of antibodies that inhibit congopain activity.

1.6 Adjuvants

Subunit vaccines are increasingly used instead of live or attenuated vaccines (Wilson-Welder *et al*., 2008). Unlike live or attenuated vaccines that are potent stimulators of the immune system, subunit vaccines are poorly immunogenic and require frequent boosting (Singh and O'Hagan, 1999; Singh and O'Hagan, 2003). Administration of subunit vaccines without immune system-

enhancing molecules (adjuvants) gives rise to an inadequate immune response (Storni *et al*., 2005).

The major functions performed by adjuvants in immune stimulation include firstly a depot effect for the antigen, allowing for sustained release of an immunogen to the immune system, resulting in prolonged antigen presentation. Slow release of an antigen from the injection site allows for transport to the draining lymph nodes, where the antigen interacts with immune cells, resulting in antigen presentation to the T cells and B cells by MHC molecules (Storni *et al*., 2005). Oil-based adjuvants such as Freund's incomplete adjuvant and aluminium salts are efficient at maintaining an antigen at the injection site. However, the toxicity of Freund's adjuvant has prevented its application in human vaccine formulations (Singh and O'Hagan, 2003).

Secondly, an adjuvant must specifically direct the antigen to the cells of the immune system. This activates both the innate and the adaptive immune response and stimulates the production of co-stimulatory molecules by antigen presenting cells (Storni *et al*., 2005; Wilson-Welder *et al*., 2008). Cells of the innate immune response display pathogen-recognition receptors that interact with the pathogen-associated microbial patterns triggering an immune response (Fearon and Locksley, 1996). This feature is characteristic of adjuvants derived from microbial components such as pertussis toxin, lipopolysaccharide, CpG-rich motifs and lipid A (Schijns, 2000), as well as HSP70.

Heat shock protein (HSP) expression is up-regulated in stressed cells indicating their role as chaperonins (Schijns, 2000; Robert, 2003). Heat shock proteins generated during necrotic but not apoptotic death of a cell bind peptides and present them to antigen presenting cells (Basu and Srivastava, 2000). This demonstrated the capacity of HSPs to act as natural adjuvants of the immune system (Suto and Srivastava, 1995). Thus, HSPs from different organisms have been extensively studied to elucidate their role in immune stimulation against infectious agents, cancers and viruses (Suzue and Young, 1996; Rico *et al*., 1998; Harmala *et al*., 2002; Ge *et al*., 2006).

1.6.1 Heat shock proteins

Heat shock proteins are amongst the most conserved proteins in evolution (Robert, 2003; Gullo and Teoh, 2004). The capacity of the HSPs to protect the cell against various physical stressors (e.g. elevated temperatures, heavy metals, glucose deprivation, ischemia, cancer and microbial infections) is mediated by their ability to bind denatured proteins (Hartl, 1996; Bukau *et al*., 2006).

There is another group of heat shock proteins that is usually neglected by researchers, the constitutively expressed heat shock proteins that are expressed in unstressed normal cells (Pelham, 1986). These proteins play a pivotal role in "housekeeping", particularly in transmembrane transport of proteins, degradation of unstable and misfolded proteins, prevention and dissolution of protein complexes, uncoating of clathrin coated vesicles and folding and refolding of proteins (Li and Srivastava, 2004; Saibil, 2008; Jonak *et al*., 2009).

Heat shock proteins are classified into three major groups based on molecular mass, i.e. HSP90, HSP70 and HSP50-60, the later also called BiP (Gullo and Teoh, 2004). These groups of proteins can be further subdivided by localisation of the proteins within a cell (Table 1.3). The HSP70 group contains three members: *viz*. HSP70, HSC70 and glucose regulated protein 78 (GRP78), the later also called BiP (immunoglobulin heavy-chain binding protein). The HSP70 and HSC70 proteins are located within the cytosol while GRP78/BiP is located within the ER lumen (Gullo and Teoh, 2004).

HSP70 family proteins are potent stimulators of the immune system against infectious agents and cancers (Harmala *et al*., 2002). Immunisation of mice with ovalbumin genetically fused to *Mycobacterium tuberculosis* HSP70 demonstrated the adjuvant potential of mycobacterial HSP70. The immunised mice produced an ovalbumin specific CD8+ T cells response within three days post immunisation (Harmala *et al*., 2002). Rico *et al*. (1998) demonstrated the adjuvant potential of *L. infantum* HSP70 fused to *Escherichia coli* maltose binding protein. *L. infantum* HSP70 induced a strong humoral and cellular immune response against bacterial maltose binding protein. These independent studies unequivocally demonstrated the adjuvanticity of HSP70.

1.6.2 *T. congolense* immunoglobulin heavy-chain binding protein (BiP).

A 69 kDa antigen was identified in *T. congolense* lysates through probing a western blot with sera from infected trypano-tolerant N'Dama cattle and trypano-susceptible Boran cattle (Authié *et al*., 1992). The antigen was recognised by sera of both infected cattle species. However, the infected tolerant and susceptible cattle produced different anti-69 kDa antibody isotypes. Differences in the isotypes characterising the immune responses of animals differing in susceptibility to trypanosomosis triggered interest into researching the diagnostic potential of the antigen and determining its nature through molecular cloning. The cDNA of the 69 kDa antigen had been cloned by immunoscreening and sequenced. Sequencing revealed that the 69 kDa antigen is a member of the HSP70 family with most identity to mammalian heavy-chain binding protein (BiP). A similar protein has been shown to be present in many protozoan parasites including *T. brucei* (Bangs *et al*., 1993), *T. vivax* and *T. cruzi* (Boulangé and Authié, 1994) as well as *Toxoplasma gondii* (Kang *et al*., 2004). Trypanosomal BiP is expressed in all parasite life cycle stages (Boulangé and Authié, 1994).

Table 1.3. Classification of heat shock proteins into three major groups (Gullo and Teoh, 2004).

Family member	Location within cell	Function
Hsp110	Cytosolic	Binds full length proteins
GPR94	ER lumen	Substrate folding and stabilisation
GRP78	ER lumen	Molecular ratchet
Hsc70	Cytosolic, nuclear	Protein folding
Hsp70	Cytosolic, nuclear	Immunogenic for adaptive and innate responses
Hsp60	ER lumen	Adjuvant property
Hsp55	ER	Chaperone

A 2.35 kb cDNA contains the complete BiP open reading frame encoding 653 amino acid residues (Boulangé and Authié, 1994). The ORF of *T. congolense* BiP had been successfully amplified and cloned into a bacterial expression vector (Boulangé and Authié, 1994). The protein

(Fig. 1.8) is composed of a localisation signal sequence, a highly conserved ATPase domain (44 kDa), a hinge region with a protease sensitive site, a peptide binding domain (18 kDa) and a G/P rich C-terminal domain containing the trypanosome specific MDDL retention signal (Wang *et al*., 2002; Daugaard *et al*., 2007). The ATPase domain has two lobes with a deep cleft that binds ATP (Wang *et al*., 2002). The 18 kDa peptide binding domain consists of two subdomains composed of an antiparallel β-sheet sandwich that interacts with the substrate and an α-helical domain that interacts with misfolded proteins (Wang *et al*., 1993).

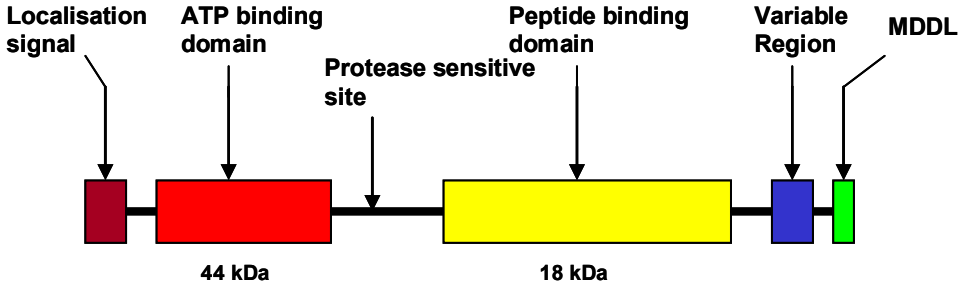

Figure 1.8. Heat shock protein 70 structure (Daugaard *et al*., 2007). The protein is composed of a localisation signal sequence, ATPase domain, peptide binding domain, variable region and the MDDL retention signal.

T. congolense BiP has a high degree of identity with other members of the HSP70 family except in the C-terminal domain (Boulangé and Authié, 1994). The C-terminal region contains epitopes unique to *T. congolense* BiP and could therefore potentially be used in species-specific diagnosis. The diagnostic potential of *T. congolense* BiP has been explored (Boulangé *et al*., 2002). Different regions of BiP were cloned and heterologously expressed in a bacterial expression system. Dot-blot assays were conducted on recombinant proteins to identify protein regions recognised by monoclonal antibodies, rabbit anti-BiP antibodies and sera from infected and uninfected N'Dama cattle (Boulangé *et al*., 2002). Antibodies predominantly recognised epitopes located in the C-terminal region of the proteins (Boulangé *et al*., 2002). A C-terminal antigen was used in an indirect ELISA to detect anti-BiP antibodies from sera of infected and uninfected cattle. These experiments showed that the C-terminal antigen-based ELISA has limited sensitivity for detecting primary infections but accurately detected secondary infections (Boulangé *et al*., 2002).

1.6.2.1 Adjuvant property of heat shock protein 70

The adjuvant potential of HSP70 of *M. tuberculosis*, *L. infantum* and *T. cruzi* have been widely documented (Suzue and Young, 1996; Rico *et al*., 1998; Planelles *et al*., 2002; Ge *et al*., 2006). Suzue and Young (1996) demonstrated the adjuvanticity of *M. tuberculosis* HSP70 fused with recombinant HIV p24 antigen. *M. tuberculosis* HSP70 elicited a strong humoral and cellular immune response against HIV p24 antigen. The capacity of HSP70 to stimulate the production of C-C chemokines by immune cells has been shown to be crucial for their adjuvanticity against a whole pathogen (Lehner *et al*., 2000). C-C chemokines play a role in attracting monocytes, B cells and T cells to the site of infection (Dieu *et al*., 1998). Moreover, chemokines signal for maturation of dendritic cells into professional APCs that migrate into the draining lymph nodes. In the lymph nodes, professional APCs interact with T cells and B cells resulting in a stronger cellular and humoral immune response (Lehner *et al*., 2000). Ge *et al*. (2006) conducted a study where mice were immunised with Japanese encephalitis virus (JEV) E-protein alone, mycobacterial HSP70 alone and E-protein fused to mycobacterial HSP70. Analysis of an immune response of mice revealed that mycobacterial HSP70 fused to E-protein elicited a stronger humoral and cellular immune response than E-protein or mycobacterial HSP70 alone. These results were corroborated by a study performed using foot and mouth disease virus (FMDV) antigens administered alone or fused with mycobacterial HSP70 (Su *et al*., 2007). Thus, covalent linkage of an antigen to the HSP70 molecule is a prerequisite for efficient stimulation of an immune response against an antigen.

Studies were done to identify the domain of HSP70 responsible for the adjuvant function. Ge *et al*. (2006) compared an immune response of mice immunised with JEV E-protein fused to full-length mycobacterial HSP70 and E-protein fused to the peptide binding domain of mycobacterial HSP70. The fusion protein consisting of E-protein and the peptide binding domain was more efficient at stimulating the immune response than E-protein fused to mycobacterial HSP70 (Ge *et al*., 2006). Furthermore, the E-protein-peptide binding domain fusion protein stimulated monocytes to produce IL-12, tumour necrosis factor-α (TNF-α) and enhanced the production of C-C chemokines (Ge *et al*., 2006). The ATPase domain of *M. tuberculosis* HSP70 plays a suppressive role in rats by inducing the production of IL-10 and TGF-α (Wendling *et al*., 2000; Wang *et al*., 2002). Therefore, it can be concluded that the peptide binding domain of HSP70 is responsible for the adjuvant property.

The efficiency of HSP70 to stimulate an appropriate immune response seems to be associated with their capacity to be rapidly internalised by APCs especially macrophages and DCs following binding of HSP-peptide complexes to their receptors (Castellino *et al*., 2000).

1.6.2.2 Receptor mediated endocytosis of HSP70-peptide complex

Mycobacterium tuberculosis HSP70 has been shown to specifically bind to CD40 on the surface of APC and stimulates the production of chemokines (Wang *et al*., 2002). C40 plays a crucial role in T-cell dependant B-cell activation, differentiation and immunoglobulin-class switching (van Kooten and Banchereau, 1997). Lazarevic *et al*. (2003) generated $CD40^{-/-}$ mice and demonstrated that stimulation of these mice with mycobacterial HSP70 failed to activate APC or activate $CD8^{+}$ and $CD4^{+}$ T-cells. It can be concluded that CD40 is the major receptor for microbial HSP-peptide complexes on the surfaces of APCs (Lazarevic *et al*., 2003).

Mammalian HSP70 has been shown to induce maturation of DCs through interaction with toll-like receptor-4 (TLR-4), CD91 and LOX-1 (Asea *et al*., 2000; Basu and Srivastava, 2000). There is no evidence showing interaction of mycobacterial HSP70 with TLR-4 and CD91 receptors (Harmala *et al*., 2002).

1.6.2.3 Mechanism of antigen presentation by HSP70

The mechanism through which HSP70 induces an immune response has not been unequivocally elucidated. However, most researchers have proposed that HSP70 presents antigens to the APCs via a process known as cross-presentation or cross-priming (Suto and Srivastava, 1995; Castellino *et al*., 2000). Presentation via the MHC class I pathway involves degradation of the proteins by the proteasome in the cytosol, followed by the transport of peptides by TAP to the ER where they bind to the MHC class I molecule (Germain, 1994; Pamer and Cresswell, 1998). Castellino *et al*. (2000) conducted a study to determine the pathway by which the peptides complexed to the HSP70 molecule are presented to the MHC class I molecule. The study showed that presentation of peptides linked to the HSP70 molecule by the MHC class I molecule is independent of the proteasome activity and TAP.

Initially, it was thought that B cells were unable to perform cross-presentation of HSP-peptide complexes (Suto and Srivastava, 1995). However, more recent work by Tobian *et al*. (2005) demonstrated the capacity of B cells to cross-present ovalbumin complexed to mycobacterial HSP70 to MHC class I molecules. This enhanced priming of $CD8^+$ T cells and improved antibody production by the B cells (Tobian *et al*., 2005). The capacity of HSP70 to induce a prominent antibody response against infectious agents makes it an attractive adjuvant for inducing an immune response against trypanosomes antigens.

1.7 Objectives of the present study

Congopain is a major cathepsin L-like cysteine protease of *T. congolense*. The protease was shown to be recognised by sera from infected cattle differing in their susceptibility to trypanosomosis. The trypano-tolerant N'Dama cattle were shown to mount a prominent IgG response to congopain during an infection with *T. congolense* compared to trypano-susceptible Boran cattle (Authié *et al*., 1993a). In addition, studies conducted using F1 crosses between N'Dama and Boran cattle revealed that a correlation exists between anti-congopain antibodies produced and the degree of susceptibility to trypanosomosis (Authié *et al*., 1993b). Hence, a hypothesis was proposed that efficient inhibition of cysteine proteases *in vivo* is an effector of trypanotolerance. In addition, it was shown that congopain exists as a dimer, and that protective epitopes are dimer-associated (Boulangé *et al*., submitted).

Immunisation trials were conducted in cattle using the baculovirus expressed catalytic domain of congopain (C2) in a proprietary, saponin-based adjuvant called RWL (Authié *et al*., 2001; Boulangé *et al*., 2001). The animals were challenged with *T. congolense* and the immune response was assessed and compared to that of control animals. The immunised animals maintained or gained weight during infection and showed less severe anaemia than the control animals during the chronic phase of the disease (Authié *et al*., 2001). The haematological and immunological features of immunised animals were similar to those shown by trypano-tolerant cattle. However, attempts to reproduce these results using yeast-expressed recombinant C2 with ISA206 oil-in-water adjuvant from Seppic, were unsuccessful (E. Authié and A. Boulangé, personal communication). The failure of the study was attributed to poor antigenicity of C2, the expression system used to produce the antigen and the type of adjuvant used (E. Authié and A. Boulangé, personal communication). Therefore, the main objective of the present study was to

improve the antigenicity of congopain using three approaches. The first approach involved validation of the proposed dimerisation model of congopain using PCR-based site directed mutagenesis. Understanding the dimerisation mechanism of congopain is crucial for the development of an effective anti-disease vaccine. The second objective was to determine the 3-D structure of the catalytic domain of congopain using X-ray crystallography in order to identify protective epitopes exposed on the surface of the protease. The last approach involved improving the delivery of C2 to the cells of the immune system while maintaining the antigen in an appropriate conformation to induce the production of activity inhibiting antibodies, by genetically fusing it to recombinant *T. congolense* BiP.

Recombinant C2 expressed in *P. pastoris* is a dimer at physiological pH. Moreover, sera from trypano-tolerant cattle detected epitopes associated with the dimeric conformation of the protease. Therefore, it is thought that protective epitopes are associated with the dimeric conformation. In order to develop an effective anti-disease vaccine based on congopain, it was crucial to understand the dimerisation mechanism of the protease. A dimerisation model was proposed based on the crystal structure of cruzipain, a monomer and the theoretical dimeric model for cathepsin W (Brinkworth *et al*., 2000; Prof. J. Hoebeke, IBMC, Strasbourg, France, personal communication). A stretch of charged and uncharged amino acid residues that form salt bridges depending on pH was identified to be involved in the dimerisation process. Amino acid residues located within the dimerisation motif of congopain were altered to mimick either cathepsin W [C2 (H43W)] or cruzipain [C2 (K39F, E44P)]. The generated C2 mutants were heterologously expressed in *P. pastoris* and their capacity to dimerise was assessed by PhastGel® SDS-PAGE. Furthermore, the effect of mutations on enzymatic characteristics of the mutants was investigated. The results obtained in this part of the study are presented in Chapter 2.

In the second approach, the aim was to determine the crystal structure of the catalytic domain of congopain using X-ray crystallography in order to elucidate major epitopes located on the surface of the protease and assist in the understanding of the dimerisation mechanism of congopain. The information acquired from the crystallisation of C2 was going to be utilised to design and develop mimotopes with potential use in vaccination studies. High level expression of C2, mutant proenzyme [proC2 (C25A)] and C2 mutants [C2 (H43W) and C2 (K39F, E44P)] was achieved in yeast. The proteins were purified to crystallisation standards. Crystallisation of these

proteins is currently underway. In the absence of crystallographic data, we embarked on *in silico* homology modelling of the 3-D structure of congopain. The results obtained in this part of the study are presented in Chapter 3.

The aim of the final part of the study was to assess the capability of *T. congolense* BiP to stimulate the production of congopain activity inhibiting antibodies while maintaining the antigen in an appropriate conformation. To this end, constructs were designed that had the catalytic domain of congopain fused to BiP. The constructs were designed in a way to allow for their cloning into pPic9 and pMal expression vectors for their expression in both *P. pastoris* and *E. coli*. The proregion of congopain was included in most constructs because of its importance in the proper folding of the protease (Lalmanach *et al*., 1998; Sanderson *et al*., 2000). On the other hand, the 10 kDa C-terminal domain of BiP was omitted in most constructs because of its high variability and antigenicity (Boulangé and Authié, 1994). The BiP-C2 chimeras were successfully expressed in both eukaryotic and bacterial expression systems. Due to problems associated with poor expression levels of the chimeras in yeast and difficulty with the purification of bacterial expressed chimeras, the chimeras were not tested in laboratory rodents. The results obtained in this part of the study are presented in Chapter 4. All the results are discussed in the final chapter.

CHAPTER 2

INVESTIGATION OF THE DIMERISATION MECHANISM OF CONGOPAIN USING PCR-BASED SITE-DIRECTED MUTAGENESIS

H. H. Ndlovu[1], A. Boulangé[1,2] and T. H. T. Coetzer[1]

[1]Department of Biochemistry, University of KwaZulu-Natal (Pietermaritzburg Campus), Private Bag X01, Scottsville, 3209, South Africa
[2]UMR 017 IRD-Cirad Trypanosomes, Campus international de Baillarguet, 34398 Montpellier Cedex 5, France

ABSTRACT

Congopain is the major cathepsin L-like cysteine protease of *Trypanosoma congolense*. Congopain has been implicated in the disease process, thus has vaccine potential and is a possible target for chemotherapy. The catalytic domain of congopain (C2) has been recombinantly expressed in *P. pastoris* and shown to dimerise at physiological pH, an unusual feature amongst cathepsin L-like cysteine proteases. Of direct relevance to vaccine design, antibodies from trypano-tolerant cattle better recognised the dimer conformation, advocating for protective epitopes to be dimer-associated. A dimerisation model was proposed based on the 3D-structure of cruzipain (always a monomer) and cathepsin W (always a dimer). The model relies on a stretch of charged and uncharged amino acid residues that form salt bridges depending on the pH. To validate the model, K^{39} and E^{44} were substituted with F^{39} and P^{44}, to mimick cruzipain while H^{43} was substituted with W^{43} to mimick cathepsin W using PCR-based site-directed mutagenesis in the presence of a thermostable DNA ligase. The mutated congopain variants were functionally expressed in *P. pastoris* and purified. Analysis of the purified mutants [C2 (H43W) and C2 (K39F; E44P)] by PhastGel® SDS-PAGE revealed unexpected electrophoretic mobility compared to wild-type C2. The non-boiled and non-reduced C2 (H43W) migrated at 60 kDa while C2 (K39F; E44P) migrated at 50 kDa and wild-type C2 displayed a mobility shift at 42 kDa. In denaturing conditions or acidic pH, the C2 mutants and wild-type C2 migrated as monomers at 27 kDa. C2 mutants and wild-type C2 displayed similar pH profiles, but different kinetic constants for the hydrolysis of Z-Phe-Arg-AMC. It can be concluded that these amino acid residues indeed play a role in the dimerisation mechanism of congopain; however, these are not the only participating residues.

Keywords: congopain; dimerisation; site-directed mutagenesis; expression; kinetic constants.

2.1 INTRODUCTION

African animal trypanosomosis or nagana is a major constraint to livestock production in sub-Saharan Africa. The disease is estimated to cost $ 1.3 to 5 billion per year in Africa alone (Kristjanson *et al*., 1999). The disease is caused by tsetse transmitted protozoan parasites including *Trypanosoma congolense, T. vivax* and to a lesser extent *T. b. brucei* (Stevens and Brisse, 2004). Generally, the disease manifests itself by weight loss, severe anaemia, immunosuppression, and is often fatal if untreated.

Currently, control of trypanosomosis relies on three principal strategies: vector eradication/control, trypanocidal drugs (chemotherapy) and trypano-tolerant cattle (McDermott and Coleman, 2001). Complete eradication of the vector is improbable; hence, control of the disease at herd or individual level relies on trypanocidal drugs (Holmes *et al*., 2004). The three widely used trypanocidal compounds are isometamidium (Samorin®), diminazene aceturate (Berenil®) and homidium bromide (Ethidium®). Prolonged exposure of the parasites to these drugs has culminated in the emergence of drug resistance (Geerts *et al*., 2001). Hence, there is an urgent need to develop new, effective and safe trypanocidal drugs. Although potential new drugs have been identified (McKerrow *et al*., 1995; McKerrow *et al*., 2006), they are far from commercialisation.

Vaccination offers an attractive alternative strategy for the treatment of trypanosomosis. However, it is significantly limited by the exposed, variable surface glycoprotein (VSG) coat of the parasite. VSG undergoes rapid antigenic variation during an infection cycle (Donelson, 2003), hence, cannot be regarded as a vaccine target. This obstacle facilitated a paradigm shift towards targeting accessible parasite invariant antigens. Immunisation trials with flagellar pocket antigen (Mkunza *et al*., 1995) and tubulin (Lubega *et al*., 2002) showed that these antigens confer partial protection against the parasite. However, the results are still subject to controversy within the scientific community.

Trypano-tolerant cattle, such as the West African N'Dama, Muturu and Dahomey breeds, possess an ability to resist the disease caused by an infection with trypanosomes (Authié, 1994; Murray *et al*., 2004). These animals are able to limit parasitemia and control the pathogenic effects of infection. Western blotting of parasite lysates with sera from trypano-tolerant and trypano-

susceptible cattle revealed that trypano-tolerant cattle specifically recognised a 33 kDa immunodominant antigen (Authié *et al*., 1992). In an independent study, a 33 kDa antigen was purified from *T. congolense* bloodstream form lysates using thiopropyl-Sepharose (Mbawa *et al*., 1992). After determination of the enzymatic characteristics of the two antigens, it was concluded that the two proteases are actually one and the same protease, later named congopain (Authié, 1994). Congopain belongs to the papain superfamily of cysteine proteases and exhibits the highest identity to mammalian cathepsin L (Authié *et al*., 1992).

Congopain is present in the bloodstream of infected animals where it is thought to contribute to trypanosome-induced disease pathology (Authié, 1994). To test this hypothesis, immunisation/challenge experiments were conducted in cattle, using baculovirus-expressed C2, the catalytic domain of congopain (Authié *et al*., 2001; Boulangé *et al*., 2001). It was shown that antibodies directed against congopain control trypanosome-induced pathology. Animals immunised with C2 developed less severe anaemia and gained weight after a few months of being challenged with *T. congolense*. Although the effects were limited; this initial experiment showed that a vaccine strategy based on targeting the disease rather than the parasite, that is called an "anti-disease vaccine" approach, seems to be a viable and promising avenue for preventing trypanosomosis in cattle (Schofield, 2007).

Congopain is composed of a signal peptide for secretion, a propeptide that governs proper folding, a central domain that bears the catalytic site, and an unusual C-terminal extension, present only in the cathepsin L-like cysteine proteases of *Trypanosomatidae*, the function of which remains to be elucidated. The catalytic domain of congopain (C2) was expressed in *Pichia pastoris* (Boulangé *et al*., submitted), an expression system known to produce higher yields than the baculovirus expression system. Western blot analysis of congopain revealed that the protease exists in two conformations, depending on pH. The protease is monomeric at acidic pH and dimeric at neutral pH (Boulangé *et al*., submitted). Antibodies from trypano-tolerant cattle better recognised the dimeric form of the protease (Boulangé *et al*., submitted). It seems that protection against the disease is associated with dimer-specific epitopes, thus it is crucial to understand the dimerisation mechanism of congopain.

Based on the 3D structures of cruzipain (Eakin *et al*., 1993; McGrath *et al*., 1995), the congopain homologue from *T. cruzi* that is known to exist as a monomer, and of cathepsin W (Brinkworth *et al*., 2000), that exists as a dimer, a dimerisation model was proposed (Prof. J. Hoebeke, IBMC, Strasbourg, France, personal communication). The model is based on a stretch of charged and uncharged amino acid residues that can form salt bridges depending on the pH. In congopain, K^{39} and E^{44} were suggested to be the amino acid residues essential for dimerisation (Fig. 2.1).

At acidic pH, the side chains of K^{39} and H^{43} are positively charged, whereas, E^{44} is negatively charged. Thus, the salt bridge could be formed between H^{43} and E^{44} resulting in the formation of an α-helix. Therefore, the protease exhibits the monomer conformation that is poorly immunogenic. At physiological pH, H^{43} is deprotonated allowing for salt bridge formation between K^{39} and E^{44} resulting in β-turn formation that constitutes the dimeric conformation of the protease. In cruzipain, K^{39} and E^{44} are substituted with F^{39} and P^{44}, respectively. This prevents salt bridge formation, thus, the protease exists as a monomer regardless of pH. In cathepsin W, H^{43} is substituted with W^{43} preventing competition between R^{39} and D^{44} for the formation of the salt bridge at acidic pH. Thus, the protease exists as a dimer regardless of pH. The interaction of R^{39} and D^{44} is equivalent to the interaction between K^{39} and E^{44} in congopain. Thus, substitution of H^{43} by W^{43} in congopain is expected to produce the mutant mimicking cathepsin W. Therefore, the mutant congopain was expected to always exist as a dimer regardless of pH. The mutant congopain containing the substitution of K^{39} and E^{44} with F^{39} and P^{44}, respectively was expected to behave like cruzipain and be monomeric regardless of pH.

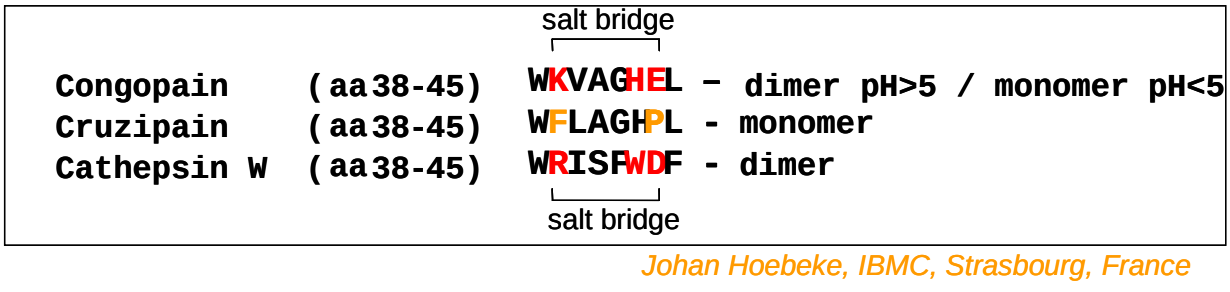

Figure 2.1. Schematic representation of the region thought to be responsible for the dimerisation of congopain. The congopain dimerisation model was proposed by Prof. J. Hoebeke (IBMC, Strasbourg, France, personal communication). The model is based on the capacity of certain amino acid residues to form salt bridges depending on pH. In congopain, K^{39}, H^{43} and E^{44} were suggested to be crucial for dimerisation.

The present study attempts to validate the proposed dimerisation model of congopain by replacing K^{39} and E^{44} with F^{39} and P^{44} using PCR-based site-directed mutagenesis to resemble

cruzipain. To resemble cathepsin W, H^{43} in congopain was replaced with W^{43}. Mutagenesis was performed through incorporation of a 5' phosphorylated mutagenesis primer by the flanking central domain primers in the presence of a thermostable DNA ligase during the standard PCR reaction (Michael, 1994). The mutated open reading frame (ORF) was expressed in *P. pastoris* and the capacity of the mutant proteins to dimerise assessed by PhastGel® SDS-PAGE. In addition, the enzymatic characteristics of the purified mutants were investigated.

2.2. MATERIALS AND METHODS

2.2.1. Materials

Molecular Biology: EcoRI, NotI, SacI, shrimp alkaline phosphatase (SAP), T4 DNA ligase, dNTP mix, X-gal, MassRuler™ DNA Ladder Mix, O'GeneRuler ™ DNA Ladder Mix 1kb, GeneJet™ Plasmid Miniprep kit and TransformAid™ Bacterial transformation kit were purchased from Fermentas (Vilnius, Lithuania). *Taq* DNA polymerase, 10× PCR buffer B and 25 mM magnesium chloride were purchased from Solis Biodyne (Tartu, Estonia). The pGEM-T® vector cloning kit was purchased from Promega (Madison, WI, USA) and the DNA Clean and concentrator kit was purchased from ZymoResearch (Orange, CA, USA). The E. Z. N. A® Gel Extraction kit was purchased from peQLab (Erlangen, Germany). Competent *Escherichia coli* cells (NEB 5-α and JM 109) were purchased from New England Biolab (Ipswitch, MA, USA). The *P. pastoris* yeast strain (GS 115) and the vector pPic9 were purchased from Invitrogen (Carlsband, CA, USA). The BioRad GenePulser™ electroporator was from BioRad (Hercules, CA, USA).

PhastGel® electrophoresis: PhastGel® homogenous 12.5% gels, PhastGel™ SDS-Buffer Strips, PhastGel™ Blue R tablets were purchased from Amersham Biosciences (Uppsala, Sweden). The PhastSystem Separation and Control unit was purchased from Amersham Biosciences (Uppsala, Sweden)

Enzyme Kinetics: The fluorogenic substrate benzoyloxycarbonyl-phenylalanine-arginine-7-amino-4-methylcoumarin (Z-Phe-Arg-AMC) and irreversible inhibitor *trans*-epoxysuccinyl-L-leucyl-amido (4-guanidino) butane (E-64) were purchased from Sigma. The FLUOStar Optima spectrophotometer was purchased from BMG Labtech (Offenburg, Germany).

2.2.2 Methods

2.2.2.1 Isolation of plasmid DNA

A glycerol stock of *E. coli* containing pPic9-proC2 (Fig. 2.2) was streaked on 2× YT plates [1.6% (w/v) tryptone, 1% (w/v) yeast extract and 0.5% (w/v) NaCl] containing ampicillin (50 µg/ml). The plates were incubated overnight at 37ºC. Liquid 2× YT (10 ml, 50 µg/ml ampicillin) was inoculated with a single colony and the culture was grown at 37ºC overnight. Plasmid DNA was isolated using the GeneJet® Miniprep kit (Fermentas) according to manufacturer's instructions. Isolated plasmid DNA was subjected to electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer (40 mM Tris-Cl buffer, 0.1 mM EDTA, 20 mM acetic acid) run at 80 V for 30 min.

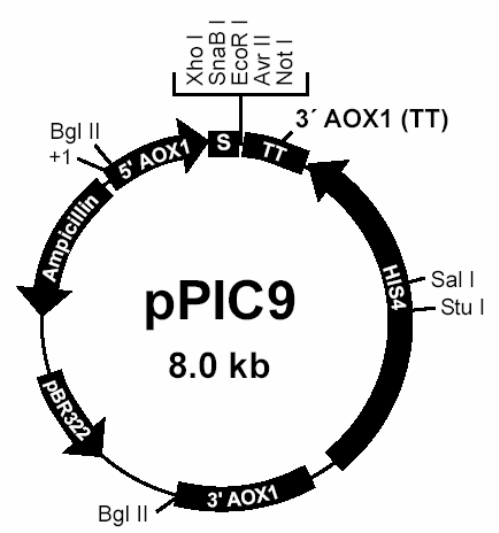

Figure 2.2. pPic9 expression vector map (Invitrogen). The vector contains the following features for efficient expression: ColE1 origin of replication (pBR322), 5' AOX fragment (5'AOX), multiple cloning sites, α-factor secretion signal (S), *c-myc* epitope (TT), ampicillin resistance gene, and the his4 gene for the selection by complementation in the his4 yeast strain.

2.2.2.2. PCR-based-site-directed mutagenesis

Site-directed mutagenesis was performed following the protocol described by Michael (1994). The isolated plasmid DNA (pPic9-proC2) was used as the template for site-directed mutagenesis. Mutations were introduced by a mutagenesis primer (H43W or K39F; E44P) incorporated during a standard PCR reaction using two flanking primers (Table 2.1). The mutagenesis primers were 5'-phosphorylated (Proligo, France) to allow for efficient ligation of the two newly synthesised single stranded fragments into a full length product by the thermostable DNA ligase (Fig. 2.3A).

The flanking primers (CD primers) were used in combination with the mutagenesis primer (H43W or K39F; E44P) at a ratio of 1:10 (Fig. 2.3B).

Table 2.1. Sequences of primers used for the mutagenesis of C2 open reading frame (ORF). The central domain (CD) primers were used as flanking primers to amplify the catalytic domain of congopain (C2). The mutagenesis primer (H43W or K39F; E44P) was used to introduce the desired mutations.

Primer	Sequence	Restriction site
CDFw	5'CC**GAATTC**TGCTTTCCCGTGGGCCGTTGC3'	EcoRI
CDRv	5'T**GCGGCCGC**CGTGCCGCACGAGCCGGGCGG3'	NotI
H43W	5'CAGTGGTTCG**TTG**CAGGCATCCGCTGACG3'	None
K43F; E44P	5'AAGGTTGCAGGC**TGGGAC**GCCTGACG3'	None

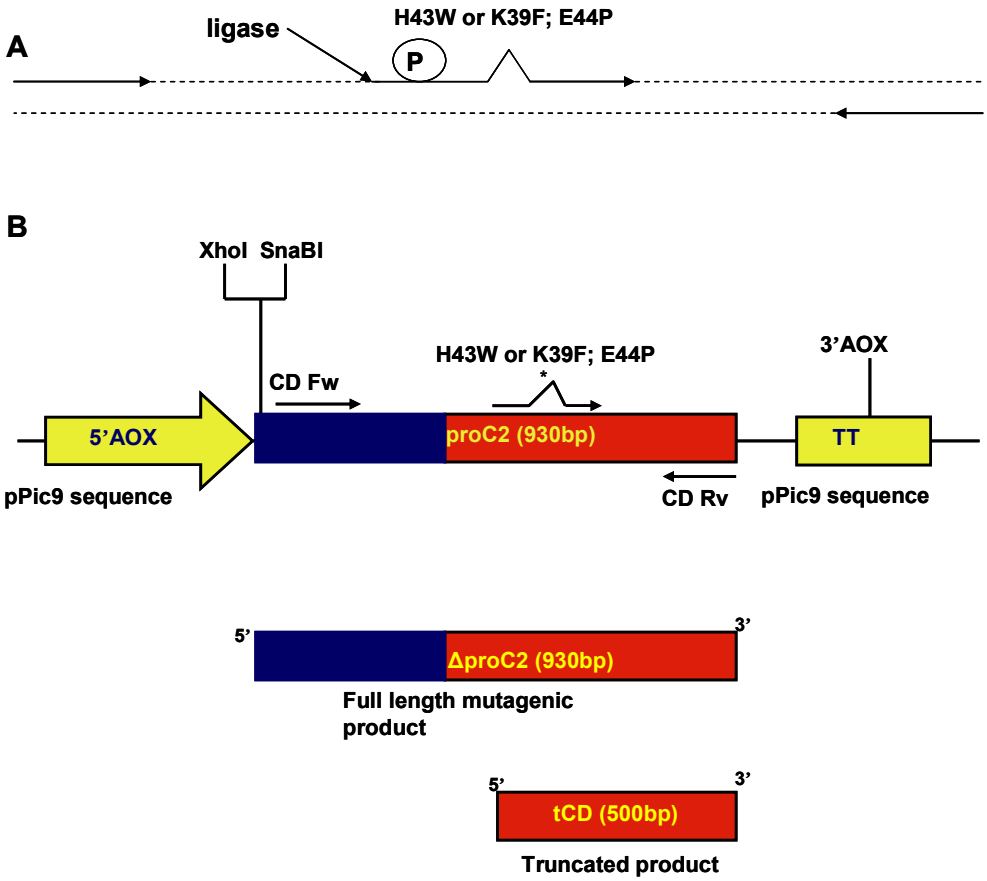

Figure 2.3. Schematic diagram showing the mutagenesis reaction and the mutagenic products expected from the mutagenesis reaction. Panel A, schematic diagram showing the mutagenesis reaction. Panel B, products expected from the mutagenesis reaction. The flanking central domain primers (CDFw and CDRv) were used to incorporate the mutagenesis primer (H43W or K39F; E44P) during a standard PCR reaction. Two expected products, the full-length product containing the mutation of interest, and a truncated product resulting from the amplification from the mutagenesis primer to the downstream primer

The PCR reaction was carried out in a 100 µl reaction volume, containing 0.5 ng plasmid DNA (pPic9-proC2), 1 µM flanking primers (CD), 10 µM mutagenesis primer (H43W or K39F; E44P), 200 µM dNTP mix, 2.5 mM Mg^{2+}, 1 mM DTT, 1 mM NAD^{+}, 2.5 U *Taq* DNA ligase (NEB), and 2.5 U FirePol® DNA polymerase (Solis Biodyne) using the 1× *Taq* polymerase buffer B [80 mM Tris-Cl buffer, pH 9.4, 20 mM $(NH_4)_2SO_4$, 0.02% (v/v) Tween 20]. The initial denaturation step (95ºC for 5 min) was followed by 25 cycles composed of denaturation at 95ºC for 1 min, primer annealing at 45ºC for 1 min, and elongation at 65ºC for 3 min. Finally, elongation at 65ºC for 6 min was performed to allow for ligation of the fragment primed with mutagenesis primer into the full length product.

2.2.2.3. Cloning of the full-length mutated product

Cloning into pGEM-T vector

The full-length PCR product (930 bp) was purified from the smaller fragment and other contaminants using the peQLab Gel extraction kit. Pure mutated full-length product was ligated into the pGEM-T vector (Fig. 2.4) in a ratio of 3:1 using T4 DNA ligase in a 10 µl reaction volume. The ligation mix (2 µl) was used to transform NEB 5-α competent *E. coli* cells (NEB) according to the manufacturer's instructions. The transformed cells (100 µl) were plated onto 2× YT containing ampicillin (50 µg/ml), X-gal (20 µg/ml) and IPTG (10 µg/ml) to allow for blue and white selection of recombinants. The plates were incubated overnight at 37ºC

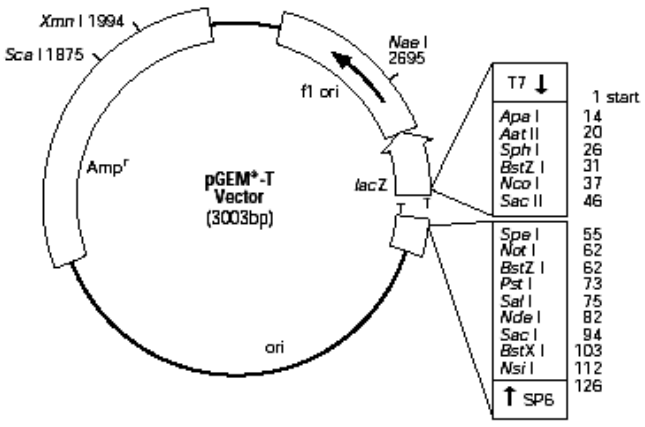

Figure 2.4. pGEM-T® vector map (Promega). The vector contains a multiple cloning site, ampicillin resistance gene for selection of recombinants, and the lacZ gene coding region for blue and white colony screening.

White colonies were selected from a plate, and colony PCR was performed using the T7 and the SP6 primers (Table 2.2). The template was denatured for 1 min at 95ºC, followed by 25 cycles consisting of denaturation at 95ºC for 30 s, primers annealing at 45ºC for 30 s, elongation at 72ºC for 1 min, and finally elongation at 72ºC for 7 min. Recombinants were grown in selective media overnight at 37ºC. Plasmid DNA was isolated using the GeneJet® Miniprep kit (Fermentas).

Cloning into pPic9 expression vector

Recombinant pGEM-T-proC2 (H43W), pGEM-T-proC2 (K39F; E44P) and pPic9 were restricted with EcoRI and NotI in 1× Orange buffer™ (0.05 M Tris-Cl buffer, pH 7.5, 0.01 M $MgCl_2$, 0.1 M NaCl, 0.1 mg/ml BSA). The dropped out insert was purified using the E. Z. N. A® Gel Extraction kit (peQLab). The vector was dephosphorylated with 1 U of SAP and concentrated using the ZymoResearch™ DNA Clean and Concentrator kit. The insert was ligated into linear pPic9 vector as described in Section 2.2.2.3. Competent *E. coli* cells (NEB 5-α) were transformed and recombinant colonies screened by colony PCR using AOX primers (Table 2.2). Positive colonies were grown overnight in selective media and plasmid DNA isolated.

Table 2.2. Sequences of primers used for screening of recombinants in pGEM-T vector and pPic9. The T7 and SP6 primers were used to screen for pGEM-T vector recombinants. The AOX primers were used to screen for pPic9 recombinants.

Primer	Sequence	Tm (ºC)
T7	5'TAATACTGACTCATATAGGG3'	46
SP6	5'CATATGATTTAGGTGACACTATA3'	48
AOX Fw	5'ACTGGTTCCAATTGACAAGC3'	50
AOX Rv	5'GCAAATGGCATTCTGACATCC3'	52

2.2.2.4. Transformation of *Pichia pastoris* (GS 115 strain)

Pichia pastoris strain GS 115 cells were treated as described by Wu and Letchworth (2004) to improve the efficacy of transformation. Briefly, *P. pastoris* cells were grown in 500 ml YPD [1% (w/v) yeast extract, 2% (w/v) peptone, and 2% (w/v) dextrose] until they reached an OD_{600} between 1 and 2. The cells were pelleted (2000×g, 20 min, 4ºC) and resuspended in 400 ml of 100 mM lithium acetate, 10 mM DTT, 0.6 M sorbitol, and 10 mM Tris-Cl buffer, pH 7.5, and incubated for 30 min at room temperature (RT). The cells were subsequently washed with 75 ml

1 M sorbitol, and finally resuspended in 1.5 ml of 1 M sorbitol. The final concentration of the cells was 10^{10} cells/ml. Recombinant pPic9-proC2 (H43W) and pPic9-proC2 (K39F; E44P) were linearised with SacI, to improve the efficiency of integration of the insert into the *P. pastoris* genome by homologous recombination. Linear plasmid DNA (200 ng) was used to transform competent GS 115 yeast cells (200 µl) by electroporation (1.5 kV, 25 µF, 186 Ω) using a BioRad Gene Pulser™ (BioRad) in pre-chilled 2 mm electroporation cuvettes. The cells were immediately resuspended in chilled 1 M sorbitol (1 ml). The cells were concentrated by centrifugation (2000×g, 10 s, RT) and plated on MD plates [1.34% (w/v) yeast nitrogen base without amino acids, 0.0004% (w/v) biotin, 2% (w/v) glucose, 15 g/l bacteriological agar] containing ampicillin (50 µg/ml). The plates were incubated at 30ºC until the colonies were visible. Colony PCR using the AOX primers (Table 2.2) was performed on the *P. pastoris* genome to confirm integration of foreign DNA (Ayra-Pardo *et al*., 1998). Fresh single colonies (not more than a week old) were screened using a protocol adapted from Ayra-Pardo *et al*. (1998). The protocol was adapted by changing the quantity of primers used in the reaction. The AOX primers were used at a 10× higher concentration as compared to the original protocol. Positive colonies were grown in selective media (Section 2.2.2.5) for protein expression.

2.2.2.5 Expression of C2 mutants in *Pichia pastoris*

Liquid YPD (100 ml) containing ampicillin (50 µg/ml) was inoculated with a single recombinant colony and grown at 30ºC in an orbital shaking incubator for two to three days. Ampicillin was added to the media to prevent contamination of the culture. These cells were transferred into 1 L BMGY [1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer, pH 6.5, and 1.34% (w/v) YNB]. The culture was grown at 30ºC for a further two days to increase biomass. The cells were pelleted (6000×g, 10 min, RT) and resuspended in 500 ml buffered minimal medium [BMM, 100 mM potassium phosphate buffer, pH 6.5, 1.34% (w/v) YNB, 0.0004% (w/v) biotin, and 5% (v/v) methanol]. The culture was grown in a baffled flask at 30ºC in a shaking incubator. Methanol (0.5%) was added daily during the four to six days expression period. The daily addition of methanol to the culture media induces continuous expression of the protein of interest by *P. pastoris* that secretes the protein of interest into the media. Cells were pelleted (6000×g, 10 min, RT) and the supernatant kept at -20ºC until required for purification.

2.2.2.6 Purification of the C2 mutants

Three phase partitioning (TPP)

The expression supernatants were thawed and filtered through Whatman No 1 filter paper. The recombinantly expressed C2 mutants were purified by three phase partitioning (TPP) (Pike and Dennison, 1989). The protease was autocatalytically processed into the mature form by lowering the pH to 4.2 with phosphoric acid. Tertiary butanol (1/3 final volume, 250 ml) was added to 500 ml supernatant. The mutants were precipitated with the kosmotrope ammonium sulfate (40% of total volume, 300 g). The salt was dissolved by gentle stirring on a magnetic stirrer until completely dissolved. The precipitated protein mixture was centrifuged (6000×g, 10 min, 4ºC) using a pre-chilled swing-out rotor. The protein layer was collected at the interface between the tertiary butanol and aqueous phases and the precipitate was dissolved in dialysing buffer (PBS, pH 7.2). The precipitated protein was dialysed (PBS, pH 7.2) using 10 kDa cut-off dialysis membrane (Pierce, Rockford, IL, USA) overnight at 4ºC to remove traces of tertiary butanol and salts. Solid polyethylene glycol M_r 20 000 was used to concentrate the dialysed proteins in the dialysis bag to approximately 5 ml.

Molecular Exclusion Chromatography (MEC)

A 25 × 840 mm column was packed with Sephacryl S-300 HR resin (Sigma) and equilibrated with one column volume of 50 mM sodium acetate buffer, pH 4.2 (25 $cm.h^{-1}$, 4ºC). The column was calibrated with a calibration mixture (5 ml) comprising of blue dextran (2000 kDa), phosphorylase b (97 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (34.3 kDa), and soybean trypsin inhibitor (25 kDa). The availability constant (K_{av}) was determined for each protein, where the elution volume (V_e) of blue dextran denotes the void volume (V_o).

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

Where, K_{av} is the availability constant,

V_e is the elution volume,

V_o is the void volume,

V_t is the total column volume (Dennison, 1999).

The Fischer's plot, K_{av} vs $logM_r$ (Fig. 2.5) was used as a standard curve to estimate M_r of eluted protein (Dennison, 1999). The TPP concentrated protein sample (2.5% of column volume, 8 ml) was applied to the column and eluted with 50 mM sodium acetate buffer, pH 4.2. Elution of the proteins was monitored by measuring absorbance at 280 nm (A_{280nm}).

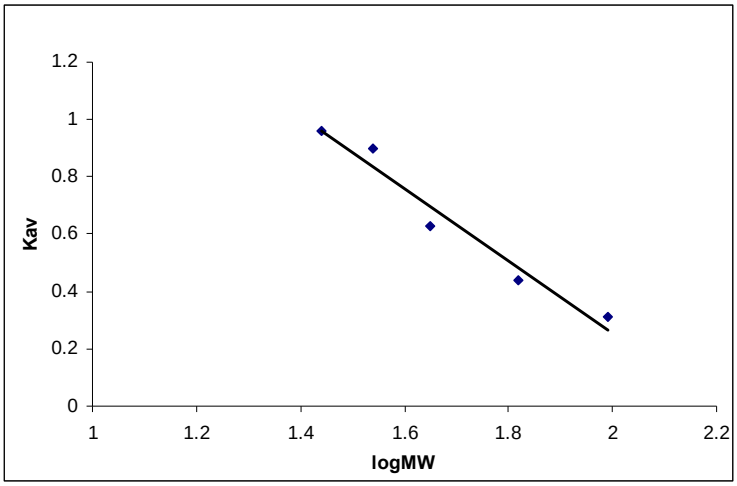

Figure 2.5. Fischer's plot for the estimation of protein M_r from MEC data. The calibration mixture was applied to a Sephacryl S-300 column (25 × 840 mm) and eluted with 50 mM sodium acetate buffer, pH 4.2 (25 $cm.h^{-1}$, 4ºC). The availability constant (K_{av}) was determined for each protein standard.

2.2.2.7 Protein assay of the dimer C2 mutants

The BCA™ Protein Assay (Pierce, Rockford, IL, USA) was used to determine protein concentration. Working reagent solution was made by combining reagent B with reagent A in a 1:20 ratio. Bovine serum albumin (BSA) standards were mixed with working reagent solution (200 µl) in a Nunc® 96 microtiter plate and incubated at 37ºC for 30 min. Proteins interacted with bicinchoninic acid in the working reagent to produce a purple coloured product that was measured at 562 nm. A standard curve was constructed to determine unknown protein concentration (Fig. 2.6).

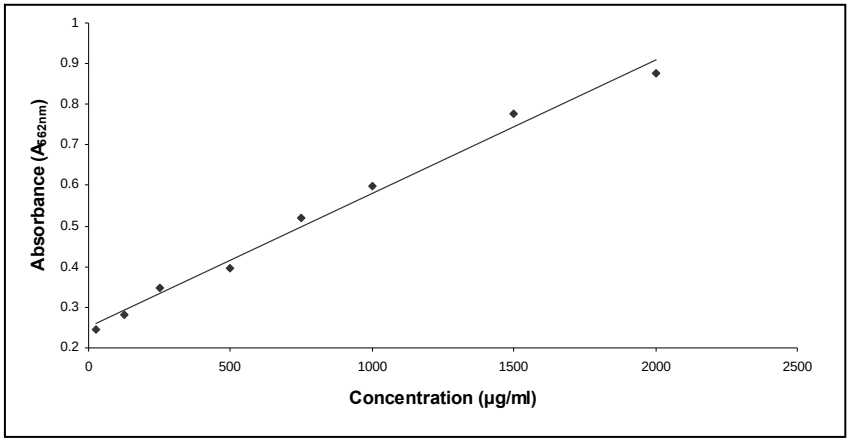

Figure 2.6. Calibration plot for determination of protein concentration. Proteins of unknown concentration were mixed with working reagent solution and A_{562nm} measured. The absorbance was used to extrapolate protein concentration from the standard curve. The equation of the trend line is y = 0.00032x + 0.252, with a correlation coefficient of 0.994.

2.2.2.8 Analysis of the dimer C2 mutants

Laemmli SDS-PAGE

Discontinuous SDS-PAGE was carried out using the method described by Laemmli (1970). The proteins were resolved through two gel types, a stacking gel (pH 6.8) and a smaller pore size resolving gel (pH 8.8) in Tris-Cl buffer. The samples were treated with either the non-reducing treatment buffer [125 mM Tris-Cl buffer, pH 6.8, 4% (w/v) SDS, and 20% (v/v) glycerol] or reducing treatment buffer (10% (v/v) 2-mercaptoethanol added to non-reducing buffer) in a 1:1 ratio (Laemmli, 1970). Reduction and presence of SDS are essential for the determination of M_r of proteins. SDS tightly binds to proteins giving them a net negative charge. Therefore, when an electric field is applied, proteins will be separated solely based on their size. The reducing agent, 2-mercaptoethanol, reduces all the disulfide bridges in the proteins giving rise to a more unfolded protein structure. The proteins were run at 20 mA/gel for 2 h in tank buffer [250 mM Tris-Cl buffer, 192 mM glycine, 0.1% (w/v) SDS, pH 8.3]. There is an inversely proportional relationship between relative mobility of a protein in an electric field and M_r. This relationship can be exploited to determine M_r of unknown proteins. A standard curve was constructed using proteins of known M_r (Fig. 2.7).

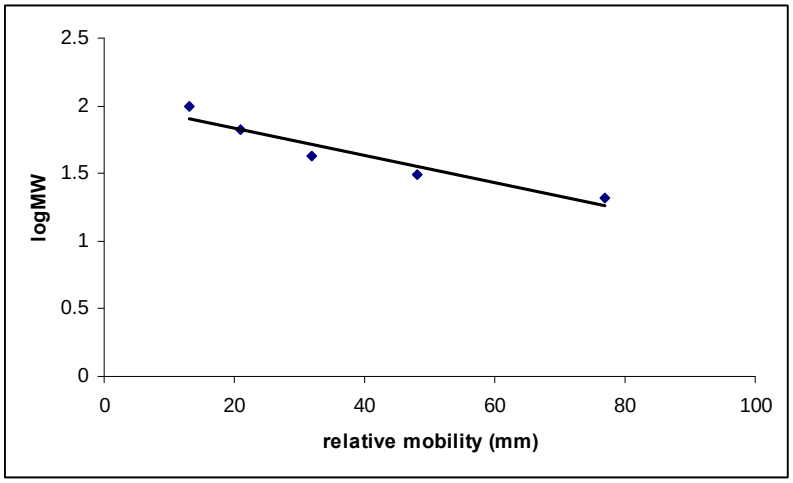

Figure 2.7. Standard curve for estimation of protein M_r by SDS-PAGE. The BioRad low range marker is composed of phosphorylase b (97 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), soybean trypsin inhibitor (31 kDa), and lysozyme (14 kDa). The equation of the trend line is y = -0.0147x + 1.9845, with correlation coefficient of 0.997.

Protein visualisation can be achieved by staining the gel with Coomassie blue R-250 [0.125% (w/v) Coomassie blue R-250, 50% (v/v) methanol, and 10% (v/v) acetic acid]. The detection limit of Coomassie blue is approximately 50-100 ng of protein. When more sensitive detection is required, silver staining is employed (Blum *et al*., 1987). This method has a detection limit of 1-10 ng of protein.

PhastGel® SDS-PAGE

Dimerisation of the C2 mutants was evaluated by PhastGel® SDS-PAGE. The proteins were resolved on a 12.5% PhastGel® homogeneous gel buffered with PhastGel™ SDS-buffer strips. Electrophoresis was carried out using the pre-set program (10 mA, 100 V, 30 min) on a PhastSystem Separation and Control unit (Amersham Biosciences). The gel was stained with PhastGel™ Blue R tablet solution [10% (v/v) acetic acid, 50% (v/v) methanol] for 1 hour.

2.2.2.9 Western blot analysis of dimer C2 mutants

Proteins were transferred from the Laemmli SDS-PAGE gels to a nitrocellulose membrane using a semi-dry blotter (Towbin *et al*., 1979). Afterwards, the nitrocellulose was transiently stained with Ponceau S [0.1% (w/v) Ponceau S in 1% (v/v) glacial acetic acid]. Protein marker bands were marked with a pencil after destaining the membrane with distilled water. The membrane was blocked with low fat milk [5% (w/v) low fat milk in TBS (20 mM Tris-Cl buffer, 200 mM

NaCl, pH 7.4)] for 1 hour. The membrane was washed with TBS (3 × 5 min) and incubated in primary antibody (chicken anti-C2 N-terminal peptide antibody) diluted in 0.5% BSA in TBS for 2 h. The membrane was further washed with TBS (3 × 5 min) and probed with horse radish peroxidase (HRPO) conjugated rabbit anti-chicken antibodies for 1 hour. Finally, the membrane was washed with TBS and incubated with substrate solution [0.006% (w/v) 4-chloro-1-naphthol, 0.1% (v/v) methanol, and 0.0015% (v/v) H_2O_2] until bands became visible.

2.2.2.10 Gelatin SDS-PAGE analysis of dimer C2 mutants

The standard Laemmli SDS-PAGE gel was modified as described by Heussen and Dowdle (1980) to allow for detection of protease activity. Gelatin [1% (m/v)] was added to the running gel buffer and dissolved by heating the mixture (Heussen and Dowdle, 1980). The mixture was added to the rest of the solutions required to cast a SDS-PAGE (12%) gel. SDS-PAGE was carried out as usual. After electrophoresis, the running gel was soaked in two changes of 2.5% (v/v) Triton X-100 for one hour at room temperature. The gel was incubated in congopain assay buffer [100 mM Bis-Tris buffer, pH 6.0, 4 mM EDTA, 0.02% (w/v) NaN_3, 6 mM DTT] for 3 h at 37ºC. The gel was stained with amido black solution [0.1% (w/v) amido black, 30% (v/v) methanol, 10% (v/v) acetic acid, 60% (v/v) water] for one hour. The gel was destained with several changes of destaining solution [30% (v/v) methanol, 10% (v/v) acetic acid, 60% (v/v) water] until clear bands indicating proteolytic activity were visible.

2.2.2.11 Enzymatic characterisation of dimer C2 mutants

Active site titration

The active concentration of cysteine proteases was determined by active site titration with the irreversible inhibitor E-64 (Barrett and Kirschke, 1981; Barrett *et al*., 1982). The enzyme [1 µM diluted in 0.1% (w/v) Brij-35] was combined with E-64 [0.1 to 1 µM diluted in 0.1% (w/v) Brij-35] in congopain assay buffer [100 mM Bis-Tris buffer, 4 mM EDTA, 0.02% (w/v) NaN_3, 6 mM DTT, pH 6.0] and incubated at 37ºC for 30 min. Activity against Z-Phe-Arg-AMC (20 µM) was measured (excitation 360 nm; emission 460 nm) using a FLUOSstar Optima spectrophotometer (BMG Labtech, Offenburg, Germany). A standard curve for AMC product (Fig. 2.8) was prepared by diluting AMC (0-15 nM) in congopain assay buffer.

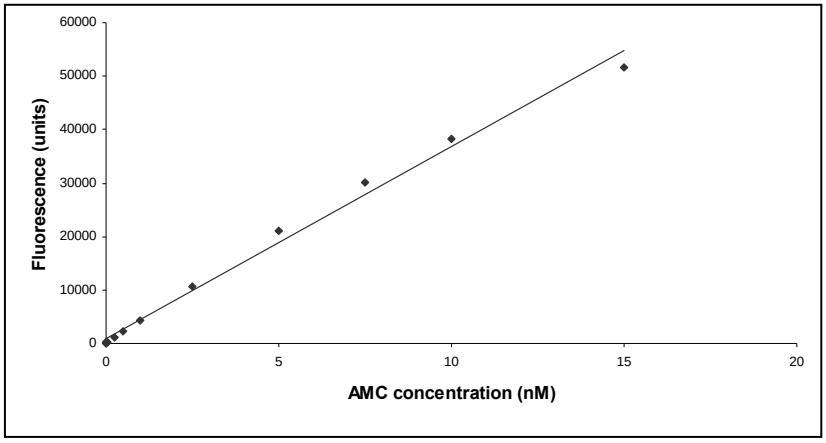

Figure 2.8. Standard curve for AMC product released. AMC stock (1 mM) solution was diluted (0-15 nM) in congopain assay buffer. Fluorescence of the AMC product (100 µl, triplicates) was read (Ex_{360nm}; Em_{460nm}). The equation of the trend line is y = 3.35 + 0.000547x, with a correlation coefficient of 0.998.

pH profiling

Constant ionic strength AMT buffers (100 mM Na-acetate, 200 mM Tris-Cl, 100 mM MES, 6 mM DTT, 4 mM EDTA) were used to determine pH optimum of the cysteine protease (Ellis and Morrison, 1982). The buffers were titrated from pH 4.0 to pH 9.0 in 1.0 pH increments by addition of HCl or NaOH. The enzyme (1.5 ng diluted in 0.1% (w/v) Brij-35) was combined with each buffer at different pH-values and incubated at 37ºC for 1 min. The substrate Z-Phe-Arg-AMC (20 µM) was added and the fluorescence measured (Ex_{360nm}; Em_{460nm}).

Activity Assays

Activity assays were conducted to determine the kinetic constants (K_m and k_{cat}) of the proteases. Enzyme (1.5 ng diluted in 0.1% (w/v) Brij-35) was incubated in congopain assay buffer containing 6 mM DTT for 15 min at 37ºC. The substrate Z-Phe-Arg-AMC (0-30 µM) was added and fluorescence measured (Ex_{360nm}, Em_{460nm}) using a FLUOStar Optima spectrophotometer (BMG Labtech, Offenburg, Germany). The Hyper32© software package (1991-2003, J. S. Easterby, University of Liverpool, UK) was used to plot the Michaelis-Menten curve.

2.3 RESULTS

2.3.1 Mutagenesis, cloning and expression of C2 mutants

The open reading frame (ORF) encompassing the truncated wild-type catalytic domain of congopain (proC2) had been cloned into pPic9 expression vector (pPic9-proC2) and expressed in

P. pastoris (Boulangé *et al*., submitted). Analysis of the expressed truncated C2 revealed dimerisation of the protease at physiological pH, an unexpected feature amongst cathepsin L-like cysteine protease. Moreover, sera from trypano-tolerant cattle recognised the dimer conformation of the protease. Thus, it is important to understand the dimerisation mechanism of congopain in order to gain access to the protective epitopes.

Recombinant plasmid DNA (pPic9-proC2) was isolated using the GeneJet® plasmid miniprep kit (Fig. 2.9, lane 2). The recombinant plasmid DNA was observed as a single band of about 10 kb as expected. The plasmid was linearised with EcoRI (Fig. 2.9, lane 3) and NotI (Fig. 2.9, lane 4). The linear plasmid DNA migrates slower compared to the covalently closed circular plasmid DNA. Double digestion of plasmid DNA (pPic9-proC2) with EcoRI/NotI dropped out a 0.93 kb product (Fig. 2.9, lane 5). The product corresponds to the cloned wild-type catalytic domain of congopain (proC2).

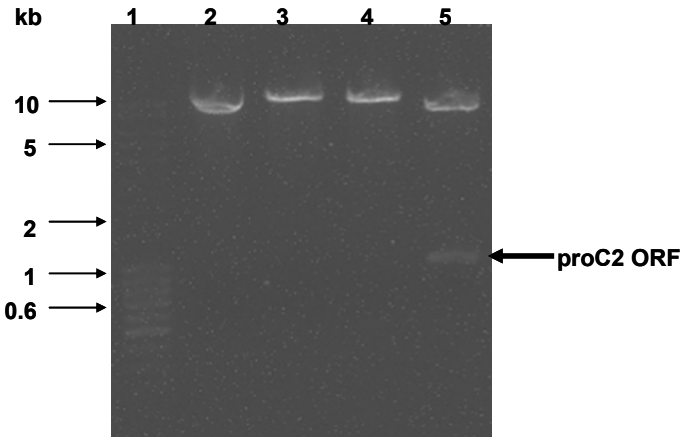

Figure 2.9. Isolation of recombinant plasmid DNA (pPic9-proC2). Plasmid DNA was analysed by electrophoresis on a 1% agarose gel in 1× TAE buffer. Lane 1, MassRuler™ DNA Ladder Mix; lane 2, uncut pPic9-proC2; lane 3, EcoRI cut pPic9-proC2; lane 4, NotI cut pPic9-proC2 and lane 5, EcoRI/NotI cut pPic9-proC2. The arrow indicates the cloned proC2 ORF.

The isolated plasmid DNA (pPic9-proC2) was used as the template for mutagenesis. Mutagenesis was performed through incorporation of the 5' phosphorylated mutagenesis primer (H43W or K39F; E44P) by the flanking central domain (CD) primers during a standard PCR reaction in the presence of a thermostable DNA ligase (Fig. 2.10A).

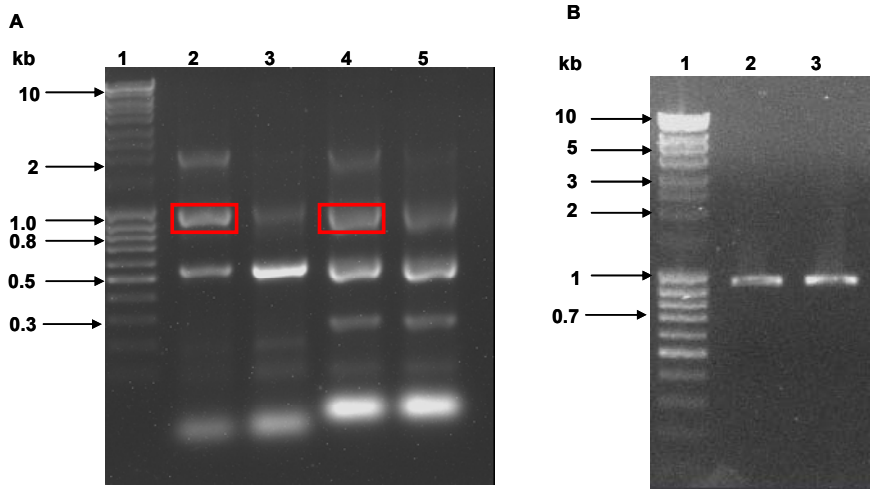

Figure 2.10. Mutation of the proC2 ORF by PCR-based site-directed mutagenesis. Panel A, PCR-based site-directed mutagenesis of proC2 ORF using mutagenesis primers (H43W or K39F; E44P) in the presence of a thermostable DNA ligase. The mutagenesis primer (H43W and K39F; E44P) was incorporated into the full-length product by the flanking (CD) primers in the presence of *Taq* DNA ligase (A). The reaction mixture was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, MassRuler™ DNA Ladder Mix; lane 2, H43W mutagenesis reaction; lane 3, H43W control reaction (without *Taq* DNA ligase); lane 4, K39F; E44P mutagenesis reaction and lane 5, K39F; E44P control reaction (without *Taq* DNA ligase). Panel B, purification of the mutated PCR products. The purified PCR product was analysed by electrophoresis on a 1% (w/v) agarose gel in 1 × TAE buffer. Lane 1, MassRuler™ DNA Ladder Mix; lane 2, purified H43W PCR product and lane 3, purified K39F; E44P PCR product.

After mutagenesis PCR amplification, three prominent bands were observed on the gel in the lanes containing mutagenesis products (Fig. 2.10A, lanes 2 and 4), compared to the two weak intensity bands in the control mutagenesis lanes (Fig. 2.10A, lanes 3 and 5). Two bands were expected in the mutagenesis reaction lanes (Michael, 1994). The bands should be corresponding to the full-length mutagenesis product (0.93 kb) and the truncated product from the mutagenesis primer to the downstream primer (0.6 kb). The third product (2.0 kb) was rather unexpected (Fig. 2.10A, lanes 2 and 4). This is thought to be a combination of the full-length product and the truncated product ligated by the thermostable DNA ligase present in the reaction mix. This hypothesis is confirmed by the absence of this particular band in the control reaction lanes that lack the thermostable DNA ligase (Fig. 2.10A, lanes 3 and 5).

The mutagenesis technique employed in this study involved addition of the mutagenesis primer in ten-fold excess to lower the risk of obtaining a non-mutated full-length product. The reaction mix was supplemented with NAD^+ and DTT, co-factors of the *Taq* DNA polymerase (Michael, 1994; Michael *et al*., 1997). This mutagenesis technique is simple, rapid, reproducible and cost effective. It only involves one PCR reaction and one mutagenesis primer. This is distinct from

other PCR-based mutagenesis techniques like the "megaprimer" (Sarkar and Sommer, 1990) and "overlap extension" (Urban *et al*., 1997) that involve two or more rounds of PCR. Thus, these techniques have a high incidence of errors. Traditional techniques, based on the original protocol described by Kunkel and co-workers are more tedious, costly and of uncertain results (Kunkel, 1985; Kunkel *et al*., 1987). An attempt made with the GeneEditor™ *in vitro* site directed mutagenesis kit was unsuccessful.

The 0.93 kb band corresponding to the expected mutagenesis PCR product was excised from the gel and purified using the peQLab gel extraction kit. The purified product was observed as a band at 0.93 kb (Fig. 2.10B, lanes 2 and 3). No contaminating bands were observed on the gel. The intensity of both the purified bands [proC2 (H43W) and proC2 (K39F; E44P)] was the same. This means that the quantity of DNA contained in each band is the same.

The mutated PCR product [proC2 (H43W) and proC2 (K39F; E44P)] was subcloned into the pGEM-T vector. White colonies were screened with T7 and SP6 universal primers to amplify the cloned mutated PCR product (Fig. 2.11). Plasmid DNA from recombinant colonies produced a 0.93 kb PCR product that migrated slightly higher than expected.

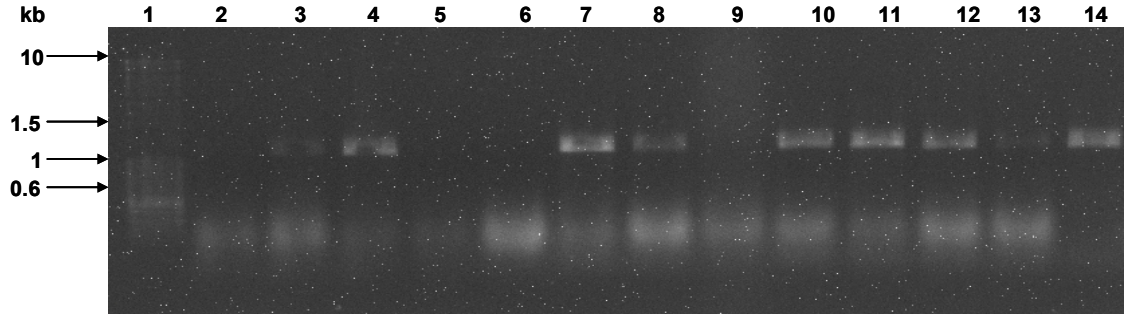

Figure 2.11. Screening of recombinant colonies by colony PCR using T7 and SP6 universal primers. The reaction mixture was analysed by electrophoresis on a 1% agarose gel in 1× TAE buffer. Lane 1, MassRuler™ DNA Ladder Mix, lanes 2-8, H43W clones 1-7 and lanes 9-14, K39F; E44P clones 1-6.

Two recombinant colonies per mutagenesis experiment were grown in selective media overnight. Recombinant plasmid DNA [pGEM T-proC2 (H43W) and pGEM-T-proC2 (K39F; E44P)] was isolated and sequenced to check for the presence of the desired mutations. All clones contained the desired mutations in the correct positions without any extra random mutations (Appendix). This exceptional high success rate (100%) demonstrates the power of the technique. The length

of the amplicon limits the risk of introducing unwanted mutations by *Taq* DNA polymerase. *Taq* DNA polymerase is a low fidelity, less versatile enzyme with an error rate per nucleotide per cycle of approximately 1-2/12000 (Solis Biodyne). The use of high fidelity thermostable polymerases like Vent® or *Pfu*® DNA polymerase could subjugate the risk.

Recombinant pGEM-T vector was double-digested with EcoRI/NotI (Fig. 2.12A, lanes 3, 5, 7 and 9). The digestion produced two prominent bands at 3.2 and 0.93 kb corresponding to pGEM-T vector and the mutated insert respectively. The 0.93 kb band was excised and purified (Fig. 2.12B, lanes 2-5).

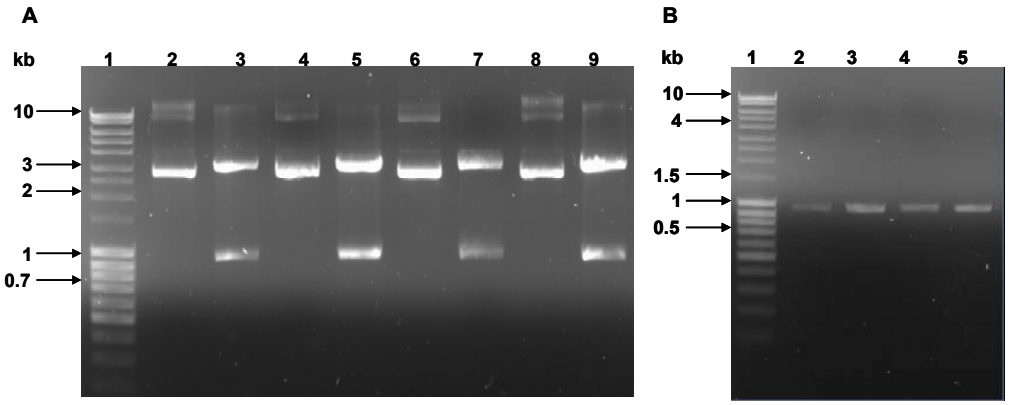

Figure 2.12. Screening of recombinant pGEM-T vector DNA by restriction digestion. Panel A, restriction of recombinant pGEM-T vector DNA with EcoRI and NotI. The digestion mix was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, MassRuler™ DNA Ladder Mix; lane 2, uncut pGEM T-proC2 (H43W) clone 1; lane 3, EcoRI/NotI cut pGEM T-proC2 (H43W) clone1; lane 4, uncut pGEM T-proC2 (H43W) clone 2; lane 5, EcoRI/NotI cut pGEM T-proC2 (H43W) clone 2; lane 6, uncut pGEM-T-proC2 (K39F; E44P) clone 1; lane 7, EcoRI/NotI cut pGEM-T-proC2 (K39F; E44P) clone 1; lane 8, uncut pGEM-T-proC2 (K39F; E44P) clone 2 and lane 9, EcoRI/NotI cut pGEM-T-proC2 (K39F; E44P) clone 2. Panel B, purified mutated insert. Lane 1, MassRuler™ DNA Ladder Mix; lanes 2-3, proC2 (H43W) and lanes 4-5, proC2 (K39F; E44P).

The next step was to ligate the mutated insert [proC2 (H43W) and proC2 (K39F; E44P)] into an expression vector (pPic9). Prior to ligation of the mutated insert into pPic9, the vector was digested with EcoRI/NotI and dephosphorylated with shrimp alkaline phosphatase (SAP) to prevent recircularisation of the vector during ligation. The ligation mix was transformed into competent *E.coli* cells. Alcohol dehydrogenase (AOX) primer pairs were used to screen for recombinants. The recombinant colonies produced a slightly larger band (1.5 kb) due to amplification of some vector sequence by the AOX primers.

Recombinant colonies [pPic9-proC2 (H43W) and pPic9-proC2 (K39F; E44P)] were grown in selective media overnight and recombinant plasmid DNA isolated. Recombinant plasmid DNA was linearised with SacI and used to transform *P. pastoris* GS 115 cells by electroporation. SacI digestion of plasmid DNA drastically improved the transformation efficiency (Wu and Letchworth, 2004).

Transformed *P. pastoris* cells were grown in minimal media until colonies were visible. Colony PCR screening with AOX primers was performed on transformants using the adapted Ayra-Pardo *et al.* (1998) protocol to check integration of plasmid DNA into the *P. pastoris* genome. The clones containing the integrated plasmid DNA produced a 1.5 kb product (Fig. 2.13, lanes 2-7). All the clones screened were positive (Fig. 2.13, lanes 2-7).

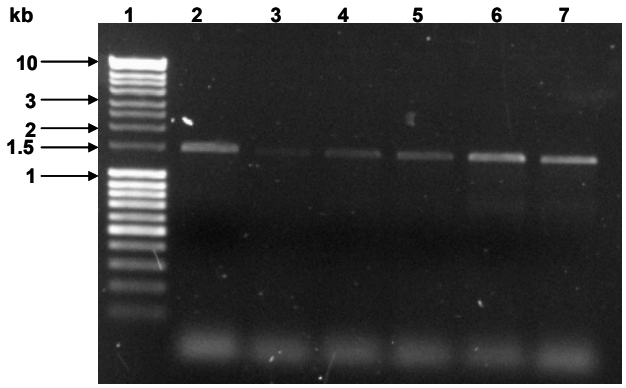

Figure 2.13. Screening of *P. pastoris* clones with AOX primers to confirm integration of mutated DNA in the *P. pastoris* genome. The reaction mix was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE. Lane 1, MassRuler™ DNA Ladder Mix; lanes 2-4, H43W clones 1-3 and lanes 5-7, K39F; E44P clones 1-3.

The PCR products were sequenced in both directions with AOX primers. Comparison of the PCR product sequences with the wild-type C2 sequence confirmed the presence of the introduced mutation in the correct positions. Therefore, the next step was to express the C2 mutants in *P. pastoris*.

Two transformants per construct were grown in minimal media and tested for protein expression. Alongside the expression of the mutants, wild-type C2 was also expressed. All clones expressed recombinant mutant protein at the same level within a given construct (Fig. 2.14, lanes 3-6). The mutants [C2 (H43W) and C2 (K39F; E44P)] and wild-type C2 migrated as prominent bands at 27 kDa. The intensity of the prominent band for C2 [(K39F; E44P)] and C2 was the same.

Therefore, it can be deduced that the level of expression of both proteins is the same. Higher molecular weight, weak intensity bands were observed in all the lanes (Fig. 2.14, lanes 2-6).

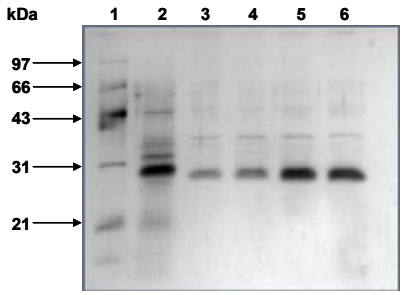

Figure 2.14. Silver stained reducing SDS-PAGE (12%) gel showing the expression of the C2 mutants by *P. pastoris*. Expression supernatants were collected by pelleting the cells at 6000×g for 10 min. Lane 1, BioRad low molecular weight marker; lane 2, C2 supernatant; lane 3, C2 (H43W) supernatant from clone 1; lane 4, C2 (H43W) supernatant from clone 2; lane 5, C2 (K39F; E44P) supernatant from clone 1 and lane 6, C2 (K39F; E44P) supernatant from clone2.

2.3.2 Purification of C2 (H43W) and C2 (K39F; E44P)

Initially, the expression supernatants for the mutants were subjected to three phase partitioning (TPP) as a purification/concentration step. The percentage of ammonium sulfate required to precipitate wild-type C2 was optimised to be 40% (data not shown). This was done by subjecting the expression supernatant to 10% increments of ammonium sulfate until saturation. Hence, 40% ammonium sulfate was applied to precipitate mutants in the presence of tertiary butanol. The precipitated mutants were observed as a prominent band of 27 kDa (Fig. 2.15A and B).

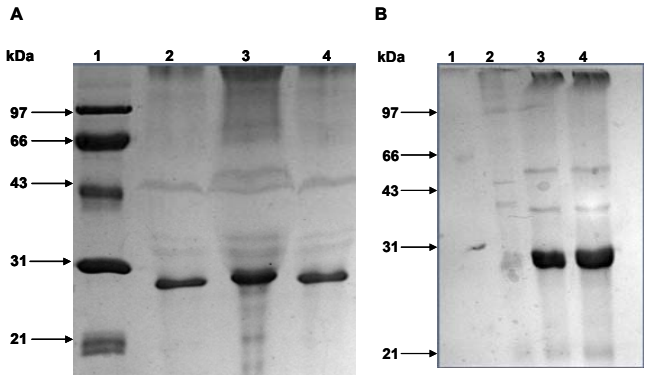

Figure 2.15. Reducing SDS-PAGE evaluation of the C2 mutants after three phase partitioning (TPP) purification. (A) Purification of C2 (H43W) mutant and (B) purification of C2 (K39F; E44P) mutant. The mutants were analysed by 12% reducing SDS-PAGE and the gel was stained with Coomassie blue. Panel A, lane 1, BioRad LMWM; lanes 2-4, different C2 (H43W) TPP fractions after dialysis and concentration. Panel B, Lane 1, BioRad LMWM and lanes 2-4, different C2 (K39F; E44P) TPP fractions after dialysis and concentration.

The next step was to remove the higher molecular weight contaminants that were still present after TPP. Sephacryl S-300 molecular exclusion chromatography was used to achieve this. The different TPP preparations of C2 (K39F; E44P) were pooled, concentrated and applied to the Sephacryl S-300 column. Three elution peaks were observed for this mutant preparation (Fig. 2.16A). This observed elution pattern is similar to that of C2 (H43W) (Fig. 2.16C). The first peak corresponds to higher molecular weight contaminants. The second elution peak corresponds to the contaminants that are approximately 97 kDa in size. The C2 (K39F; E44P) mutant eluted in the last peak at 45 kDa whereas C2 (H43W) eluted at 66 kDa. The introduced mutations might be playing a role in this observed apparent different elution pattern of mutants on the column. Analysis of the peak fractions for both mutants by SDS-PAGE revealed that purification was effected by MEC. Only one band at 27 kDa was observed on both gels (Fig. 2.16B and D). The peak fractions were pooled and concentrated by ultrafiltration using an Amicon PM 10 filter.

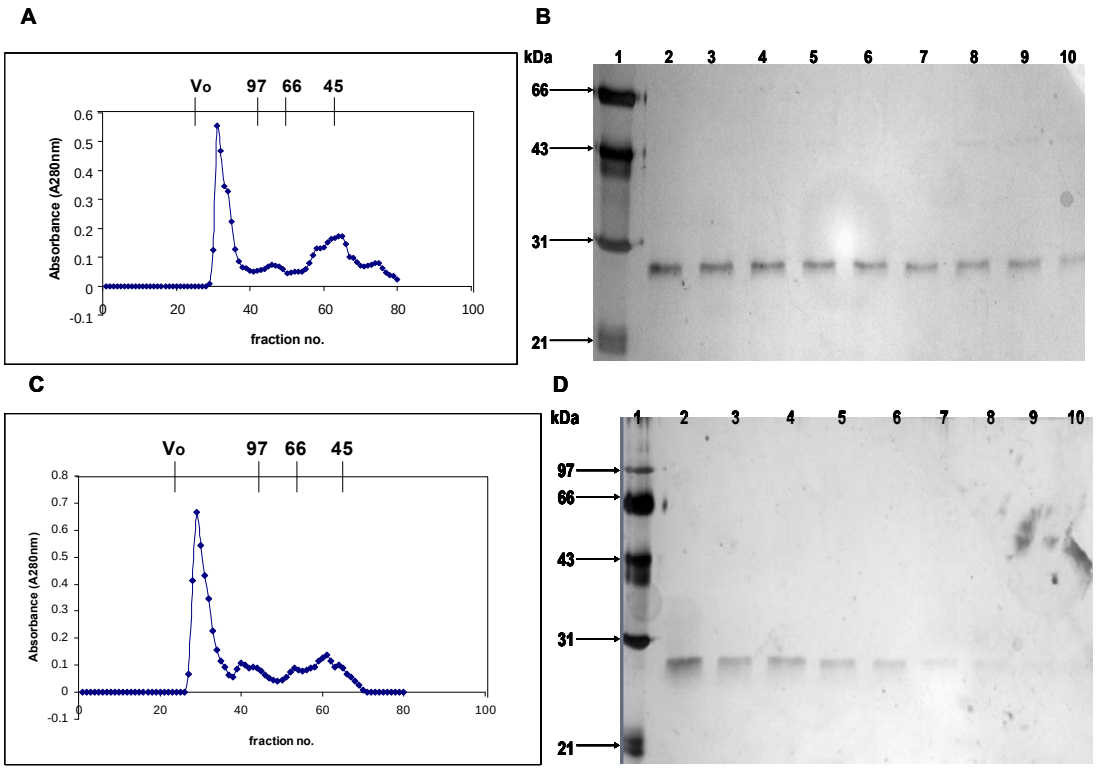

Figure 2.16. Purification of the C2 mutants by molecular exclusion chromatography and analysis by reducing SDS-PAGE. Panels A and C, elution profiles of C2 (K39F; E44P) and C2 (H43W) on a Sephacryl S-300 HR column. Sephacryl S-300 HR (25 × 840 mm, 25 cm.h^{-1}) was equilibrated with 50 mM sodium acetate buffer, pH 4.2. Panels B and D, reducing SDS-PAGE analysis of C2 (K39F; E44P) and C2 (H43W) peak fractions. The fractions were boiled and reduced prior to loading. Panel B, lane 1, BioRad LMWM; lanes 2-10, C2 (K39F; E44P) peak fractions. Panel D, lane 1, BioRad LMWM; lanes 2-10, C2 (H43W) peak fractions.

Wild-type C2 was also purified on the Sephacryl S-300 column. The wild-type C2 eluted in three major peaks as was observed for the C2 dimer mutants (Fig. 2.17A). The protein eluted as a dimer at 45 kDa. This elution profile is similar to that of C2 (K39F; E44P) mutant. Analysis of C2 peak fractions revealed that purification was achieved. The higher molecular weight contaminating bands were no longer present after MEC (Fig. 2.17B). The peak fractions were pooled and concentrated.

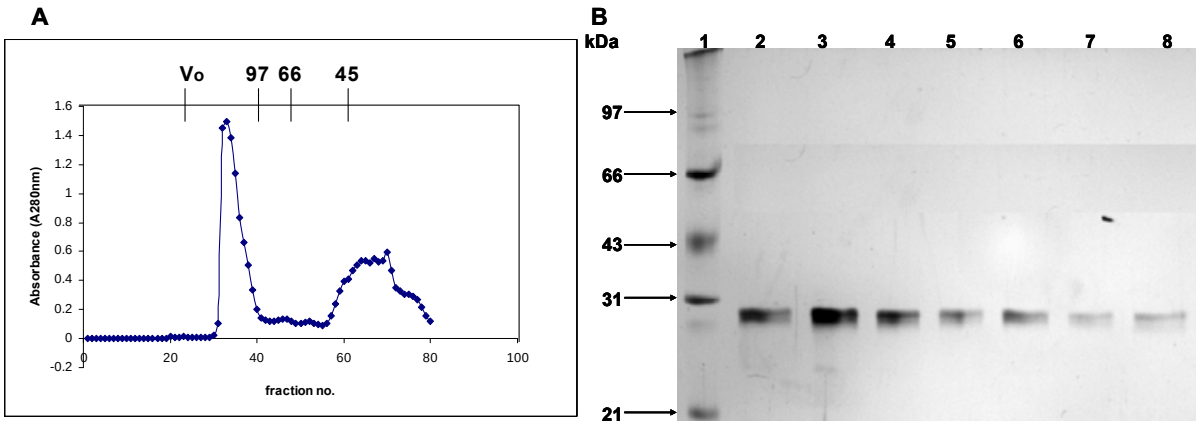

Figure 2.17. Purification of wild-type C2 by MEC. Panel A, elution profile of wild-type C2 after purification on the Sephacryl S-300 HR column (25 × 840 mm, 25 cm.h^{-1}). Panel B, reducing SDS-PAGE (12%) analysis of C2 peak fractions. The fractions were boiled and reduced prior to loading. Lane 1, BioRad LMWM and lanes 2-8, C2 fractions.

The concentrated C2 mutants were observed as a prominent band at 27 kDa (Fig. 2.18, lanes 2 and 4). The C2 (K39F; E44P) mutant band intensity was stronger than that of the C2 (H43W) mutant. This means that removal of the higher molecular weight contaminants was achieved by MEC. Since the contaminants were successfully removed, a dye-binding technique (BCA™ kit) was used to estimate protein concentration. The concentration of C2 (H43W) was estimated to be 10 mg/ml and 14 mg/ml for C2 (K39F; E44P).

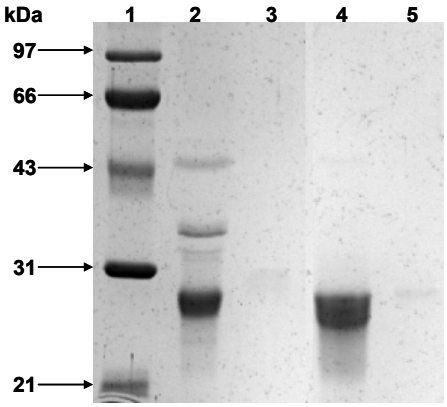

Figure 2.18. Coomassie stained reducing SDS-PAGE gel of the concentrated mutants after MEC. The mutants were boiled and reduced prior to loading. Lane 1, BioRad LMWM; lane 2, concentrated C2 (H43W); lane 3, C2 (H43W) filtrate; lane 4, concentrated C2 (K39F; E44P) and lane 5, C2 (K39F; E44P) filtrate.

Western blotting of the mutants after treatment under non-boiled and non-reduced conditions revealed that the C2 mutants migrate as a smear on a standard Laemmli SDS-PAGE gel (Fig. 2.19, lanes 2 and 5). Therefore, the C2 dimer cannot be resolved on a standard Laemmli gel. There is no plausible explanation for this phenomenon. However, it can be proposed that the SDS present in Laemmli buffers disrupt dimerisation. This is because dimerisation has been proposed to occur through non-covalent interaction of charged amino acid residues. Boiling and/or reduction disrupt dimerisation of the mutants. Hence, the mutants resolved as a conspicuous band at 27 kDa (Fig. 2.19, lanes 3 and 4 and lanes 6 and 7).

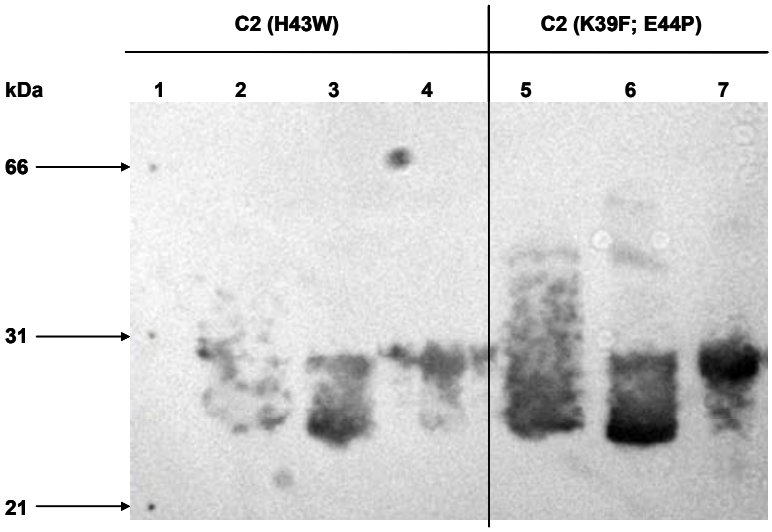

Figure 2.19. Western blotting of the C2 mutants treated under different conditions. Following 12% reducing SDS-PAGE of the C2 mutants, the gel was electroblotted to a nitrocellulose membrane using a semi-dry-blotter run at 7 V overnight. The blot was probed with chicken anti-C2 N-terminal peptide antibodies. Lane 1, BioRad LMWM; lanes 2 and 5, non-boiled and non-reduced C2 mutants; lanes 3 and 6, boiled and non-reduced C2 mutants and lanes 4 and 7, boiled and reduced C2 mutants.

2.3.3. Assessment of C2 mutants' capacity to dimerise

The C2 mutants' capacity to dimerise was assessed by PhastGel® SDS-PAGE. Dimerisation of the mutants is observed on the PhastGel® (Fig. 2.20). The non-boiled and non-reduced C2 (H43W) was observed as a multimer with a mobility shift at 60 kDa (Fig. 2.20, lane 2). The electrophoretic mobility of the C2 (H43W) mutants was unexpected as this mutant is supposed to be a dimer mimicking cathepsin W. Therefore, it is expected to exhibit electrophoretic mobility similar to that of wild-type C2 and migrate at 42 kDa (Fig. 2.20, lane 8). The non-boiled and non-reduced C2 (K39F; E44P) was observed as a multimer at 50 kDa. This relative shift is also peculiar as this mutant is supposed to be a monomer regardless of pH, mimicking cruzipain. Hence, it is expected to migrate at 27 kDa like monomeric wild-type C2. Multimeric conformations of the mutants was abolished by either boiling or reduction with 2-mercaptoethanol. The monomers were observed as doublets at 24 kDa (Fig. 2.20, lanes 3 and 4 and lanes 6 and 7).

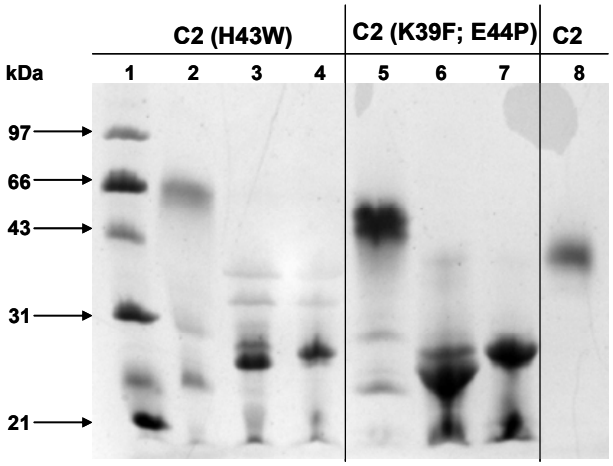

Figure 2.20. Assessment of C2 mutants' capacity to dimerise by PhastGel® SDS-PAGE. The mutants were analysed by 12.5% PhastGel® SDS-PAGE and the gel was stained with Coomassie blue R. Lane 1, BioRad LMWM; lanes 2 and 5, non-boiled and non-reduced C2 mutants; lanes 3 and 6, boiled and non-reduced C2 mutants and lanes 4 and 7, boiled and reduced C2 mutants and lane 8, non-boiled and non-reduced wild-type C2.

Dimerisation of wild-type C2 is a pH dependant phenomenon (Boulangé *et al*., submitted). Hence, it was necessary to subject the C2 mutants and wild-type C2 to different pH-values to assess this phenomenon. Multimerisation was observed to occur at physiological/neutral pH (Fig. 2.21, lanes 2, 4 and 6). Subjecting both mutants and wild-type C2 to acidic pH (pH 3.5) completely abolished multimerisation. Thus, the proteins migrated as monomers at 24 kDa (Fig. 2.21, lanes 3, 5 and 7).

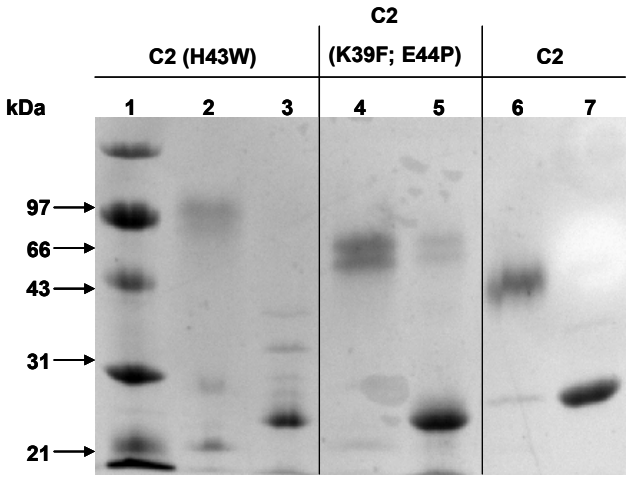

Figure 2.21. pH treatment of the C2 mutants and wild-type C2. The C2 mutants and wild-type C2 samples were analysed by electrophoresis on a 12.5% homogeneous PhastGel® SDS-PAGE and the gel was stained with Coomassie blue R. Lane 1, BioRad LMWM; lanes 2, 4 and 6, samples incubated at pH 7.4 and lanes 3, 5 and 7, samples incubated at pH 3.5.

Gelatin SDS-PAGE analysis was performed to determine the activity of the C2 mutants against gelatin. Both C2 mutants were observed to be active as doublets at 41 kDa (Fig. 2.22, lanes 2 and 3) represented by the clearing on the gel. The C2 (K39F; E44P) mutant was observed to be more active than the C2 (H43W) mutant as indicated by the relative clearing on the gel (Fig. 2.22, lanes 2 and 3). C2 activity was observed as the single band clearing at 41 kDa (Fig. 2.22, lane 4).

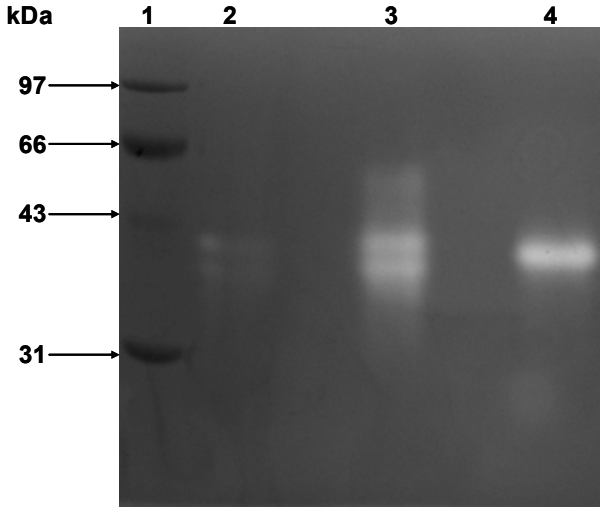

Figure 2.22. Gelatin-containing non-reducing SDS-PAGE (12%) showing activity of the C2 mutants. Non-boiled and non-reduced samples were subjected to electrophoresis on a SDS-PAGE (12%) gel containing 1% gelatin. Lane 1; BioRad LMWM; lane 2, C2 (H43W); lane 3, C2 (K39F; E44P) and lane 4, wild-type C2.

2.3.4 Enzymatic characterisation of the C2 mutants.

Enzymatic characterisation of the C2 mutants was performed to assess the effects of the mutations on activity compared to activity of wild-type C2. Firstly, the active concentration of the enzymes was determined by active site titration using E-64. The C2 (H43W) mutant exhibited 20% activity whereas the C2 (K39F; E44P) mutant contained 50% active enzyme. Wild-type C2 contained 10% active enzyme. Activity of the mutants was assayed against a standard substrate for cathepsin L-like cysteine proteases, Z-Phe-Arg-AMC.

Michaelis-Menten curves were plotted for both the C2 mutants [C2 (H43W) and C2 (K39F; E44P)] and wild-type C2 (Fig. 2.23). These curves were used to determine the affinity constant (K_m) of each protease form for Z-Phe-Arg-AMC. The K_m of C2 (H43W) was determined to be 11.7 μM, 10.3 μM for C2 ($K^{39}E^{44}\Delta FP$) and 9.82 μM for wild-type C2 (Table 2.4). Therefore, it can be concluded that the mutants and wild-type C2 have similar K_m-values for the hydrolysis of Z-Phe-Arg-AMC.

The specificity constant (k_{cat}/K_m) of the C2 mutants and wild-type C2 for the hydrolysis of Z-Phe-Arg-AMC was determined. The C2 mutants displayed similar specificity constants of 1.02 $s^{1}.\mu M^{-1}$ for C2 (H43W) and 0.62 $s^{-1}.\mu M^{-1}$ for C2 (K39F; E44P) while wild-type C2 displayed a very different specificity constant for the hydrolysis of Z-Phe-Arg-AMC of 90.51 $s^{-1}.M^{-1}$ (Table 2.3). Differences in the rate of substrate turnover by the C2 mutants compared to wild-type C2 suggest that the mutations did affect the enzymatic characteristics of the protease.

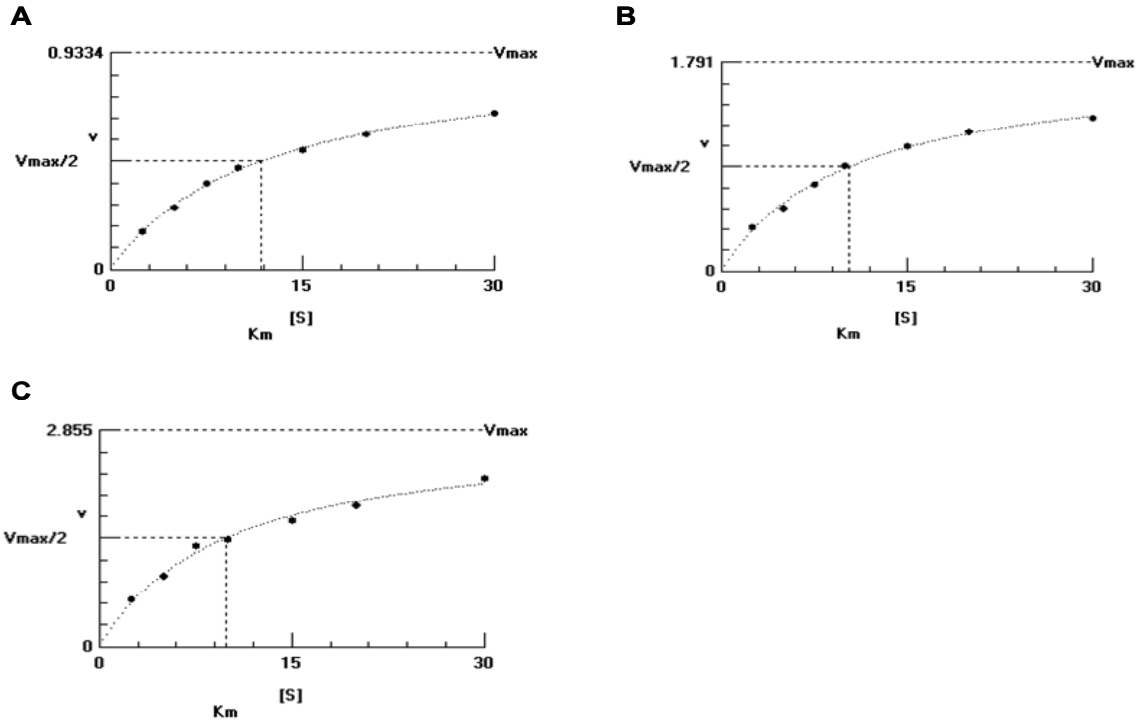

Figure 2.23. Michaelis-Menten curves for the hydrolysis of Z-Phe-Arg-AMC by C2 mutants and wild-type C2. (A) C2 (H43W), (B) C2 (K39F; E44P) and (C) wild-type C2 Michaelis-Menten curves. The activity of the proteases was assayed against Z-Phe-Arg-AMC in congopain assay buffer [100 mM Bis-Tris buffer, pH 6.0, 4 mM EDTA, 0.02% (w/v) NaN_3, 6 mM DTT]. The curves were plotted using the Hyper® 32 software (J.S. Easterby). The curves were used to determine the kinetic constants (K_m, k_{cat} and k_{cat}/K_m).

Table 2.3. Kinetic constants for the hydrolysis of Z-Phe-Arg-AMC by the C2 mutants and wild-type C2. Assays were carried out in congopain assay buffer [100 mM Bis-Tris buffer, pH 6.0, 4 mM EDTA, 0.02% (w/v) NaN_3, 6 mM DTT] with 1.5 ng of active enzyme at 37ºC. The kinetic constants were calculated using the Hyper® 32 software (J. S Easterby).

Enzyme	K_m (μM)	V_{max} (μM.s^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}.μM^{-1})
C2	9.82	2.85	0.89	90.57
C2 (H43W)	11.7	0.93	0.012	1.02
C2 (K39F; E44P)	10.3	1.79	0.0064	0.62

The optimum pH for maximal activity of both mutants and wild-type C2 was determined using constant ionic strength AMT buffers at different pH-values (Ellis and Morrison, 1982) before measuring enzyme activity against Z-Phe-Arg-AMC. The pH activity profiles for C2 (H43W), C2 (K39F; E44P) and wild-type C2 are shown (Fig. 2.24). All the C2 forms share a pH optimum of 6.5. The proteases exhibited activity over a broad pH range from pH 4.0 to pH 7.5 with 80%

activity remaining at physiological pH. Activity of the proteases is completely abolished at pH-values greater than 7.5.

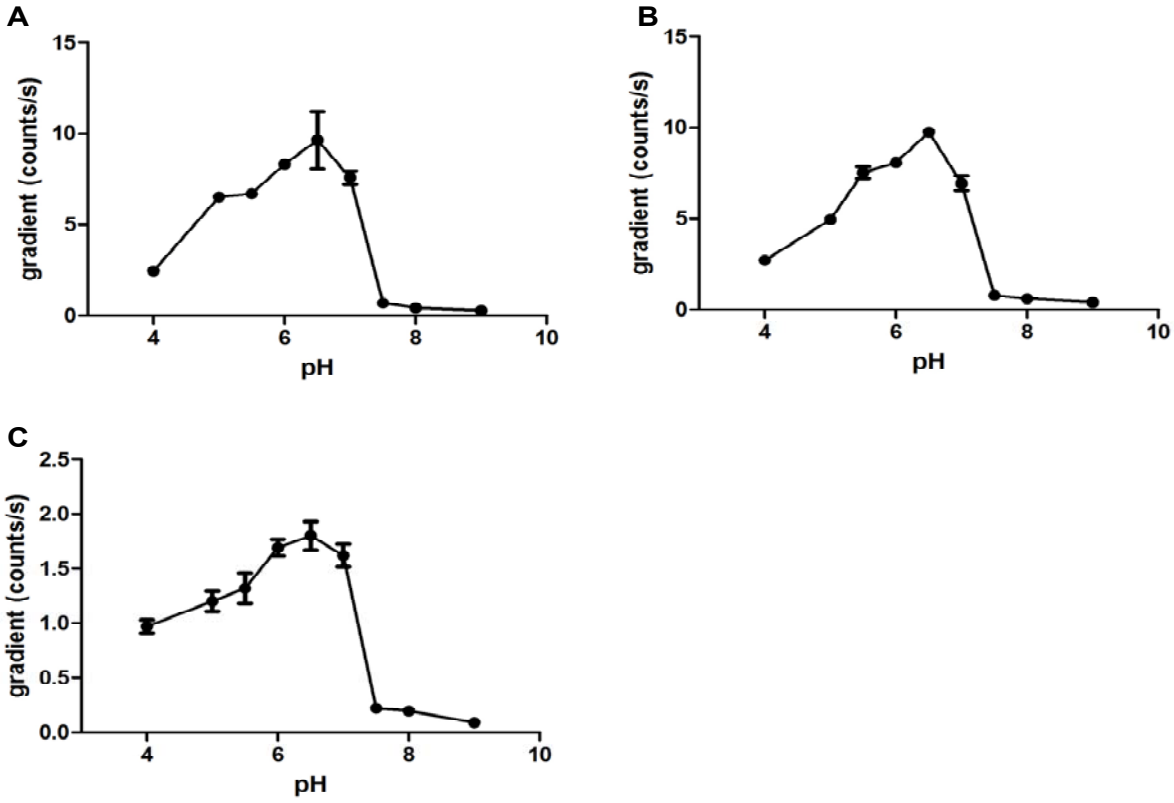

Figure 2.24. pH activity profiles of the C2 mutants and wild-type C2 over a pH range of 4.0-9.0. (A) C2 (H43W), (B) C2 (K39F; E44P) and (C) C2 pH profiles. The constant ionic strength AMT buffers were used to determine optimum pH of the proteases. The curves were plotted using the GraphPad® Prism software.

2.4 DISCUSSION

Vaccination of animals with the pathogenic factors secreted or released by dying parasites has been shown to be the most viable avenue for controlling trypanosomosis in cattle. Authié *et al*. (2001) conducted an immunisation trial in cattle using the recombinant catalytic domain of congopain (C2) as an antigen. Although the protective effects of immunisation were limited, the study revealed the role played by cysteine proteases in the disease process. Analysis of *P. pastoris* expressed C2 by PhastGel® SDS-PAGE revealed unexpected dimerisation of the protease at physiological pH. More relevant to vaccine development was the fact that antibodies from trypano-tolerant cattle specifically recognised the dimeric conformation. Therefore, a hypothesis was formulated that protective epitopes are associated with the dimer. For these reasons, it was important to identify the dimer-associated protective epitopes of congopain and

elucidate the dimerisation mechanism of congopain in order to develop an effective anti-disease vaccine for trypanosomosis.

A dimerisation model was proposed based on the crystal structure of cruzipain, a monomer and cathepsin W that is always a dimer. In the present study the author embarked on mutating certain amino acid residues located in the dimerisation motif of congopain to mimick either cruzipain or cathepsin W using PCR-based site-directed mutagenesis. The mutants were successfully expressed in the *P. pastoris* expression system and their dimerisation capacity was assessed by PhastGel® SDS-PAGE. In addition, the enzymatic characteristics of the mutants were determined to assess the effects of mutations on enzyme catalysis.

In the present study, high level expression of the C2 mutants and wild-type C2 was achieved in the *P. pastoris* expression system. In addition, each protease form remained stable in expression supernatants over an expression period. Although full-length congopain has been expressed in high yields (2 to 4 mg of protein per litre of culture) in yeast, the protease was shown to be highly unstable in the culture supernatants (Boulangé *et al*., submitted). This is the major obstacle encountered with the expression of the full length protein in eukaryotic systems. The stability problem observed with congopain seems to occur in all *Trypanosomatidae* cathepsin L-like cysteine proteases with the exception of cruzipain (Alvarez *et al*., 2002). Expression of functionally active full-length cruzipain was achieved in insect cells, although the yield was low (Alvarez *et al*., 2002). The catalytic domain of parasite cysteine proteases had been successfully expressed in the *P. pastoris* expression system. These include the cathepsin L-like cysteine protease of *Gnathostoma spinigerum*, a mammalian parasite (Kongkerd *et al*., 2008), cathepsin B of *T. congolense* (Mendoza-Palomares *et al*., 2008), rhodesain, a major cysteine protease of *T. b. rhodesiense* (Caffrey *et al*., 2001), and multiple isoforms of cathepsin B-like cysteine proteases from *Trichobilharzia regenti* (Dvořák *et al*., 2005).

Heterologous expression of functionally active cysteine proteases has been achieved in *E. coli* cells. These include full-length cruzipain, the major cysteine protease of *T. cruzi* (Eakin *et al*., 1993), a cysteine proteinase B of *Leishmania braziliensis* (Lanfranco *et al*., 2008), and cysteine proteinase B of *L. mexicana* (Sanderson *et al*., 2000). Attempts to express functionally active congopain in *E. coli* were unsuccessful even after several rounds of renaturation (Boulangé *et al*.,

2001). Functionally active truncated congopain was successfully expressed in a baculovirus expression system with very low yields (Boulangé *et al*., 2001). Therefore, these systems could not be used in this study due to the above mentioned limitations.

Expression of the truncated protease devoid of the highly immunogenic C-terminal domain drastically improved the expression efficiency of cruzipain (Eakin *et al*., 1993; McGrath *et al*., 1995). Similar findings were obtained with heterologous expression of truncated congopain in yeast (Boulangé *et al*., submitted). Truncated congopain was expressed at five fold higher levels than the full-length protease (Boulangé *et al*., submitted). The truncated form of congopain retained its activity, indicating that the C-terminal domain is not essential for activity of the parasite cysteine proteases (Boulangé *et al*., submitted). Sanderson *et al*., (2000) showed that the recombinant cysteine protease B of *L. mexicana* is active in the absence of the C-terminal domain. Thus, the C-terminal domain of parasite cysteine protease plays no role in activity of the enzymes.

The *P. pastoris* expression system allows for post-translational modification i.e. proteolytic processing, folding, disulfide bond formation and glycosylation of proteins (Wu and Letchworth, 2004). This is the main advantage of this system over the bacterial system. Bacterial systems express cysteine proteases as insoluble inclusion bodies that require solubulisation and renaturation for recovery of active protein (Sanderson *et al*., 2000; Aloulou *et al*., 2006; Lanfranco *et al*., 2008). The yeast system allows for extracellular secretion of the proteins, thus significantly simplifying the purification process. This desirable feature is brought about by the presence of the *Saccharomyces cerevisae* α-factor signal peptide in *P. pastoris* expression vectors.

Analysis of *P. pastoris* expression supernatants revealed processing of the proregion of the C2 mutants yielding an active mature enzyme. Processing of the proregion was expected to occur autocatalytically at acidic pH as shown for papain (Vernet *et al*., 1995). The pH of the expression supernatant decreased to pH 5.0. Therefore, the drop in pH of the expression supernatants facilitated the cleavage of the proregion by changing the tertiary structure of the protease, circumventing the inhibition of active site residues by the propeptide (Menard *et al*., 1998). Maturation of congopain has also been shown to occur at pH 8.0 in the presence of dextran

sulfate, an unusual feature amongst papain-like enzymes (Serveau *et al*., 2003). This suggests that congopain might be processed in the bloodstream of the infected host to its active form and can contribute to the pathogenesis of the disease (Serveau *et al*., 2003).

The propeptide of cysteine proteases has been shown to play a dual role in the protease. The propeptide has been empirically shown to be essential for the correct folding of the parent enzyme and inhibits proteolytic activity of the parent enzyme, thereby maintaining it in the inactive zymogen state (Vernet *et al*., 1995; Lalmanach *et al*., 1998; Menard *et al*., 1998; Brooks *et al*., 2000; Sanderson *et al*., 2000; Lanfranco *et al*., 2008). The inhibitory property of the proregion against congopain was shown to be conferred by the conserved "YHNGA" motif. A similar motif "YLNGA" is present in the cysteine protease B of *L. mexicana* (Sanderson *et al*., 2000). The inhibitory motif of trypanosomal and mammalian cysteine proteases was exploited to derive potent inhibitors against the protease with potential chemotherapeutic use (McGrath *et al*., 1995; McKerrow *et al*., 1995; Chowdhury *et al*., 2002).

Processing of the propeptide takes place at several cleavage sites within the protease as shown for the central domain of CP1 (C1), cruzipain and cysteine protease B of *L. mexicana* (Eakin *et al*., 1993; Sanderson *et al*., 2000; Boulangé *et al*., 2001). C1 was observed to migrate as a doublet of 27 kDa (Boulangé *et al*., 2001). The monomeric forms of the C2 mutants exhibited similar behaviour in the present study migrating as doublets of 24 kDa. N-terminal sequencing of C1 revealed inconsistency in the *in vitro* maturation site of the proenzyme (Boulangé *et al*., 2001). The N-terminal sequence of the top band was TGKAPPA and KAPPA for the bottom band (Boulangé *et al*., 2001). The N-terminal sequence of cruzipain consists of APAA while that for the cathepsin L-like cysteine protease of *T. carassii* comprises of APDE (Eakin *et al*., 1993; Ruszczyk *et al*., 2008). Therefore, this evidence confirms that maturation of parasite cysteine proteases occur at different maturation sites, although there are some parasites that share the same maturation site (Lanfranco *et al*., 2008).

Dimerisation of the C2 mutants has been shown to be a pH dependant phenomenon (Boulangé *et al*., submitted). In the present study, both C2 mutants exhibited a higher molecular weight than wild-type C2 at physiological pH. However, at acidic pH, the C2 mutants and wild-type C2 displayed the monomeric conformation. Although the electrophoretic mobility of the C2 mutants

was different from that of wild-type C2, K^{39}, H^{43} and E^{44} are indeed involved in the dimerisation process. However, the mobility shift of the C2 mutants on a PhastGel® suggests that more amino acid residues are involved in the dimerisation mechanism of congopain. This hypothesis was corroborated by the proposed dimerisation mechanism of cathepsin W (Brinkworth *et al*., 2000). Brinkworth *et al*. (2000) identified amino acid residues located at positions 1-10, 40-80 and 170-200 in each monomer to be involved in the dimerisation process. Furthermore, the motif consisting of $R^{40}I^{41}S^{42}F^{43}W^{44}D^{45}F^{46}$ was shown to form part of the dimer interface (Brinkworth *et al*., 2000). The work done by Chen *et al*. (2008) in elucidating the dimerisation mechanism of the SARS 3C-like protease confirms involvement of many amino acid residues in the dimerisation process. Seven amino acid residues including S^{1}, F^{3}, R^{4}, S^{10}, E^{14}, S^{139} and F^{140} were shown to be involved in the dimerisation process of the protease (Chen *et al*., 2008; Shi *et al*., 2008). Substitution of these amino acid residues with alanine was shown to alter the dimer-monomer equilibria of the protease and catalytic activity of the 3C-like protease.

The proposed dimerisation mechanism of congopain was based on the 3D structure of cathepsin W and cruzipain (Prof J. Hoebeke, IBMC, Strasbourg, France, personal communication). Although cruzipain has been demonstrated to exist as a monomer *in vitro*, there is no evidence demonstrating the dimerisation of cathepsin W *in vitro* (McGrath *et al*., 1995; Brinkworth *et al*., 2000). In fact, it was assumed that cathepsin W is a dimer based on a theoretical model that was constructed from the cruzipain central domain (C1) structure (Brinkworth *et al*., 2000). Mammalian cathepsin C is the only cathepsin that has been shown to exist as a multimer *in vitro* (Dahl *et al*., 2001; Horn *et al*., 2002). Analysis of cathepsin C isolated from bovine spleen by gel chromatography revealed that the mature protease is a tetramer (Horn *et al*., 2002). These findings were corroborated by crystallisation of a baculovirus expressed cathepsin C in the presence of the inhibitor Gly-Phe-diazomethane (Molgaard *et al*., 2007). Thus, such flaws in the proposed dimerisation model of congopain explain the unexpected electrophoretic mobility of the C2 mutants on a PhastGel®.

Sera from trypano-tolerant cattle were previously shown to better recognise the dimeric conformation of congopain than the monomeric conformation (Boulangé *et al*., submitted). Elucidation of the dimerisation mechanism of congopain was expected to assist in the identification of protective epitopes associated with the dimer. This information would have

allowed for the design of peptidomimotopes that specifically target the dimer interface of congopain. A similar strategy has been explored for developing inhibitors against the SARS 3C-like protease (Chen *et al*., 2008). An inhibitor targeting the N-terminal finger of the SARS 3C-like protease was shown to specifically bind and competitively inhibit the activity of the protease (Ding *et al*., 2005). Therefore, understanding of the dimerisation mechanism of congopain is crucial for the design and development of an effective inhibitor against the protease.

The C2 mutants exhibited affinity constants (K_m) similar to wild-type C2. The mutants [C2 (H43W) and C2 (K39F; E44P)] and wild-type C2 used as a control had affinity constants around 10.61 µM. These compared favourably to the K_m-values obtained previously for native congopain of 1.5 µM, 4.4 µM, and 7.3 µM for the hydrolysis of Z-Phe-Arg-AMC (Authié *et al*., 1992; Mbawa *et al*., 1992; Chagas *et al*., 1997). Similarity in the affinity constants of wild-type C2, C2 mutants and native congopain suggests that the mutations did not affect the affinity of the protease for the substrate. However, the mutation did affect the proteolytic activity of the protease as indicated by significant differences in the specificity constants for the hydrolysis of Z-Phe-Arg-AMC by the C2 mutants compared to that by wild-type C2. A two-fold difference in the specificity constant was observed between the C2 mutants and wild-type C2.

The C2 mutants exhibited pH profiles similar to that of wild-type C2. They were all active over a broad pH range with optimum activity at pH 6.5 and 80% activity remaining at physiological pH. Similar findings were previously obtained with native congopain (Mbawa *et al*., 1992; Chagas *et al*., 1997). Significant activity at physiological pH suggests that the protease is active in the bloodstream of infected cattle (Serveau *et al*., 2003). This bears interesting consequences for host-parasite interactions.

The focus of the present study was to elucidate the dimerisation mechanism of congopain with the aim of identifying protective epitopes associated with the dimer conformation. Amino acid residues located in the dimerisation motif of congopain (K^{39}, H^{43} and E^{44}) were mutated successfully by PCR-based site-directed mutagenesis to mimick cruzipain and cathepsin W. The mutant forms of C2 were heterologously expressed in *P. pastoris* and purified by three phase partitioning and molecular exclusion chromatography. The effects of mutations on the electrophoretic mobility of the C2 mutants were examined by PhastGel® SDS-PAGE. The

mutations altered the electrophoretic mobility of the C2 mutants compared to wild-type C2. The C2 (H43W) mutant was expected to migrate as a dimer of 42 kDa like wild-type C2 but displayed a mobility shift at 60 kDa. The C2 (K39F; E44P) mutant did not behave as postulated in the model as it was observed to migrate at 50 kDa in a PhastGel®. Multimerisation of the C2 mutants was prevented by subjecting C2 mutants to denaturing conditions and acidic pH. The C2 mutants and wild-type C2 displayed similar pH profiles but different kinetic constants for the hydrolysis of Z-Phe-Arg-AMC. Although this study did not completely elucidate the dimerisation mechanism of congopain, it confirmed the involvement of the dimerisation motif in the dimerisation process. Crystallisation of congopain is required in order to identify surface located protective epitopes associated with the dimer conformation.

CHAPTER 3

RECOMBINANT EXPRESSION AND PURIFICATION OF CONGOPAIN FOR CRYSTALLOGRAPHIC STUDIES

H. H. Ndlovu[1], A. Boulangé[1,2] , F. Vellieux[3] and T. H. T. Coetzer[1]

[1]Department of Biochemistry, University of KwaZulu-Natal (Pietermaritzburg Campus), Private Bag X01, Scottsville, 3209, South Africa
[2]UMR 017 IRD- Cirad Trypanosomes, Campus international de Baillarguet, 34398 Montpellier Cedex 5, France
[3]Institut de Biologie Structurale, Grenoble, France

ABSTRACT

Congopain, the major cathepsin L-like cysteine protease of *Trypanosoma congolense* is an attractive candidate for an "anti-disease" vaccine, and target for the design and development of chemotherapeutic drugs. Structural studies are the most powerful tool for gaining access to the 3-D structure of a protein. The 3-D structure of congopain has not yet been elucidated. Hence, wild-type catalytic domain of congopain (C2), as well as mutant forms, were heterologously expressed in *P. pastoris* and purified by a combination of three phase partitioning, molecular exclusion chromatography and cation exchange chromatography, to crystallisation purity levels. The catalytically active wild-type C2 and the C2 dimer mutants were inhibited with E-64 to prevent self-degradation of the proteases during the purification process. Then, each protease form was concentrated to 5-10 mg/ml per batch, and sent for crystallisation. As crystallisation of these proteases is still underway, we embarked on virtual 3-D modelling. The 3-D structure of congopain was constructed based on the crystal structure of cruzipain using *in silico* homology modelling. The 3-D structure of congopain is similar to the structure of papain-like enzymes, with α-helices and β-sheets forming the L and the R domains. The dimerisation motif was shown to be located on the surface of the protease. While awaiting the crystallographic results, molecular docking studies can be conducted using the 3-D model of congopain to define and construct mimotopes for immunisation studies, or design and development of structure-based inhibitors of the protease.

Keywords: congopain, three phase partitioning, crystallisation, *in silico* homology modelling.

3.1 INTRODUCTION

Trypanosoma congolense is the major causative agent of African animal trypanosomosis, or nagana in sub-Saharan Africa. In search for new drug targets, cysteine proteases were extensively studied in this group of parasites (Authié *et al*., 1992; Coombs and Mottram, 1997; Mottram *et al*., 2004). Analysis of *T. congolense* cDNA revealed two genes coding for related enzymes, namely CP1 (Fish *et al*., 1995, EMBL accession Z25813) and CP2 (Jaye *et al*. 1994, EMBL accession L25130). A functionally active, CP2 type enzyme, named congopain, was purified from *T. congolense* bloodstream lysates by immuno-affinity chromatography (Authié *et al*., 1992; Mbawa *et al*., 1992).

Congopain is a cathepsin L-like cysteine protease with a classical papain-like catalytic triad consisting of C^{25}, H^{159} and N^{175} (papain numbering). The enzyme is synthesised as an inactive zymogen by the parasites (Authié *et al*., 1992; Mbawa *et al*., 1992). The propeptide is essential for correct folding of the protease and inhibits the parent enzyme (Vernet *et al*., 1995). Autocatalytic processing of the propeptide at acidic pH produces a mature enzyme consisting of a catalytic domain (217 residues) and a C-terminal domain (105 residues) (Boulangé *et al*., 2001; Serveau *et al*., 2003). The C-terminal domain is unique to *Trypanosomatidae* cysteine proteases (Mottram *et al*., 1989). Although the C-terminal domain is highly immunogenic, its exact function still remains to be elucidated (Jaye *et al*., 1994; Chagas *et al*., 1997).

Congopain is active over a broad pH range with optimal activity at physiological pH. The enzyme exhibits endopeptidolytic activity against natural protein substrates such as fibrinogen, serum albumin, immunoglobulins and trypanosome variant surface glycoprotein (Mbawa *et al*., 1992). However, the protease has little or no activity against casein and elastin (Mbawa *et al*., 1992). The protease prefers basic amino acid residues such as arginine at P_1 and bulky hydrophobic residues like phenylalanine in P_2. The choice of synthetic substrate for congopain (and all papain-like enzymes) is Z-Phe-Arg-AMC (Authié *et al*., 1992; Chagas *et al*., 1997). However, substrates such as the dibasic Z-Arg-Arg-AMC and monobasic, N-terminally unblocked Arg-AMC are not hydrolysed by congopain (Authié *et al*., 1992; Mbawa *et al*., 1992). The protease is inhibited by the mechanistic class specific inhibitor, trans-epoxysuccinyl-4-guanidinobutane (E-64) and by cystatins (Authié *et al*., 1992; Mbawa *et al*., 1992).

Congopain has been shown to be circulating in the bloodstream of infected cattle (Authié *et al*., 1992). Furthermore, it was found that antibodies from trypano-tolerant, but not trypano-susceptible, cattle recognise congopain *in vitro* (Authié *et al*., 1992). Inhibition of congopain activity by antibodies from trypano-tolerant cattle was thought to contribute to disease resistance of this breed of cattle. Therefore, congopain is a good candidate for an anti-disease vaccine for cattle trypanosomosis. Recombinant catalytic domain of congopain (C2) was heterologously expressed in a baculovirus expression system (Boulangé *et al*., 2001). Cattle were immunised with C2 and challenged with *T. congolense* (Authié *et al*., 2001). This study revealed that C2 partially protected the animals against the disease caused by an infection with *T. congolense* (Authié *et al*., 2001). This study confirmed the role played by cysteine proteases in disease pathology (Authié *et al*., 2001).

The 3-D structure of congopain is yet to be elucidated. However, cruzipain, the congopain homologue from *T. cruzi*, has been crystallised. The crystal structure of cruzipain allowed researchers to develop numerous structure-based peptide inhibitors. McGrath *et al*. (1995) showed that cruzipain activity was inhibited by peptidly-fluoromethylketone inhibitors. The inhibition of cruzipain activity by the inhibitors was correlated with death of parasites in both *ex vivo* tissue culture and *in vivo* mouse models of the disease (Harth *et al*., 1993; McGrath *et al*., 1995). Thus, it is hoped that elucidation of the 3-D structure of congopain will facilitate both the design of structure-based inhibitors of the protease for prophylaxis of animal trypanosomosis and development of mimotopes for use in immunisation studies.

The catalytic domain of congopain (C2) and the mutant proenzyme [proC2 (C25A)] were heterologously expressed in a *P. pastoris* expression system (Boulangé *et al*., submitted). Analysis of recombinant C2 by PhastGel® SDS-PAGE revealed unexpected dimerisation of the protease at physiological pH (Boulangé *et al*., submitted). Moreover, antibodies from trypano-tolerant cattle recognised dimer-specific epitopes (Boulangé *et al*., submitted). Thus, it is important to understand the dimerisation mechanism of congopain in order to design and develop an effective vaccine against the disease. In this study, both C2 and proC2 (C25A) were heterologously expressed in the *P. pastoris* expression system. The proteases were purified by three phase partitioning (TPP), molecular exclusion chromatography and cation exchange

chromatography. Crystallisation of the different protease forms of congopain is currently underway at the Institut de Biologie Structurale (Grenoble, France).

3.2 MATERIALS AND METHODS

3.2.1. Materials

Glycerol stocks of *P. pastoris* containing recombinant pPic9-proC2 and pPic9- proC2 (C25A) were obtained from Dr A Boulangé (University of KwaZulu-Natal, Pietermaritzburg). *P. pastoris* containing the pPic9-proC2 (H43W) and pPic9-proC2 (K39F; E44P) were generated as described in Chapter 2. Sephacryl S-300 and SP Sephadex C-25 were purchased from Sigma (Munich, Germany). Anti-peptide antibodies were produced in chickens against a peptide corresponding to the 22 N-terminal residues of congopain (Mkhize, 2003). IgG anti-dimer C2 antibodies were raised in rabbits against the dimeric form of C2 that was expressed in *P. pastoris* (A. Boulangé, unpublished). Goat anti-rabbit IgG-HRPO conjugate was obtained from Jackson Immunochemicals (USA). The rabbit anti-IgY-HRPO conjugate was obtained from Sigma (Steinhem, Germany).

3.2.2 Methods

3.2.2.1 Expression of C2 and proC2 (C25A) in *P. pastoris*

The glycerol stocks of *P. pastoris* strain GS 115 containing pPic9-proC2 and pPic9-proC2 (C25A) were streaked on YPD plates [1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose, 15 g/l bacteriological agar] containing ampicillin (50 µg/ml). It is important to mention that addition of ampicillin was purely to minimise the risk of bacterial contamination of the culture. The plates were incubated at 30ºC for 2-3 days until the colonies were sizable. Liquid YPD (100 ml) containing ampicillin (50 µg/ml) was inoculated with a single colony of yeast. The culture was incubated at 30ºC in a baffled flask with agitation for two days. This pre-culture was transferred into 1 L BMGY [1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer, pH 6.5, containing 1.34% (w/v) yeast nitrogen base without amino acids (YNB)]. The culture was grown at 30ºC for a further two days to increase biomass. The cells were pelleted (6000×g, 10 min, RT) and resuspended in 500 ml buffered minimal media [BMM, 100 mM potassium phosphate buffer, pH 6.5 containing 1.34% (w/v) YNB, 0.0004% (w/v)

biotin, 5% (v/v) methanol]. The culture was grown in a baffled flask covered with cheesecloth to facilitate aeration, in an orbital shaking incubator at 30ºC for four to six days to allow for protein expression. The culture was supplemented with 0.5% (v/v) methanol daily to induce protein expression. *P. pastoris* secretes proteins into the expression media. Cells were pelleted (6000×g, 10 min, RT) and supernatants were kept at -20ºC until required for purification.

3.2.2.2 Purification of the proteins

Three phase partitioning (TPP)

Three phase partitioning (TPP) was carried out as described by Pike and Dennison (1989). The expression supernatants were thawed and filtered through Whatman No 1 filter paper. Pro-C2 was autocatalytically processed to mature C2 by lowering the pH to 4.2 with phosphoric acid. Tertiary butanol was added to 500 ml supernatant (1/3 final volume, 250 ml). The proteins were precipitated with 40% (w/v) ammonium sulfate (40% of final volume, 300 g) (Pike and Dennison, 1989). Ammonium sulfate was dissolved by gentle stirring with a magnetic stirrer. The precipitated protein mixture was centrifuged (6000×g, 10 min, 4ºC) using a pre-chilled swing-out rotor. The protein layer was collected at the interface between the tertiary butanol and aqueous phases and re-dissolved in dialysis buffer (PBS, pH 7.2). Excess salt and tertiary butanol were removed from the purified protein by overnight dialysis at 4ºC (PBS, pH 7.2) using a 10 kDa cut-off dialysis membrane. The diluted protein was concentrated by dialysis against solid polyethylene glycol M_r 20 000 (PEG 20 000) until 5-10 ml remained in the dialysis bag.

Molecular Exclusion Chromatography

Molecular exclusion chromatography (MEC) was used to remove the higher molecular weight contaminants. The Sephacryl S-300 HR column (25 × 840 mm) was equilibrated with one column volume of 50 mM sodium acetate buffer, pH 4.2 (25 $cm.h^{-1}$, 4ºC). The column was calibrated as described in Section 2.2.2.6. The TPP concentrated protein sample (2.5% of column volume, 8 ml) was applied to the column and eluted with 50 mM sodium acetate buffer, pH 4.2. Protein elution was monitored by measuring absorbance at 280 nm.

Cation exchange chromatography

Cation exchange chromatography was performed as a further purification step for C2 and proC2 (C25A). The sulfopropyl (SP) Sephadex C-25 cation exchange column (10 ml) was equilibrated

with 50 mM sodium acetate buffer, pH 4.2. The pooled MEC fractions were cycled overnight through the column at 4ºC (0.5 cm/min). The column was washed with 50 mM sodium acetate buffer, pH 4.2 (5 column volumes) and eluted with a linear gradient of NaCl from 0 to 1 M over 5 column volumes. The A_{280nm} was measured for all the eluted fractions (1.5 ml). The fractions containing C2 or proC2 (C25A) were pooled and concentrated by ultrafiltration using an Amicon PM 10 membrane. The column was washed with 1 M NaCl (5 column volumes) to remove traces of bound proteins. The column was regenerated using one column volume each of 0.1 M HCl and 0.1 M NaOH and finally re-equilibrated with Na-acetate buffer (50 mM, pH 4.2, 5 column volumes).

3.2.2.3. Protein Quantitation

Two methods were used to determine protein concentration, i.e. BCA™ Protein Assay kit (Pierce, Rockford, IL, USA) and direct visualisation on reducing SDS-PAGE. The BCA assay was performed following the manufacturer's instructions. This quantification method is based on the binding of bicinchoninic acid to proteins. Briefly, the working reagent was prepared by combining reagent B with reagent A in a 1:20 ratio. Bovine serum albumin (BSA) standards were prepared as described in the manual. The standards were mixed with working reagent solution (200 µl) in a Nunc® 96 microtiter plate and incubated at 37ºC for 30 min. Development of the purple colour was monitored by measuring absorbance at 562 nm (A_{562nm}). A standard curve was constructed to determine the concentration of test samples (Fig. 3.1).

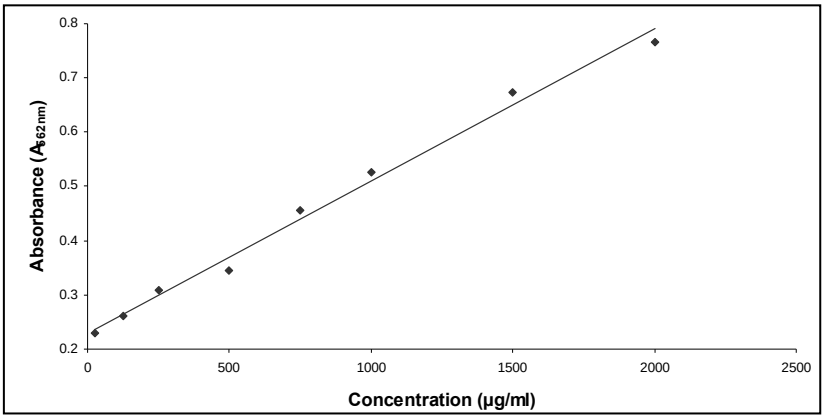

Figure 3.1. Standard curve for the determination of protein concentration. Increasing concentrations of BSA were plotted against absorbance at 562 nm. The equation of the trend line is y = 0.000279 + 0.230, with a correlation coefficient of 0.995.

For the second method of quantification, BSA standards (0.5 mg/ml to 2 mg/ml) were prepared. Equal volumes of the standards and proteins of unknown concentration (5 µl) were run on a reducing 12% Laemmli SDS-PAGE gel. The gel was stained with Coomassie blue R-250 [0.1255 (w/v) Coomassie blue R-250, 50% (v/v) methanol, 10% (v/v) glacial acetic acid]. Protein concentration was estimated by comparing band intensity of proteins of unknown concentration with the standards (Fig. 3.2).

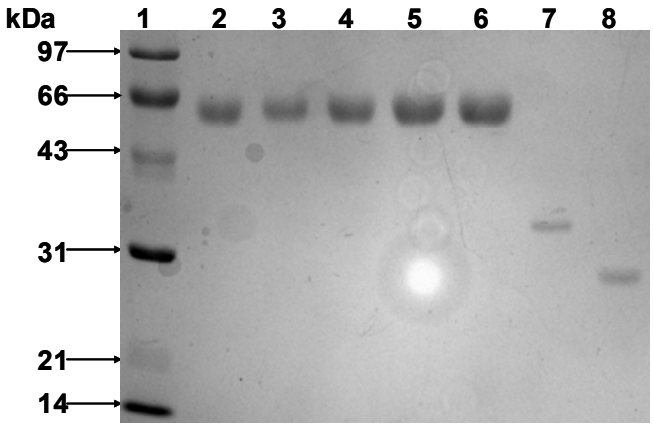

Figure 3.2. Coomassie stained 12% reducing SDS-PAGE gel for direct estimation of protein concentration. The BSA standards were run alongside the proteins of unknown concentration. Comparison of band intensity was used to estimate protein concentration. Lane 1, BioRad LMWM; lane 2, 0.5 mg/ml BSA; lane 3, 0.75 mg/ml BSA; lane 4, 1 mg/ml BSA; lane 5, 1.5 mg/ml BSA; lane 6, 2 mg/ml BSA; lane 7, 1:100 diluted proC2 (C25A) and lane 8, 1:100 diluted wild-type C2.

3.2.2.4 Protein Visualisation

Laemmli SDS-PAGE

Reducing SDS-PAGE was performed using a discontinuous buffer system consisting of running (pH 8.8) and stacking (pH 6.8) Tris-glycine buffers (Laemmli, 1970). The samples were treated with either non-reducing treatment buffer [125 mM Tris-Cl buffer, pH 6.8, 4% (w/v) SDS, 20 (v/v) glycerol] or reducing treatment buffer (10% (v/v) 2-mercaptoethanol added to the non-reducing treatment buffer) in a 1:1 ratio. Electrophoresis was conducted at 20 mA/gel in tank buffer [250 mM Tris-Cl buffer, pH 8.3, 192 mM glycine, 0.1% (w/v) SDS]. The proteins were stained with Coomassie blue R-250 [0.125% (w/v) Coomassie blue R-250, 50% (v/v) methanol, 10% (v/v) glacial acetic acid].

Western blotting

Proteins were transferred to a nitrocellulose membrane from the Laemmli SDS-PAGE gels using a semi-dry blotter (Sigma, Steinhem, Germany), run at 7 V overnight (Towbin *et al*., 1979). The nitrocellulose was transiently stained with Ponceau S [0.1% (w/v) Ponceau S in 1% (v/v) glacial acetic acid] and the positions of molecular weight marker bands marked with a pencil after destaining the membrane with distilled water. The unoccupied sites on the membrane were blocked with low fat milk [5% (w/v) low fat milk in TBS (20 mM Tris-Cl buffer, pH 7.4, 200 mM NaCl)] for one hour. The membrane was washed with TBS (3 × 5 min) and incubated in primary antibody diluted in 0.5% (w/v) BSA-TBS for two hours. The membrane was washed with TBS (3 × 5 min) and incubated with horse radish peroxidase (HRPO) conjugated secondary antibody for one hour. Finally, the membrane was washed with TBS and incubated in substrate solution [0.006% (w/v) 4-chloro-1-napthol, 10% (v/v) methanol, 0.0015 (v/v) H_2O_2] until bands became visible.

3.2.2.5 Deglycosylation of proC2 (C25A)

The proC2 (C25A) sample was incubated with endoglycosidase H according to Caffrey *et al*. (2001). Mutant proenzyme was mixed with 5 M NaCl containing 5% (w/v) SDS. The mixture was reduced with 10% (v/v) 2-mercaptoethanol and boiled for 5 min. The mixture was allowed to cool to room temperature and 0.5 M sodium citrate (one-tenth volume) was added. The samples were split into two tubes and endoglycosidase H (500 U) was added to one of the tubes. Both tubes were incubated at 37ºC for three hours to allow for hydrolysis by endoglycosidase H. Following incubation, the samples were analysed on a 12% reducing SDS-PAGE gel as described in Section 3.2.2.4.

3.2.2.6 *In silico* homology modelling of congopain

The 3-D structure of the catalytic domain of congopain (C2, GeneBank accession number L 25130) was predicted using comparative modelling methods. Firstly, the template protein with protein data bank (PDB) coordinates was chosen for construction of the 3-D structure. In this case, cruzipain (PDB accession number 2aim) was chosen as a template (Gillmor *et al*., 1997). The C2 amino acid sequence was aligned with the template sequence using the ClustalW2™ alignment algorithm (Thompson *et al*., 1994). The alignment file was submitted to the SWISS-

MODEL™ server for prediction of protein structure (Schwede *et al*., 2003; Kopp and Schwede, 2004). The programme models the 3-D structure of the target based on the resolved X-ray crystallographic structure of the template. The predicted model of congopain based on the crystal structure of cruzipain was referred to as C2 and the positive control model was referred to as cruzipain+. The SWISS-MODEL™ server verified the accuracy of the predicted model using ANOLEA™ and GROMOS™ algorithms (Melo and Feytmans, 1998). The verified model was downloaded as a PDP file with PDB coordinates. The predicted model was viewed using the SWISS-PDP viewer (spdbv) programme that was downloaded from the SWISS-MODEL™ server (www.swissmodel.expasy.org//SWISS-MODEL.html, accessed on 16/02/2009). The predicted model was refined using the spdbv programme by performing two steepest descent energy minimisation steps (1:20 and 1:100) to improve the stereochemical geometry of the predicted model. The PROCHECK™ algorithm available on the PDB web site (http://deposit.pdb.org/adit, accessed on 16/02/2009) was used to validate the refined 3-D model.

3.3 RESULTS

3.3.1 Expression and purification of C2 and proC2 (C25A)

High level expression of C2 and proC2 (C25A) was achieved in *P. pastoris*. C2 was observed as a band of approximately 27 kDa while proC2 (C25A) was observed as a prominent band of 36 kDa (Fig. 3.3, lanes 2 and 3). C2 was processed into a mature enzyme through autocatalytic cleavage of the propeptide at acidic pH (Vernet *et al*., 1995; Boulangé *et al*., 2001; Serveau *et al*., 2003). Functional activity of the enzyme is necessary for processing of the propeptide (Menard *et al*., 1998; Boulangé *et al*., submitted). The bulk of proC2 (C25A) existed as a zymogen of 36 kDa (Fig. 3.3, lane 3). Two other bands corresponding to different forms of proC2 (C25A) were observed at 58 kDa and 78 kDa (Fig. 3.3, lane 3).

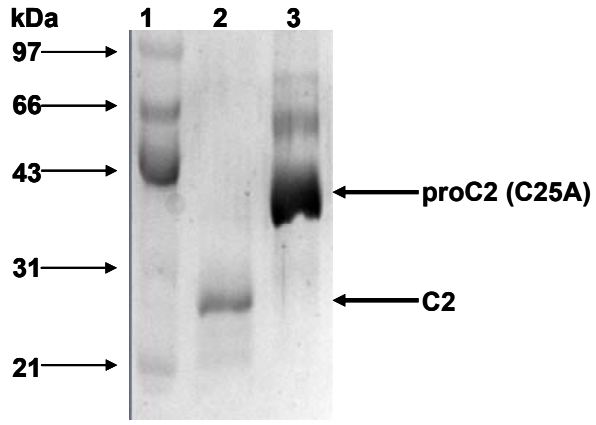

Figure 3.3. Coomassie stained 12% reducing SDS-PAGE gel showing the expression of C2 and proC2 (C25A) by *P. pastoris*. The protein samples were boiled and reduced prior to loading. Lane 1, BioRad low molecular weight marker (LMWM); lane 2, C2 expression supernatant and lane 3, proC2 (C25A) expression supernatant.

Higher levels of the inactive mutant [proC2 (C25A)] were expressed than the functionally active C2 (Fig. 3.3, lanes 2 and 3), as described previously (Boulangé *et al*., submitted). The two higher molecular weight bands (58 kDa and 78 kDa) observed for proC2 (C25A) were thought to correspond to different forms of the protease (Fig. 3.3, lane 3). The hazy 58 kDa band is proposed to be the glycosylated form of proC2 (C25A) and the less conspicuous 78 kDa band is proposed to be the multimeric form of proC2 (C25A).

Three phase partitioning was applied as an initial purification step. The expressed proteins were precipitated with 40% ammonium sulfate in the presence of tertiary butanol. C2 migrated at 27 kDa on a reducing SDS-PAGE gel (Fig. 3.4, lanes 2-9). The protein was observed to be pure except for the presence of higher molecular weight contaminants. Hence, further purification of C2 was still required.

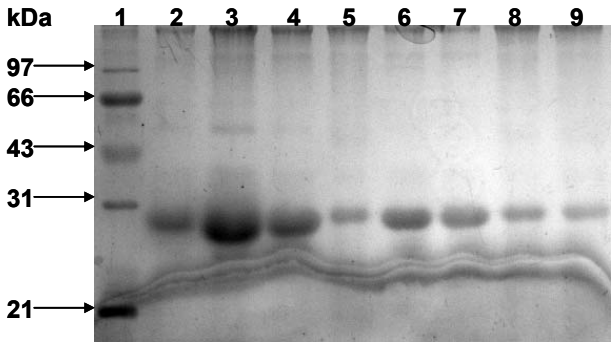

Figure 3.4. Purification of recombinantly expressed C2 by three phase partitioning (TPP). The precipitated proteins were subjected to analysis on a 12% reducing SDS-PAGE gel. Lane 1, BioRad LMWM and lanes 2-9, precipitated C2 fractions from different expression supernatants.

Concerning proC2 (C25A), four bands were observed on a SDS-PAGE gel after TPP (Fig. 3.5, lanes 2-10). The bands were observed at 27, 36, 58 and 78 kDa. The 27 kDa band corresponds to the expected size of the mature enzyme. Although processing of the inactive proenzyme was not expected (Boulangé *et al*., submitted), similar results were obtained with the inactive mutant cruzipain expressed in a bacterial system (Eakin *et al*., 1993). Although the mechanism of processing is not fully understood, it was proposed that secreted bacterial proteases might be involved in the processing of the proregion of cruzipain (Eakin *et al*., 1993). The presence of higher molecular weight contaminants required a further purification step.

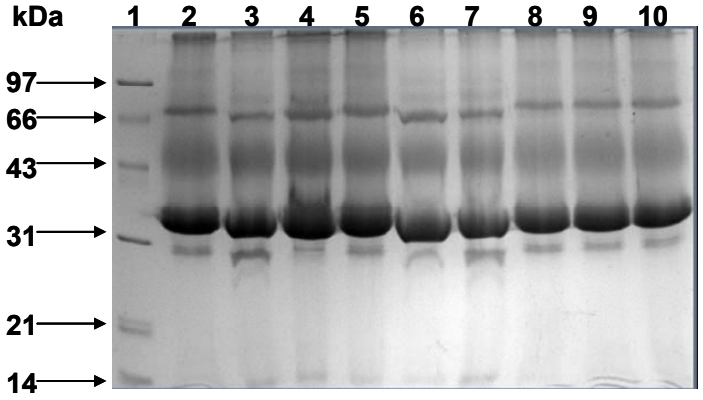

Figure 3.5. Purification of recombinantly expressed proC2 (C25A) by three phase partitioning (TPP). The precipitated proteins were analysed on a SDS-PAGE (12%) gel. Lane 1, BioRad LMWM and lanes 2-10, precipitated proC2 (C25A) fractions from different expression supernatants.

The higher molecular weight contaminants were removed by molecular exclusion chromatography. These contaminants correspond to the first peak on the elution profile of C2 (Fig. 3.6A). C2 eluted at 45 kDa from the column confirming that the protease might have a dimeric conformation (Fig. 3.6A). However, the protease of 27 kDa was shown on a reducing SDS-PAGE gel (Fig. 3.6B and C). It is likely that the presence of reducing or denaturing agents in the treatment buffer disrupts the protease dimer conformation. Analysis of the peak fractions by reducing SDS-PAGE gel confirmed the removal of higher molecular weight contaminants (Fig. 3.6B and C).

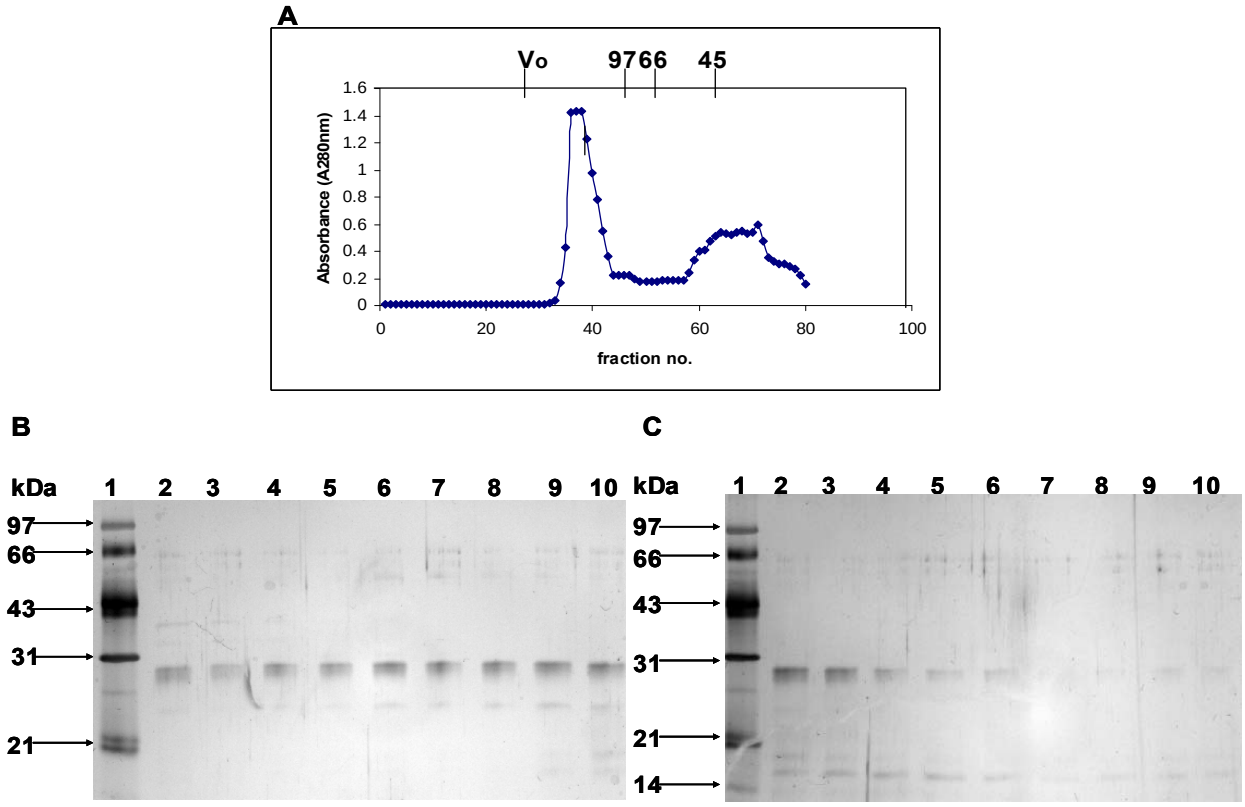

Figure 3.6. Purification of C2 by molecular exclusion chromatography on a Sephacryl S-300 HR column. Panel A, elution profile of C2 on a Sephacryl S-300 HR column (25 × 840 mm, 25 cm.h^{-1}) equilibrated with 50 mM sodium acetate buffer, pH 4.2. The column was calibrated with blue dextran (2000 kDa), phosphorylase b (97 kDa), bovine serum albumin (66 kDa) and ovalbumin (44 kDa). Panels B and C, reducing SDS-PAGE gel analysis of C2 peak fractions. The fractions were boiled and reduced prior to loading. Panel B, lane 1, BioRad LMWM and lanes 2-10, C2 MEC peak fractions 58-66. Panel C, lane 1, BioRad LMWM and lanes 2-10, C2 MEC fractions 67-75.

The MEC elution pattern of proC2 (C25A) showed two peaks (Fig. 3.7A). The first peak contained higher molecular weight contaminants while proC2 (C25A) eluted in the second peak at 66 kDa confirming the existence of a dimeric protease. Analysis of the fractions in the second peak by reducing SDS-PAGE revealed that the 58 and 78 kDa bands of proC2 (C25A) co-eluted with the 36 kDa band of proC2 (C25A) (Fig. 3.7B, lanes 4-10). This was not surprising as these proteins have a molecular weight that is close to each other. A minor 27 kDa band corresponding to the mature enzyme was observed on the reducing SDS-PAGE gels (Fig. 3.7B, lanes 3-9 and C, lanes 2-10). The mechanism of maturation of the inactive mutant proenzyme is not understood.

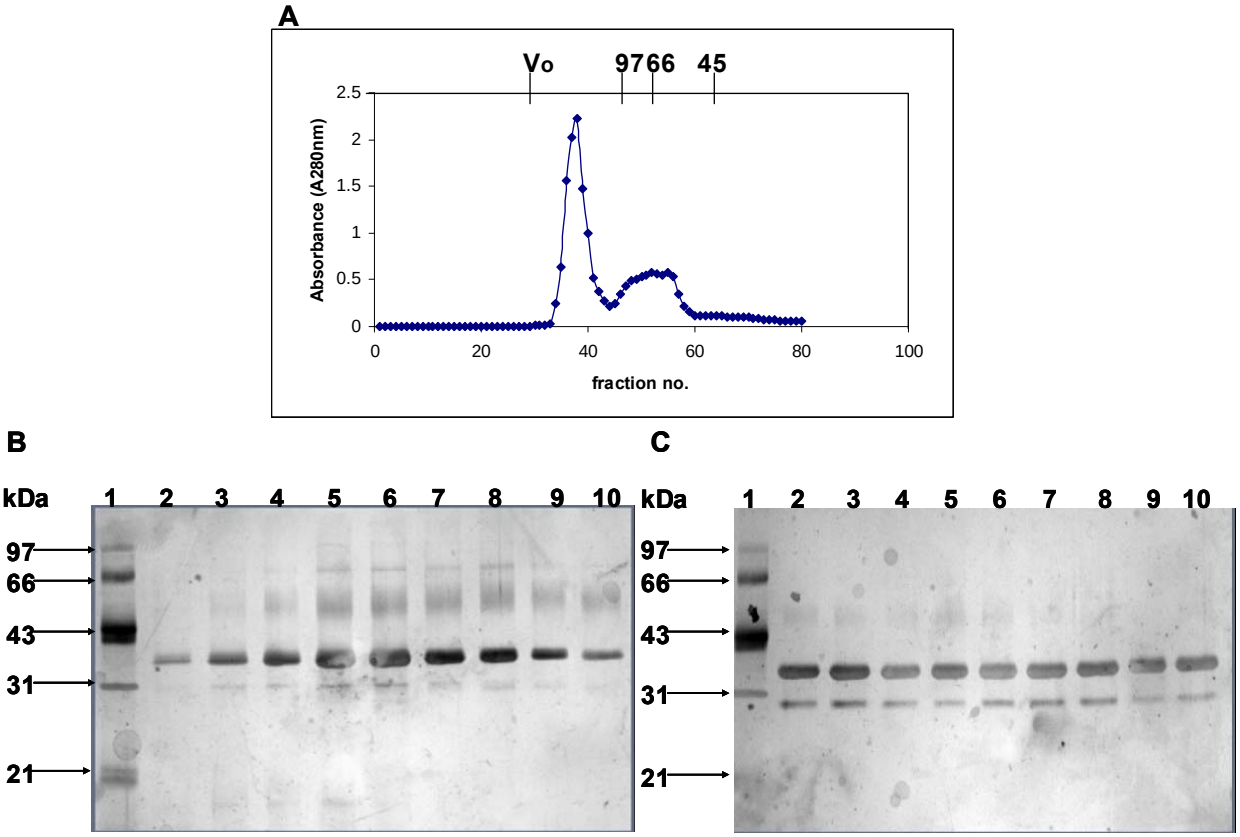

Figure 3.7. Purification of proC2 (C25A) by molecular exclusion chromatography on a Sephacryl S-300 HR column. Panel A, elution profile of proC2 (C25A) on a Sephacryl S-300 HR column (25 × 840 mm, 25 cm.h^{-1}) equilibrated with 50 mM sodium acetate buffer, pH 4.2. The column was calibrated with blue dextran (2000 kDa), phosphorylase b (97 kDa), bovine serum albumin (66 kDa) and ovalbumin (44 kDa). Panels B and C, reducing SDS-PAGE gel analysis of proC2 (C25A) peak fractions. The fractions were boiled and reduced prior to loading. Panel B, lane 1, BioRad LMWM and lanes 2-10, proC2 (C25A) MEC peak fractions 43-51. Panel C, lane 1, BioRad LMWM and lanes 2-10, proC2 (C25A) MEC peak fractions 52-60.

Upon concentration of C2 and proC2 (C25A), the protein samples appeared yellowish, suggesting non-protein contamination of the samples. This was undesirable for crystallography, hence, a third purification step, cation exchange chromatography, was performed. C2 eluted at 200 mM NaCl from a SP Sephadex C-25 column (Fig. 3.8A, fractions 4-8). SP Sephadex C-25 purification of the proteins was achieved as shown by the presence of a single band at 27 kDa on a reducing SDS-PAGE gel (Fig. 3.8B, lanes 4-10).

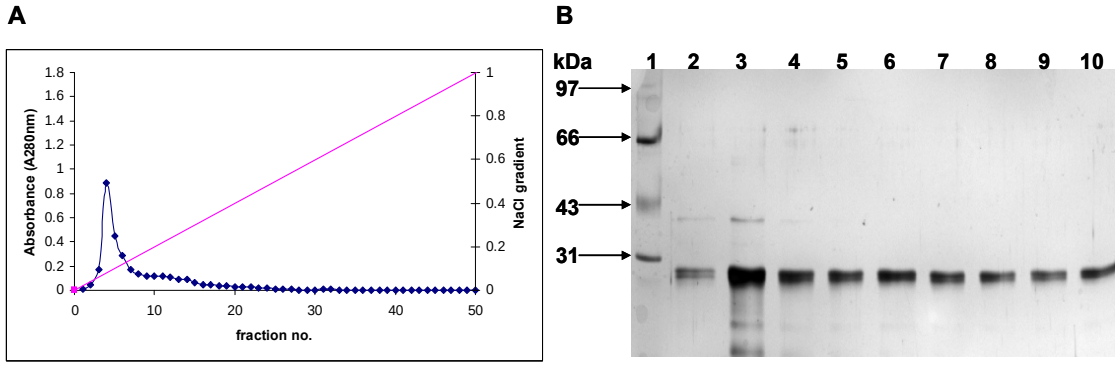

Figure 3.8. Purification of C2 by cation exchange chromatography on a SP Sephadex C-25 column. Panel A, elution of C2 from SP Sephadex C-25 column (1.5 × 10 cm, 0.5 cm/min) using a 0-1 M NaCl gradient in 50 mM sodium acetate buffer, pH 4.2. Elution of the proteins was followed by measuring absorbance at 280 nm for each fraction. Panel B, assessment of the purity of the peak fractions by reducing SDS-PAGE (12%). The gel was stained with Coomassie blue R-250. The samples were boiled and reduced prior to loading. Lane 1, BioRad LMWM and lanes 2-10, C2 IEC peak fractions.

The mutant proenzymes eluted with 300 mM salt from the SP Sephadex C-25 column (Fig. 3.9A, fractions 8-16). A single band corresponding to the mutant proenzyme was observed at 36 kDa on a SDS-PAGE gel (Fig. 3.9B, lanes 2-10). The two larger protein bands (58 kDa and 78 kDa) were not observed. There is an apparent difference in the pI of C2 and proC2 (C25A), indicated by the difference in elution profiles. This is in line with the theoretical pI of the two proteins, of 4.6 for C2 and 8.66 for proC2 (C25A). The higher pI of the proenzyme is due to the presence of a large number of positively charged amino acid residues in the propeptide (Boulangé *et al.*, submitted).

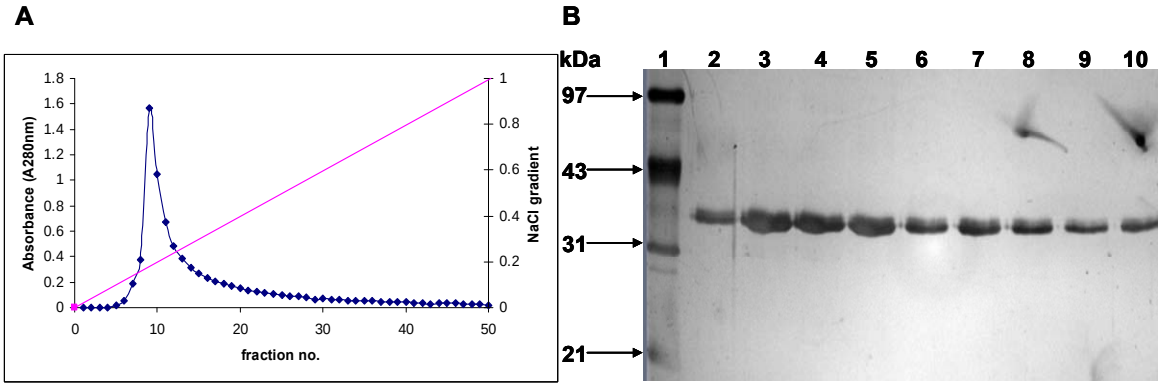

Figure 3.9. Purification of proC2 (C25A) by cation exchange chromatography on a SP Sephadex C-25 column. Panel A, elution of proC2 (C25A) from SP Sephadex C-25 column (1.5 × 10 cm, 0.5 cm/min) using 0-1 M NaCl gradient in 50 mM sodium acetate buffer, pH 4.2. Elution of the proteins was followed by measuring absorbance at 280 nm for each fraction. Panel B, assessment of the purity of the peak fractions by reducing SDS-PAGE (12%). The samples were boiled and reduced prior to loading. The gel was stained with Coomassie blue R-250. Lane 1, BioRad LMWM and lanes 2-10, proC2 (C25A) IEC peak fractions 8-16.

The cation exchange chromatography fractions of C2 and proC2 (C25A) were separately concentrated by ultrafiltration using an Amicon PM 10 membrane. The concentrate appeared colourless, indicating the separation of the non-protein contaminants. Concentrated C2 migrated as a single intense band of 27 kDa (Fig. 3.10, lane 2). The mutant proenzymes migrated at 36 kDa (Fig. 3.10, lane 4). The two higher molecular weight bands (58 kDa and 78 kDa) of proC2 (C25A) re-appeared after concentration. The concentration of C2 was determined to be 10 mg/ml and that for proC2 (C25A) was 15 mg/ml. The yield was 10× higher than that obtained for the expression of C2 in baculovirus (Boulangé *et al*., 2001).

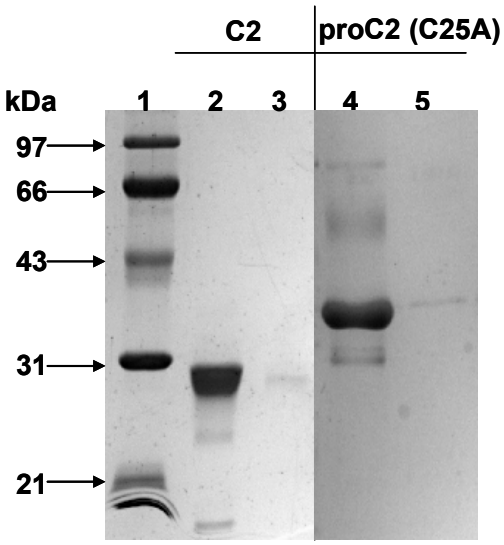

Figure 3.10. Analysis of concentrated C2 and proC2 (C25A) by 12% reducing SDS-PAGE. The samples were boiled, reduced and subjected to electrophoresis on a 12% reducing SDS-PAGE gel. Lane 1, BioRad LMWM; lane 2, concentrated C2; lane 3, C2 ultrafiltration filtrate; lane 4, concentrated proC2 (C25A) and lane 5, proC2 (C25A) ultrafiltration filtrate.

The effects of different treatment conditions on the migration of C2 and proC2 (C25A) was assessed by SDS-PAGE (Fig. 3.11A). The non-boiled and non-reduced C2 migrated as a smear from 36 kDa to 24 kDa (Fig. 3.11A, lane 2). Boiling C2 improved the resolution of the protein on a SDS-PAGE gel. The boiled and non-reduced C2 migrated at 24 kDa (Fig. 3.11A, lane 3). Boiling and reduction changed the mobility of C2 on a SDS-PAGE gel. The reduced protein migrated at 31 kDa (Fig. 3.11A, lane 4). Reduction and boiling serve to relax the structure of the protein by disrupting the secondary and the tertiary structure of a protein, thus allowing for accurate estimation of protein molecular weight (Dennison, 1999).

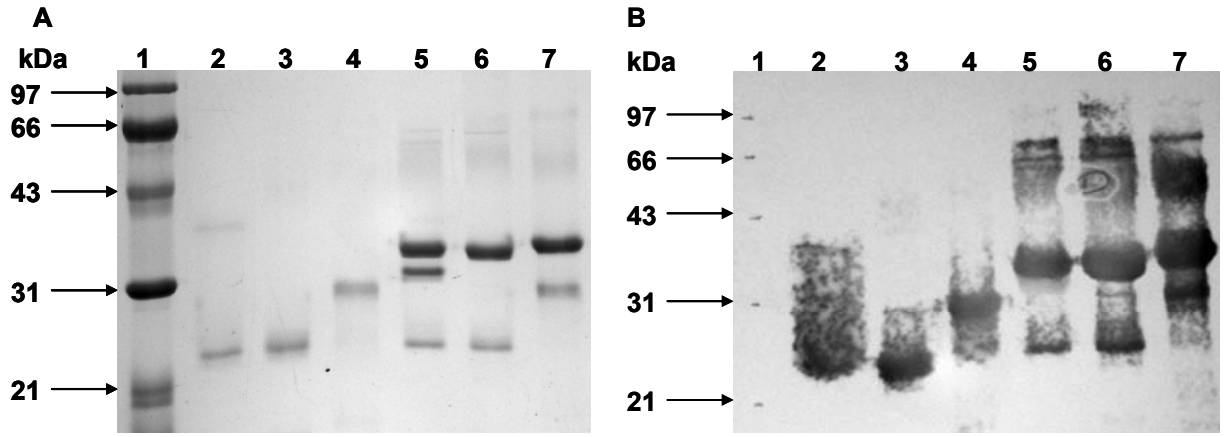

Figure 3.11. Analysis of C2 and proC2 (C25A) treated under different conditions. Panel A, reducing SDS-PAGE (12%) gel showing C2 and proC2 (C25A) subjected to different treatments. Panel B, western blot of C2 and proC2 (C25A) with congopain N-terminal peptide antibodies raised in chickens. The affinity purified peptide antibodies were used at 10 µg/ml. Lane 1, BioRad LMWM; lane 2, non-boiled and non-reduced C2; lane 3, boiled and non-reduced C2; lane 4, boiled and reduced C2; lane 5, non-boiled and non-reduced proC2 (C25A); lane 6, boiled and non-reduced proC2 (C25A) and lane 7, boiled and reduced proC2 (C25A).

The non-boiled and non-reduced proC2 (C25A) was observed as a conspicuous band at 36 kDa (Fig. 3.11A, lane 5). The band corresponding to the mature mutant enzyme was observed at 24 kDa (Fig. 3.11A, lane 5). A 33 kDa band was observed below the 36 kDa band corresponding to the zymogen proC2 (C25A). A faint 78 kDa band is proposed to correspond to the dimeric form of proC2 (C25A), although it does not disappear after boiling and reduction of the protease. Three bands with a M_r of 24, 36 and 58 kDa were observed for the boiled and non-reduced mutant (Fig. 3.11A, lane 6). Four bands were observed for boiled and reduced proC2 (C25A). The bands exhibited a M_r of 27, 36, 58 and 78 kDa (Fig. 3.11A, lane 7). The 58 and 78 kDa bands were observed as very faint bands on a SDS-PAGE gel. A band corresponding to the mature mutant protease displayed a molecular shift of 27 kDa (Fig. 3.11A, lane 7). Thus, boiled and reduced proteins migrate more slowly than non-boiled and non-reduced proteins.

Congopain N-terminal peptide antibodies recognised both C2 and proC2 (C25A) (Fig. 3.11B). The non-boiled and non-reduced C2 was observed as a smear from 36 to 24 kDa (Fig. 3.11B, lane 2). The dimer conformation of C2 was not visualised on a standard Laemmli SDS-PAGE system (Fig. 3.11B, lane 2). Boulangé *et al*., (submitted) showed that C2 does not resolve as a dimer on a standard Laemmli SDS-PAGE gel. Attempts to replicate the gel composition and buffer systems used for PhastGel® in the Laemmli system failed to resolve the dimeric form of C2 (Boulangé *et al*., submitted). The boiled and non-reduced C2 was recognised as a 24 kDa component, while boiled and reduced C2 was recognised as a 31 kDa component (Fig. 3.11B,

lanes 3 and 4). The antibodies recognised four bands of 24, 36, 58 and 78 kDa for non-boiled and non-reduced proC2 (C25A) (Fig. 3.11B, lane 5). Similar bands were recognised by antibodies in boiled and non-reduced proC2 (C25A) (Fig. 3.11B, lane 6). Boiled and reduced proC2 (C25A) showed a band of 37 kDa (Fig. 3.11B, lane 7). The anti-peptide antibodies were used at a high concentration, thus non-specific recognition of proteins was expected.

Treatment of proC2 (C25A) with endoglycosidase H (EndoH) revealed glycosylation of the mutant. In the presence of EndoH, the hazy 58 kDa band disappeared (Fig. 3.12, lane 3). However, a conspicuous 66 kDa band appeared in the EndoH treated proC2 (C25A). Caffrey *et al*. (2001) showed that rhodesain, the major cysteine protease of *T. b. rhodesiense*, is glycosylated. The glycosylation site was shown to be located in the catalytic domain. Another study showed that cruzipain, the congopain homologue from *T. cruzi* is glycosylated (Eakin *et al*., 1993). Therefore, glycosylation seems to be a common feature amongst cysteine proteases of parasitic protozoa.

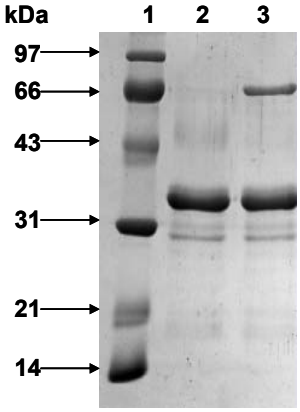

Figure 3.12. Analysis of deglycosylation of proC2 (C25A) on a reducing SDS-PAGE gel. The boiled and reduced proC2 (C25A) was treated with EndoH. Lane 1, BioRad LMWM; lane 2, untreated proC2 (C25A) and lane 3, EndoH treated proC2 (C25A).

Rabbit anti-dimer C2 antibodies recognised all the different multimer conformations of proC2 (C25A), C2 (H43W), C2 (K39F; E44P) and wild-type C2 (Fig. 3.13). These antibodies recognised the 27, 36, 58, 79 and an extra 97 kDa band of the mutant proenzyme [proC2 (C25A)]. The rabbit anti-dimer C2 antibodies were produced against *P. pastoris* expressed recombinant C2. Thus, these antibodies may recognise some of the contaminating yeast proteins. The 97 kDa band may correspond to some *P. pastoris* protein that was not removed during the

purification process. The monomeric forms of C2 (H43W) and C2 (K39F; E44P) were not recognised by these antibodies. These antibodies recognised the 97 kDa band of C2 (H43W) and did not recognise C2 (K39F; E44P).

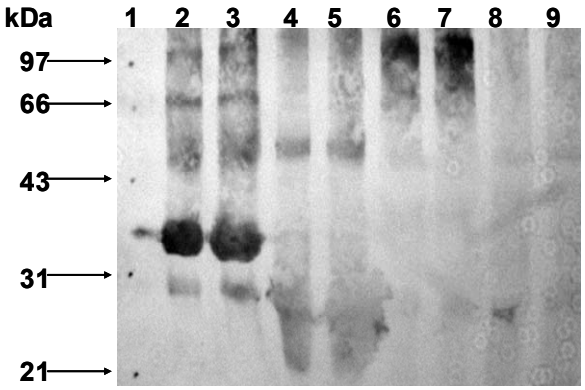

Figure 3.13. Western blot of C2, proC2 (C25A), C2 (H43W) and C2 (K39F; E44P) probed with rabbit anti-dimer C2 antibodies. The samples were transferred to a nitrocellulose membrane from a Laemmli SDS-PAGE gel using a semi-dry blotter, run at 7 V overnight. Lane 1: BioRad LMWM; lanes 2-3, proC2 (C25A); lanes 4-5, wild-type C2; lanes 6-7, C2 (H43W) and lanes 8-9, C2 (K39F; E44P).

3.3.2 Homology modeling of congopain

The primary structure of the catalytic domain of congopain (C2) was aligned with the catalytic domain of cruzipain, human cathepsin L and human cathepsin W using the CLUSTALW2™ programme. C2 displayed 68% identity with cruzipain catalytic domain, 46% identity with human cathepsin L and 37% identity with human cathepsin W (Fig. 3.14). Hence, cruzipain was chosen as a suitable template for constructing the 3-D model of congopain. The percentage identity shared by the template (cruzipain) and the target (C2) sequence is crucial for comparative modelling (Melo and Feytmans, 1998; Schwede *et al*., 2000; Schwede *et al*., 2003). In fact, the quality of the model is directly dependant on the quality of the sequence alignment (Melo and Feytmans, 1998). Alignment displaying lower than 40% identity generates errors in the modelling process (Schwede *et al*., 2000). The alignment of C2 is expected to produce a good quality model as sequence identity is above the 40% threshold limit. The CLUSTALW2™ alignment file was submitted to the SWISS-MODEL™ server for prediction of the 3-D structure of the protease. The SWISS-MODEL™ server modelled 214 amino acid residues for the C2 model based on the cruzipain 3-D structure. The number of amino acids modelled by the SWISS-MODEL™ server is dependant on the number of amino acid residues present in the template crystal structure.

```
CLUSTAL 2.0.10 multiple sequence alignment


congopain        -APEAVDWRKKG-AVTPVKDQGQCGSCWAFSAIGNIEGQWKVAGHELTSLSEQMLVSCD- 57
cruzipain        -APAAVDWRARG-AVTAVKDQGQCGSCWAFSAIGNVECQWFLAGHPLTNLSEQMLVSCD- 57
cathepsin L      -APRSVDWREKG-YVTPVKNQGQCGSCWAFSATGALEGQMFRKTGRLISLSEQNLVDCSG 58
cathepsin W      SVPFSCDWRKVAGAISPIKDQKNCNCCWAMAAAGNIETLWRISFWDFVDVSVHELLDCG- 59
                  .* : ***  .  ::.:*:* :*..***::* * :*         : .:* : *:.*.

congopain        -TNDFGCEGGLMDDAFKWIVSSNKGNVFTEQSYPYASGGGNVPTCDKSGKVVGAKIRDHV 116
cruzipain        -KTDSGCSGGLMNNAFEWIVQENNGAVYTEDSYPYASGEGISPPCTTSGHTVGATITGHV 116
cathepsin L      PQGNEGCNGGLMDYAFQYVQ--DNGGLDSEESYPYEAT---EESCKYNPKYSVANDTGFV 113
cathepsin W      -RCGDGCHGGFVWDAFITVL--NNSGLASEKDYPFQGKVR-AHRCHPKKYQKVAWIQDFI 115
                    . ** **::  **  :   ::. : :*..**: .       *  .     *   ..:

Congopain        DLPEDENAIAEWLAKNGPVAIAVDA--TSFQSYTGGVLTS----CISEHLDHGVLLVGYD 170
cruzipain        ELPQDEAQIAAWLAVNGPVAVAVDA--SSWMTYTGGVMTS----CVSEQLDHGVLLVGYN 170
cathepsin L      DIPKQEKALMKAVATVGPISVAIDAGHESFLFYKEGIYFEP--DCSSEDMDHGVLVVGYG 171
cathepsin W      MLQNNEHRIAQYLATYGPITVTINM--KPLQLYRKGVIKATPTTCDPQLVDHSVLLVGFG 173
                  : ::*  :   :*  **::::::    .   *  *:       * .: :**.**:**:.

Congopain        DTSK-------------------PPYWIIKNSWSKGWGEEGYSALRR-HNQCLMKNLPS 209
cruzipain        DSAA-------------------VPYWIIKNSWTTQWGEEGYIRIAKGSNQCLVKEEAS 210
cathepsin L      FEST-------------------------------------------------------- 175
cathepsin W      SVKSEEGIWAETVSSQSQPQPPHPTPYWILKNSWGAQWGEKGYFRLHRGSNTCGITKFPL 233


congopain        SAVVSGPPPPPPTPTFT 226
cruzipain        SAVVG------------ 215
cathepsin L      ----------------- 
cathepsin W      TARVQKPDMKPRVSCPP 250
```

Figure 3.14. ClustalW2™ alignment of congopain catalytic domain (C2). C2 was aligned with the sequences of cruzipain (PDB accession number 2aim), human cathepsin L (PDB accession number 1mhw) and human cathepsin W (GenBank accession number N 001335). * indicates 100% conserved amino acid residues.

The total energy of the C2 model before energy minimisation was -7968.2 KJ/mol, which was 85% of the total energy of the positive control, cruzipain+ (-9284.1 KJ/mol). Performance of two steps of energy minimisation using the GROMOS™ algorithm improved the stereochemical geometry of the model (Melo and Feytmans, 1998; Kopp and Schwede, 2004). The final energy of the C2 model was reduced to -9410.2 KJ/mol, resulting in 15% improvement in the packing quality of the predicted model. The total energy of the model was negative, confirming that the predicted model exists in the most relaxed conformation (Melo and Feytmans, 1998). The packing quality of the cruzipain+ model improved by 10% after energy minimisation. There was a slight improvement in the total energy of the positive control after refinement using the GROMOS™ algorithm, therefore, authenticating the modelling algorithm (Melo and Feytmans, 1998). The positive control (cruzipain+) model displayed lower total energy than the C2 model before and after energy minimisation. The total energy of the X-ray resolved crystal structure was shown to be lower than the total energy of the predicted model (Melo and Feytmans, 1998). Hence, the predicted model will always exhibit higher total energy than the template.

Ramachandran analysis of the C2 model using the PROCHECK™ algorithm revealed that 85.6% of the non-glycine and non-proline residues fall within the limits of Φ/ψ conformational angles (Fig. 3.15A). This compared very well with the cruzipain+ model that showed 84.7% of the non-proline and non-glycine residues to be located within the most favoured regions of the Ramachandran plot (Fig. 3.15B). The C2 model contained 11.7%, 0.6% and 2.2% residues in the additionally allowed regions, generously allowed regions and the disallowed regions of the Ramachandran plot respectively (Fig. 3.15A). A good quality model should contain very few amino acid residues in the disallowed regions (Kleywegt and Jones, 1996; Kleywegt, 1997). Independent studies modelling *P. falciparum* transketolase and *E. coli* oligopeptidase B showed the presence of a few amino acid residues in the disallowed regions of the Ramachandran plot (Gérczei *et al*., 2000; Joshi *et al*., 2008). Therefore, the quality of the C2 model was validated to be high.

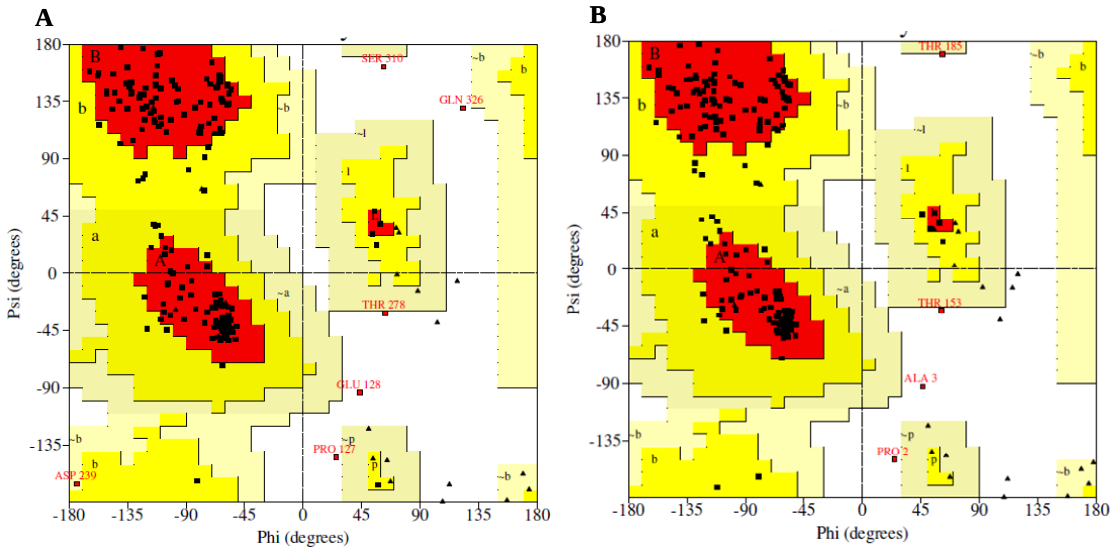

Figure 3.15. Validation of the C2 model by constructing Ramachandran plots using the PROCHECK™ algorithm. Ramachandran plots were derived after energy minimisation of (A) the C2 model and (B) the cruzipain+ model. The different coloured regions indicate the "most favoured regions" (red), "additionally allowed" (yellow), "generously allowed" (light yellow) and "disallowed" regions (white).

The predicted 3-D topology of the C2 model consisted of antiparallel β-pleated sheets and α-helices that folded to form the two domains of C2; namely the R and the L domains (Fig. 3.16A). The predicted 3-D structure of the C2 model is consistent with that of papain-like cysteine proteases (Lecaille *et al*., 2001). The secondary structure of the C2 model consisted of eight antiparallel β-pleated sheets and five α-helices that varied in length (Fig. 3.16A). This was consistent with the 3-D structure of the cruzipain+ model (Fig. 3.16B). The R domain of the C2

model was shown to be composed of six of the eight antiparallel β-pleated sheets and two of the five α-helices (Fig. 3.16A). Therefore, it can be concluded that the R domain of the C2 model consists mostly of antiparallel β–pleated sheets and the L domain consists of α–helices.

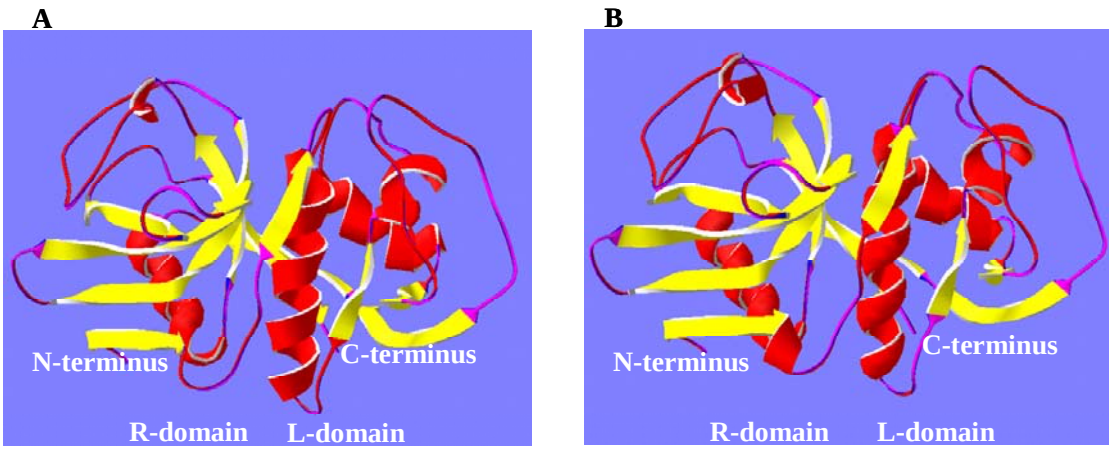

Figure 3.16. Predicted 3-D models of C2 and cruzipain+. Models, (A) C2 and (B) cruzipain+ were analysed using the spdbv viewer programme. The 3-D model was coloured according to the secondary structure. The antiparallel β-pleated sheets are represented in yellow and the α-helices in red.

The predicted 3-D structure of the C2 model followed the common framework of the positive control model (cruzipain+). This was revealed by performing an iterative magic fit between the chiral carbon (Cα) of the predicted model and the respective positive control. Root mean square deviation (RMSD) between the predicted backbone of the model and the experimental control model is another method for assessing the quality of the model (Chothia and Lesk, 1986; Wells *et al*., 2006). The C2 model deviated by 0.22 Å from the cruzipain+ model with 852 atoms superimposed. The minimal deviation between the C2 and cruzipain+ models could be due to the high percentage of sequence identity shared by the template (cruzipain+) and the target (C2). The RMSD values between the predicted model and the experimental positive control should not be more than 0.5 Å (Lesk, 1997). Therefore, the RMSD values of the C2 model further confirmed the quality of the predicted model.

Assessment of the superimposed 3-D structures of the C2 model on the cruzipain+ control structure confirmed the high quality of the model (Fig. 3.17). Traces of a green colour indicating the secondary structure of the C2 model was observed throughout the secondary structure of the positive control (yellow colour) (Fig. 3.17). This confirms that the 3-D structure of congopain is similar to that of cathepsin L-like cysteine proteases.

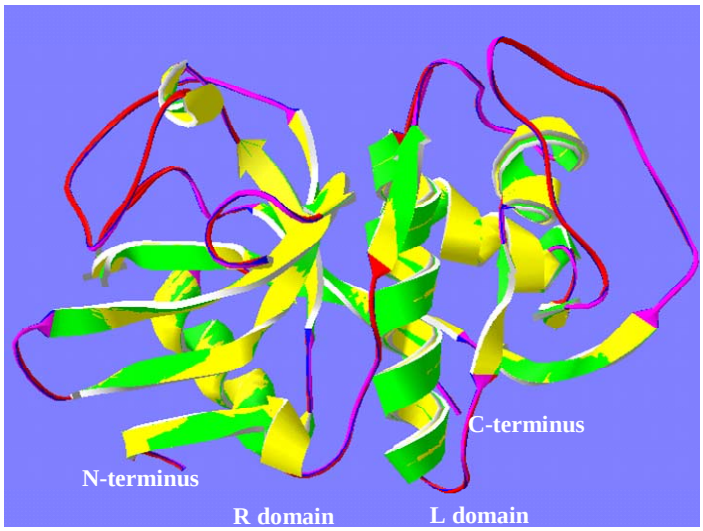

Figure 3.17. Comparison of the 3-D structure of the C2 model with the cruzipain+ control model. The C2 model was superimposed on the cruzipain+ model using the spdbv programme. The secondary structure of the C2 model is indicated by the green colour and the yellow colour represents the secondary structure of the positive control.

The predicted model of C2 obtained in this study was better than that available in the literature (Lecaille *et al*., 2001). The C2 model was constructed by homology modelling of congopain with the experimental 3-D structure of cruzipain (PDB accession number 1aim) (Lecaille *et al*., 2001). The predicted model deviated by 1.06 Å (190 Cα atoms superimposed) from the backbone of the positive control (Lecaille *et al*., 2001). The RMSD of the predicted model is above the 0.5 Å threshold limit, thus, the model could be considered inaccurate. The experimental conditions and method used to resolve the crystal structure of a protein impact greatly on the resulting 3-D structure (Harrison *et al*., 1995; Guex and Peitsch, 1997). Models constructed from experimental proteins resolved by nuclear magnetic resonance (NMR) displayed minimal deviation from the backbone structure of the template (Harrison *et al*., 1995; Guex and Peitsch, 1997). Therefore, the 3-D structures resolved by NMR are more relaxed since the protein is resolved in the solution state (Guex and Peitsch, 1997). The degree of resolution of the X-ray crystal structure influences the deviation of the predicted model backbone from the control (Kleywegt and Jones, 1996). The crystal structure of cruzipain was resolved to 2.1 Å in the presence of inhibitors (Gillmor *et al*., 1997). Judging from the degree of resolution of the crystal structure of cruzipain, the generated model was not expected to deviate to a great degree from the backbone structure of the experimental control. Inhibitors used to stabilise the 3-D structure of a protein during crystallisation also contributes to the degree of deviation of the model from the control backbone structure. Cruzipain was co-crystallised with a benzoyl-Trp-Ala-fluoromethylketone inhibitor

that might have contributed to the stabilisation of the 3-D conformation of the enzyme, resulting in a more relaxed 3-D structure of cruzipain (Gillmor *et al*., 1997). Therefore, the model of C2 constructed in the present study could be used to design structure-based inhibitors of the enzyme.

The possible dimerisation motif of congopain has been identified based on the 3-D structures of cruzipain and cathepsin W (Brinkworth *et al*., 2000; Prof. J. Hoebeke, IBMC, France, personal communication). The dimerisation motif of congopain consists of $W^{38}K^{39}V^{40}A^{41}G^{42}H^{43}E^{44}L^{45}$ (Prof. J. Hoebeke, IBMC, France, personal communication). This motif is similar to that of cathepsin W consisting of amino acid residues $R^{40}I^{41}S^{42}F^{43}W^{44}D^{45}F^{46}$ (Brinkworth *et al*., 2000). There is no empirical evidence showing surface location of the dimerisation motif in congopain. Hence, the 3-D surface structure of the C2 model was analysed by the spdbv programme to identify the location of the dimerisation motif in the enzyme. Indeed, the dimerisation motif was located on the surface of the C2 model (Fig. 3.18A). The cruzipain+ model exhibited similar results, confirming surface location of the dimerisation motif, although the protease is known to exist as a monomer (Fig. 3.18B). The dimerisation motif of the C2 model was virtually superimposable on the cruzipain+ motif with a RMSD value of 0.22 Å (data not shown). Traces of a green colour representing the secondary structure of the C2 model were observed in the backbone and side chains of the cruzipain+ model except where K^{39} was substituted with F^{39} and E^{44} substituted with P^{44} in cruzipain (data not shown).

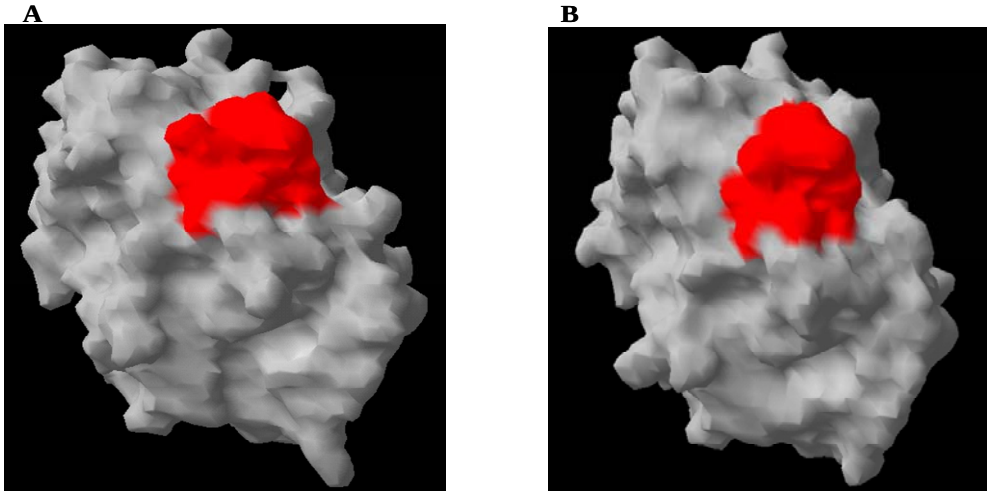

Figure 3.18. Molecular surface structures of the models showing surface location of the dimerisation motif. (A) C2 model surface structure and (B) cruzipain+ surface structure. The molecular surface of both C2 model and cruzipain+ is represented in grey and the dimerisation motif (residues 38-45) in red.

3.4 DISCUSSION

Ever since the emergence of drug resistant parasites, vaccination has been explored as an alternative method for controlling trypanosomosis in animals. However, antigenic variation displayed by trypanosomes is a major impediment to vaccine development. The concept of an anti-disease vaccine was proposed based on targeting pathogenic factors contributing to the disease, rather than targeting the parasite itself (Playfair *et al*., 1991; Authié, 1994). Cysteine proteases such as congopain were implicated as pathogenic factors involved in the development of anaemia in infected animals (Authié *et al*., 2001). Immunisation studies conducted in cattle using baculovirus expressed congopain, administered in conjunction with RWL proprietary adjuvant from SmithKline-Beecham, provided partial protection to animals following a challenge with *T. congolense* (Authié *et al*., 2001). Although there were no significant differences in the development of infection between immunised and non-immunised animals, the immunised animals produced higher titres of antibodies against congopain, maintained or gained weight during infection and developed less severe anaemia. A subsequent study aimed at reproducing these results using *P. pastoris* expressed congopain co-administered with ISA206 oil-in-water adjuvant from Seppic was unsuccessful (E. Authié and A. Boulangé, personal communication). The objective of the present study was to elucidate the 3-D-structure of congopain in order to identify protective epitopes exposed at the surface of the protease. Once these epitopes have been identified, mimotopes could be designed, constructed and used in vaccination experiments.

In the present study, the truncated wild-type catalytic domain of congopain (C2) and the mutant proenzyme [proC2 (C25A)] were heterologously expressed in the *P. pastoris* expression system for crystallisation purposes. Expression of trypanosomal cysteine proteases such as truncated cruzipain (Eakin *et al*., 1993), rhodesain (Caffrey *et al*., 2001), cathepsin B of *T. congolense* (Mendoza-Palomares *et al*., 2008) and congopain (Boulangé *et al*., 2001) had been conducted in either a bacterial or an eukaryotic system. Truncated cruzipain is the only trypanosomal cysteine protease that has been crystallised (Eakin *et al*., 1993; McGrath *et al*., 1995). Attempts to crystallise full-length cruzipain that includes the C-terminal domain have been unsuccessful (Alvarez *et al*., 2002). The difficulty to crystallise the functionally active full-length cruzipain might be due to the instability of the enzyme in culture supernatants (Alvarez *et al*., 2002).

High level heterologous expression of functionally active C2 and proC2 (C25A) was achieved in yeast. The yield of C2 was 10-20 mg/l of culture and 50-100 mg/l of culture for proC2 (C25A). The inactive mutant is expressed at levels five to ten times higher than its active counterpart. The inactive mutant is more stable in culture than active C2. Full-length active cysteine proteases such as congopain and cruzipain have been shown to be highly unstable (Alvarez *et al*., 2002).

Both C2 and proC2 (C25A) were expressed as proenzymes by *P. pastoris*. However, the proregion of both enzymes was cleaved at acidic pH yielding a 27 kDa mature enzyme. The bulk of proC2 (C25A) remained as a zymogen. Maturation of the inactive enzyme was puzzling since enzyme activity is necessary for cleavage of the proregion. Processing of inactive cruzipain (C25A) was shown to involve bacterial proteases co-purified during preparation of inclusion bodies (Eakin *et al*., 1993). On the other hand, activation of lysosomal cysteine proteases such as cathepsin X and C was shown to be performed by cathepsins L and S (Dahl *et al*., 2001; Turk *et al*., 2001). Therefore, processing of the inactive proenzyme may have been conducted by *P. pastoris* proteases that are secreted into the culture media during protein expression. The cleavage sites for removal of the proregion differ depending on the parasite species (Sanderson *et al*., 2000). Congopain displays an "APEA" N-terminal sequence after removal of the propeptide while cruzipain displays "APAA" (Eakin *et al*., 1993; Authié, 1994; Fish *et al*., 1995).

The proenzyme form of congopain [proC2 (C25A)] is glycosylated. The same is true for the cruzipain (Eakin *et al*., 1993) and rhodesain proenzymes (Caffrey *et al*., 2001). This could present major obstacles for the crystallisation of the protease. Glycosylation was shown to inhibit crystallisation of IL-10 (Hoover *et al*., 1999). This is because glycosylation alters size micro-heterogeneity, flexibility of the protein and is a potential source of charge, thus, prevents proper packing of a protein necessary for crystallisation (Wyss and Wagner, 1996). However, glycoproteins such as proteinase A of *Saccharomyces cerevisae* and HIV gp120 envelope protein have been crystallised in spite of their carbohydrate heterogeneity (Aguilar *et al*., 1997; Kwong *et al*., 1998). The glycosylation site for congopain has not been identified whereas glycosylation of cruzipain occurs on Asn^{113} (Eakin *et al*., 1993) and Asn^{295} in rhodesain (Caffrey *et al*., 2001). Although the glycosylation site of rhodesain has been shown to be located in the central domain, the glycosylation site(s) of congopain is proposed to be located in the proregion. The proregion

of congopain contains four asparagine residues (N^{67}, N^{78}, N^{102} and N^{119}) that could potentially be glycosylated.

Crystallisation of the recombinant catalytic domain of congopain (C2) had been attempted (Dr A. Boulangé, University of KwaZulu-Natal, personal communication). Although the results were encouraging, the crystallisation process could not be completed due to shortage of the protease. In order to complete the crystallisation process, highly concentrated C2 (> 5 mg/ml) was required. Subsequent attempts to produce more protease to complete the crystallisation process were unsuccessful (Dr A. Boulangé, University of KwaZulu-Natal, personal communication). The major obstacle encountered was that the ion-exchange purification step could not be reproduced. Therefore, a new purification protocol had to be developed.

To ensure homogeneity of the protease sent for crystallisation, proC2 expressed in *P. pastoris* was matured at pH 4.2 to ensure processing of the protease. Furthermore, the protease was treated with DTT to ensure complete activation of the protease. Following this, C2 as well as the C2 dimer mutants were irreversibly inhibited with E-64 to prevent self-degradation during the purification process. The purification regimen included three phase partitioning (TPP), molecular exclusion chromatography and cation exchange chromatography. Three phase partitioning described by Pike and Dennison (1989) for the purification of cathepsin L from sheep liver was a major breakthrough in the purification process. The method was successful in concentrating the congopain variants from large volumes of culture supernatant while removing the contaminating proteins. The main advantage of this technique was that it was reproducible, cheap and saved significant amount of time (Pike and Dennison, 1989). After TPP, the precipitated protein mixture appeared yellowish in colour indicating the presence of non-protein contaminants. Molecular exclusion chromatography was successfully applied to remove the non-protein contaminants as well as the higher molecular weight contaminants. Finally, cation exchange chromatography was used to ensure homogeneity of the proteins. The highly pure protein sample (> 99% purity) was concentrated to 5-10 mg/ml per batch for crystallisation. The proteins were sent for crystallisation at the Institut de Biologie Structurale as part of a collaborative agreement with Prof. F. Vellieux (Grenoble, France). During the preparation of this dissertation, crystallisation of the proteins was still underway.

While waiting for the crystallographic data, comparative homology modelling was employed to generate the 3-D structure of congopain. Crucial to comparative modelling of protein 3-D structures is the template used to construct the model (Melo and Feytmans, 1998; Schwede *et al*., 2000; Schwede *et al*., 2003). The primary structure of congopain displayed a high level of sequence identity with cruzipain, the major cysteine protease of *T. cruzi*. Moreover, the crystal structure of cruzipain had been elucidated in the presence of various inhibitors (Eakin *et al*., 1993; McGrath *et al*., 1995). Hence, cruzipain was chosen as a suitable template for the generation of the 3-D models of congopain.

The 3-D model of congopain consisted of α-helices and β-sheets that formed the L and the R-domains. The predicted 3-D structure of congopain displayed similar structural features as cruzipain. The 3-D model of congopain revealed surface location of the dimerisation motif as postulated by Prof. J. Hoebeke (IBMC, Strasbourg, France, personal communication). Although the models did not offer much information about how exactly the dimerisation of congopain unfolds, the dimerisation mechanism of cathepsin W was proposed based on a theoretical 3-D model of the protease constructed from the crystal structure of cruzipain and cathepsin K. Brinkworth *et al*. (2000) proposed that an extended loop composed of amino acid residues 40-46 (RISFWDF) forms part of the dimer interface. The mechanism involves interaction between amino acid residues located in the dimer interface (residues 40-80), residues 1-10 and residues 170-200 of each monomer facing opposite directions (Brinkworth *et al*., 2000). It was proposed that the presence of A^{11} in the cathepsin W structure contributes to dimer formation by adding twists to the structure of the protease. Crystallisation of the congopain variants (that is currently underway) and solving the crystal structures is expected to complete the puzzle regarding the dimerisation mechanism of congopain. This is critical for the development of an effective vaccine since protective epitopes have been shown to be dimer-associated (Boulangé *et al*., submitted).

The objective of this study was to determine the 3-D structure of congopain in order to map protective epitopes exposed at the surface of the protease and develop mimotopes to be utilised in vaccination studies. We embarked on large-scale heterologous expression in *P. pastoris* of the wild-type catalytic domain of congopain (C2) as well as mutant forms of congopain. The proteins were purified using a combination of three phase partitioning, molecular exclusion

chromatography and cation exchange chromatography to purity standards required for crystallisation. At the time of writing of this thesis, the crystallographic data was not yet available. Hence, *in silico* homology modelling was performed using the cruzipain crystal structure to gain access to the 3-D structure of congopain. The 3-D model confirmed that the congopain structure is similar to that of papain-like enzymes. The structure is composed of α-helices and β-sheets that fold to form the L and the R domains. Furthermore, the dimer motif identified by Brinkworth *et al*. (2000) in cathepsin W was shown to be located on the surface of the model. This presents strong evidence for its involvement in the dimerisation mechanism of congopain.

CHAPTER 4

INVESTIGATION OF THE ADJUVANT POTENTIAL OF *TRYPANOSOMA CONGOLENSE* BIP USING RECOMBINANT CONGOPAIN AS A MODEL ANTIGEN

H. H. Ndlovu[1], A. Boulangé[1,2] and T. H. T. Coetzer[1]

[1]Department of Biochemistry, University of KwaZulu-Natal (Pietermaritzburg Campus), Private Bag X01, Scottsville, 3209, South Africa
[2]UMR 017 IRD- Cirad Trypanosomes, Campus international de Baillarguet, 34398 Montpellier Cedex 5, France

ABSTRACT

Heat shock protein 70 (HSP70) family proteins are intracellular chaperones involved in the folding of denatured proteins. In addition, they have been found to be potent stimulators of innate and adaptive immune responses. A protein belonging to the HSP70 family had been identified in *T. congolense*. It was shown to be an immunodominant antigen homologous to mammalian immunoglobulin-heavy-chain binding protein (BiP). The current study aimed at elucidating the adjuvanticity of *T. congolense* BiP using C2, the recombinantly expressed catalytic domain of congopain, as a model antigen. Full-length BiP or truncated BiP devoid of the 10 kDa C-terminal domain was genetically fused to proC2 or C2. The BiP-C2 chimeras were heterologously expressed in *P. pastoris* and *E. coli*. SDS-PAGE analysis of *P. pastoris* expressed chimeras revealed poor expression levels. The quantity and purity was not sufficient for conducting immunisation experiments in mice. On the other hand, bacteria overexpressed the BiP-C2 chimeras as fusion proteins with maltose binding protein. The chimeras were expressed as insoluble inclusion bodies that required solubilisation with 8 M urea, renaturation and purification. Although the chimeras were successfully renatured, purification proved to be a cumbersome task. Hence, the BiP-C2 chimeras need to be recloned into an expression system that would combine prolific expression and easy purification of the chimeras to allow testing of their adjuvant potential in a murine model challenged with *T. congolense*. Furthermore, the capacity of the produced antibodies to inhibit the activity of congopain *in vitro* will be assessed.

Keywords: heat shock protein 70; BiP; chimeras; solubilisation; renaturation.

4.1 INTRODUCTION

Bovine trypanosomosis (nagana) is a parasitic disease caused by tsetse-transmitted *Trypanosoma congolense*, *T. vivax* and *T. b. brucei* (Stevens and Brisse, 2004). The disease is characterised by anaemia, loss of productivity and is often fatal if untreated, making it the major constraint to livestock production in tropical Africa (Eisler *et al*., 2004). Prevention and treatment of trypanosomosis relies heavily on vector control and the use of chemotherapeutic and prophylactic agents (Allsopp and Hursey, 2004). Large scale vector eradication is improbable; hence, for the time being treatment of cattle still relies on trypanocidal drugs and other measures of integrated control (Holmes *et al*., 2004). However, emergence of drug resistant parasites has been reported in Africa (Anene *et al*., 2001; Geerts *et al*., 2001). Consequently, vaccination has been considered.

Conventional vaccines based on targeting parasites' variable or invariable antigens such as variable surface glycoprotein (VSG), tubulin (Lubega *et al*., 2002; Li *et al*., 2007) and flagellar pocket antigens (Mkunza *et al*., 1995) have proven elusive. Hence, a new "anti-disease" vaccine approach has been devised based on mounting an immune response against pathogenic factors released by the parasites during infection (Playfair *et al*., 1990; Playfair *et al*., 1991; Authié, 1994). Congopain, the major cysteine protease of *T. congolense*, has been shown to act as a pathogenic factor in the disease process (Authié *et al*., 2001). Inhibition of congopain activity with antibodies has been associated with resistance to the disease, a phenomenon known as trypanotolerance (Authié, 1994). For these reasons, congopain is an attractive target for an anti-disease vaccine for cattle trypanosomosis.

The functionally active catalytic domain of congopain (C2) had been heterologously expressed in the baculovirus expression system and used in immunisation trials in cattle to assess its role in the disease process (Authié *et al*., 2001; Boulangé *et al*., 2001). The antigen was co-administered with RWL proprietary adjuvant from SmithKline-Beecham to boost the immune response. Immunised cattle showed features similar to those of trypano-tolerant cattle, i.e. cattle that have an inherited genetic ability to control the disease upon infection with *T. congolense* (Paling *et al*., 1991; Williams *et al*., 1991). These animals limited anaemia and produced a prominent IgG response against the antigen (Authié *et al*., 2001).

Initially, vaccination was conducted with live or attenuated pathogens that carried all the machinery necessary to stimulate a strong immune response (Storni *et al*., 2005). However, major drawbacks were encountered with the use of live or attenuated vaccines (Wilson-Welder *et al*., 2008). These are associated with the capability of the attenuated vaccines to undergo recombination and the threat of them reverting to the virulent forms. Molecular biology techniques have been applied to address this major problem. Now, subunit vaccines have been developed based on targeting a particular antigen of an organism (Wilson-Welder *et al*., 2008). Unlike live vaccines, subunit vaccines lack strong immunogenicity and required frequent boosting (Wilson-Welder *et al*., 2008). These vaccines were co-administered with adjuvants to enhance the immune response (Singh and O'Hagan, 2003). Adjuvants based on live or attenuated pathogens suspended in oil-water emulsion like Freund's complete adjuvant have been shown to be effective immune boosters in livestock vaccines, however, their toxicity towards the host has limited their use (Singh and O'Hagan, 2003; Wilson-Welder *et al*., 2008).

In this context, the development of safer and more effective adjuvants is necessary. Heat shock proteins (HSPs), more specifically the HSP70 family proteins, have been identified as potential stimulators of the immune response against cancers and infectious agents (Rico *et al*., 1998; Planelles *et al*., 2002; Ge *et al*., 2006). Ge *et al*. (2006) conducted an immunisation trial in mice comparing the immune response induced by immunisation with *P. pastoris* expressed Japanese encephalitis virus E-protein genetically fused with mycobacterial HSP70 and recombinant E-protein co-administered with Freund's adjuvant. The immune response stimulated by the Freund's adjuvant was more pronounced than that stimulated by the HSP70 molecule. In addition, antibodies induced with Freund's adjuvant were shown to be more effective in neutralising the growth of the virus *in vitro* (Ge *et al*., 2006). However, the immune response generated by immunisation with E-protein fused to HSP70 was better than that induced with E-protein alone, the HSP70 molecule alone and the E-protein plus HSP70. Although the immune response induced by the Freund's adjuvant was better, the low cost and low toxicity of the HSP70 molecule makes it an attractive adjuvant (Ge *et al*., 2006).

A 69 kDa antigen was identified in *T. congolense* lysates using sera from trypano-tolerant and trypano-susceptible cattle in a western blot (Boulangé and Authié, 1994). Cloning of the 69 kDa antigen open reading frame revealed that the protein belongs to the HSP70 family, with most

identity to mammalian immunoglobulin heavy-chain binding protein (BiP) (Boulangé and Authié, 1994). Hence, the antigen was named *T. congolense* BiP (Boulangé and Authié, 1994). The cDNA of *T. congolense* BiP is 2.35 kb and encodes 653 amino acid residues (Boulangé and Authié, 1994). The protein is composed of a leader peptide, a N-terminal ATPase domain (44 kDa), a protease sensitive site, a peptide binding domain (18 kDa) and a variable, immunogenic C-terminal domain (10 kDa) that harbours the trypanosome specific MDDL retention signal (Boulangé and Authié, 1994).

Previous studies showed that immunisation of animals with C2 can lead to a partially protective immune response against a *T. congolense* infection (Authié *et al*., 2001). The partially protective immune response is linked to the poor antigenicity of congopain and the type of antigen delivery system used. The protective epitopes of congopain have been shown to be dimer-associated (Boulangé *et al*., submitted). Hence, the objective of the present study was to improve the delivery of C2 to the immune cells while maintaining the dimer conformation of the protease. The capacity of *T. congolense* BiP to deliver and maintain the antigen in the appropriate conformation was investigated. BiP was genetically fused to C2 using molecular biology techniques and the chimeras were expressed in *P. pastoris* and *E. coli* expression systems. High level expression of C2 had been achieved in *P. pastoris* (Boulangé *et al*., submitted). However, bacteria expressed C2 as insoluble inclusion bodies that were difficult to renature (Boulangé *et al*., 2001). Although BiP was expressed as a soluble fusion protein with maltose binding protein in *E. coli* (Boulangé *et al*., 2002), attempts to express it in *P. pastoris* failed (Dr A. Boulangé. University of KwaZulu-Natal, personal communication). Due to these reasons, expression of the BiP-C2 chimeras was attempted in both systems.

Two constructs were designed for expression in each expression system. The constructs contained proC2 or C2 cloned 5' or 3' to either full-length BiP (R69) or truncated BiP lacking the 10 kDa C-terminal domain (R60). The proregion of cysteine proteases plays a crucial role in proper folding of the protease and inhibits proteolytic activity of the enzyme (Lalmanach *et al*., 1998; Sanderson *et al*., 2000). Consequently, it was included in most constructs. The C-terminal domain (10 kDa) of BiP is highly variable and antigenic (Boulangé and Authié, 1994). Antibodies from infected animals were shown to specifically recognise epitopes located in the C-terminal domain of BiP (Boulangé *et al*., 2002). To prevent misdirection of the immune

response, the C-terminal domain of BiP was omitted in most constructs. Here we report on the cloning and expression of the BiP-C2 chimeras in eukaryotic and bacterial expression systems.

4.2 MATERIALS AND METHODS

4.2.1 Materials

Molecular Biology: EcoRI, NotI, XbaI, shrimp alkaline phosphatase (SAP), T4 DNA ligase, dNTP mix, X-gal, MassRuler™ DNA ladder mix, Middle Range™ DNA ladder mix, GeneJet™ plasmid miniprep kit and TransformAid™ bacterial transformation kit were purchased from Fermentas (Vilnius, Lithuania). *Taq* DNA polymerase was obtained from Solis Biodyne (Tartu, Estonia). DNA Clean and Concentrator kit was obtained from ZymoResearch™ (Orange, CA, USA). The E. Z. N. A® gel extraction kit was from peQLab (Erlangen, Germany). Competent *E. coli* (JM 109) and XhoI were purchased from New England Biolabs (Ipswitch, MA, USA). The pGEM-T® vector was purchased from Promega (Madison, WI, USA). The *P. pastoris* yeast strain GS 115 and pPic9 expression vector were obtained from Invitrogen (Carlsband, CA, USA). The BioRad GenePulser™ electroporator was purchased from BioRad (Hercules, CA, USA).

Antibodies: The anti-congopain bovine polyclonal antibody was raised against the catalytic domain of CP1 (C1) that was expressed in bacteria (Boulangé *et al*., 2001). The rabbit anti-BiP polyclonal antibody raised against full-length BiP expressed in bacteria was obtained from Dr A. Boulangé (University of KwaZulu-Natal, Pietermaritzburg).

4.2.2 Methods

4.2.2.1 Cloning of BiP-C2 chimeras for expression in *Pichia pastoris*

Amplification of truncated BiP

The catalytic domain of congopain (C2) had been cloned into a pPic9 expression vector (Fig. 4.1) with its propeptide using EcoRI and NotI cloning sites (Boulangé *et al*., submitted). Immunoglobulin heavy-chain binding protein (BiP) of *T. congolense* had been cloned into the pMal-cRI vector (Boulangé *et al*., 2002). The glycerol stocks of *E. coli* containing pPic9-proC2 and pMal-R69 were streaked onto 2× YT plates [1.6% (w/v) tryptone, 1% (w/v) yeast extract,

0.5% (w/v) NaCl, 15 g/l bacteriological agar] containing ampicillin (50 µg/ml). Liquid 2× YT containing ampicillin (50 µg/ml) was inoculated with a single colony and the culture was incubated overnight at 37ºC. Plasmid DNA was isolated using the GeneJet™ plasmid miniprep kit according to the manufacturer's instructions. The quality of isolated plasmid DNA was assessed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer (40 mM Tris-Cl buffer, 0.1 mM EDTA, 20 mM acetic acid) run at 80 V for 30 min.

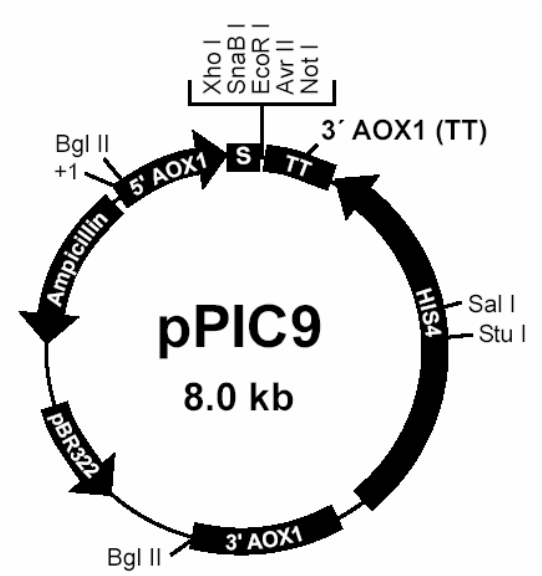

Figure 4.1. Map of pPic9 expression vector. The pPic9 expression vector contains the ColE1 origin of replication (pBR322), 5' AOX promoter (5' AOX), a multiple cloning site, α-factor secretion signal (S), *c-myc* epitope (TT), ampicillin resistance gene, and his4 gene for the selection by complementation in his4 yeast strains.

Enzymes cutting or not cutting the full-length BiP and proC2 open reading frames were identified using restriction maps (Fig. 4.2A and B). Therefore, EcoRI and NotI were chosen as restriction enzymes to be used in the cloning process. The primers were designed to contain either EcoRI or NotI sites at both ends to allow for 5' or 3' cloning of the PCR product into the pPic9-proC2 expression vector. Truncated BiP devoid of the variable and immunogenic C-terminal domain was amplified by polymerase chain reaction (PCR) using specific primers from the pMal-R69 template DNA (Table 4.1). Amplification was carried out by denaturing plasmid DNA at 95ºC for 5 min followed by 25 cycles of denaturation at 95ºC for 1 min, primer annealing at 60ºC for 1 min, elongation at 72ºC for 2 min and finally elongation at 72ºC for 7 min.

Table 4.1. Primers used for amplification of R60 from pMal-R69 plasmid DNA. The amplicon was ligated into pGEM-T vector before being transformed into competent *E. coli* cells. The restriction sites are highlighted in bold.

Primer	Sequence
R60EcoRIFw	5'-GCC**GAATTC**GCGCCCGAGAGCGGCGGGAAG-3'
R60EcoRIRv	5'-GAG**GAATTC**CCGCGCCTCCACACGCTCCCC-3'
R60NotIFw	5'-CGC**GCGGCCGC**GCCCGAGAGCGGCGGGAAG-3'
R60NotIRv	5'-GAG**GCGGCCGC**CCGCGCCTCCACACGCTCC-3'
SP6	5'-CATATGATTTAGGTGACACTATA-3'
T7	5'-TAATACTGACTCATATAGGG-3'

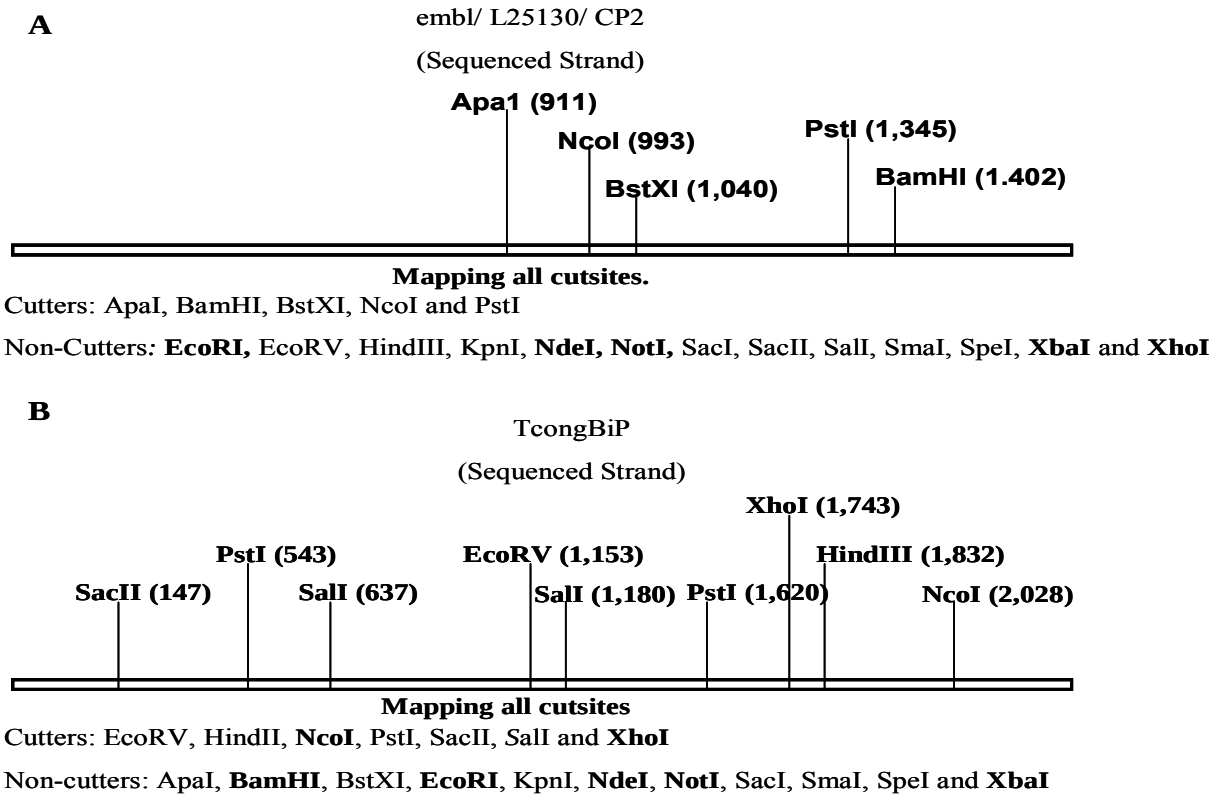

Figure 4.2. Restriction maps of CP2 (A) and BiP (B) showing cutting and not cutting enzymes. The enzymes highlighted in bold are those used in the cloning process.

Subcloning of amplicons into pGEM-T vector

The truncated PCR product (1.65 kb) was purified using the peQLab gel extraction kit. The purified amplicons were ligated into the pGEM-T vector (Fig. 4.3). The ligation mix (3 µl) was transformed into competent *E. coli* (JM 109) cells using the TransformAid™ bacterial transformation kit. The transformed cells (50 µl) were plated onto 2× YT plates containing

ampicillin (50 µg/ml), X-gal (20 µg/ml) and IPTG (10 µg/ml) to allow for blue and white colony screening. The plates were incubated overnight at 37ºC.

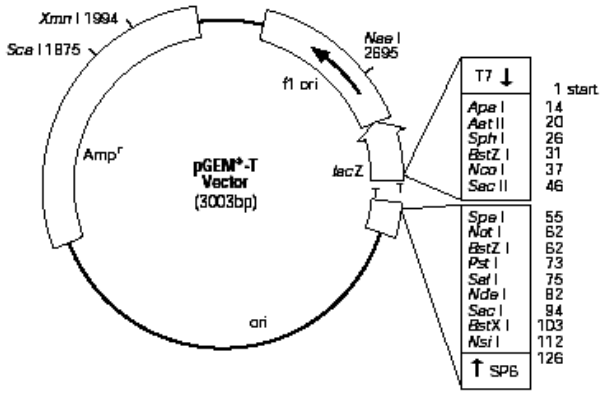

Figure 4.3. Map of the pGEM-T® vector (Promega). The vector contains a multiple cloning site, ampicillin resistance gene (Amp^r) for selection of recombinants, f1 origin of replication (f1 ori) and the lacZ gene for blue and white colony screening.

White colonies were screened by colony PCR using the vector-specific T7 and SP6 primers (Table 4.1). The amplification conditions applied for PCR screening are shown in Table 4.2. PCR-positive colonies were grown in selective media overnight at 37ºC. Plasmid DNA was isolated using the GeneJet™ plasmid miniprep kit.

Subcloning of the inserts into pPic9-proC2 expression vector

Recombinant pGEM-T vector containing R60 was digested with either EcoRI or NotI and the drop out product was purified using the E. Z. N. A.® gel extraction kit. pPic9-proC2 was also digested with either EcoRI or NotI in their respective enzyme buffers. The linearised vector (pPic9-proC2) was dephosphorylated with 1 U of SAP and concentrated using the ZymoResearch™ DNA Clean and Concentrator kit. The insert was ligated into linear pPic9-proC2 vector to generate pPic9-proC2-R60 and pPic9-R60-proC2 (Fig. 4.4). The ligation mix was used to transform competent *E. coli* cells and the cells were plated onto 2× YT plates containing ampicillin (50 µg/ml). Transformants were screened with AOX primers (Table 4.3). The PCR conditions used during colony PCR screening of transformants are shown in Table 4.2. Positive colonies were grown overnight in selective media and plasmid DNA isolated the next morning.

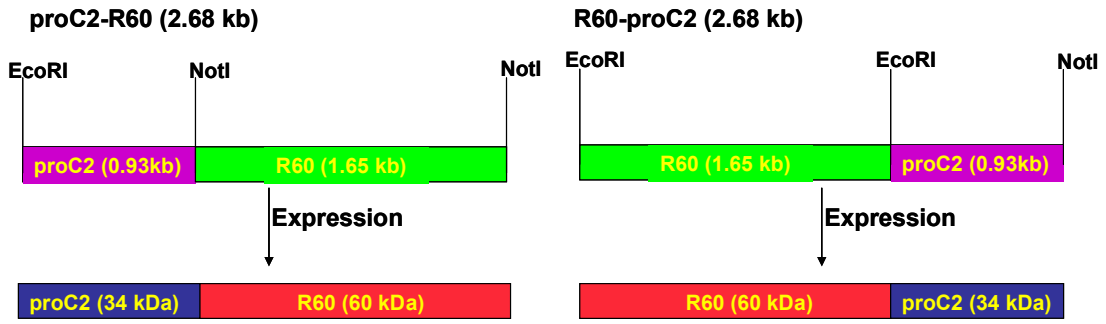

Figure 4.4. Schematic representation of the cloning process for the generation of R60-proC2 and proC2-R60 constructs in the pPic9 expression vector. The truncated BiP (R60) amplicon containing EcoRI sites was cloned 5' to proC2 ORF and the amplicon containing NotI sites was cloned 3' to proC2 ORF in pPic9-proC2 vector.

Table 4.2. Colony PCR conditions for the different experiments. The master mix composed of PCR reaction buffer (1×), Taq DNA polymerase (0.1 U), primers (1 µM), dNTPs (0.2 mM), and subsequently dispensed into PCR tubes. All the PCR reactions were cycled 25 times.

	Colony PCR (*E. coli*)		Colony PCR (*P. pastoris*)
	T7/SP6	AOX	AOX
Denaturation	95ºC; 30 s	95ºC; 1 min	95ºC; 1 min
Annealing	45ºC; 30 s	50ºC; 30 s	50ºC; 30 s
Elongation	72ºC; 1 min	72ºC; 3 min	72ºC; 3 min
Reaction volume (µl)	20	20	100

Transformation of Pichia pastoris *strain GS 115 cells*

The isolated plasmid DNA (pPic9-proC2-R60 and pPic9-R60-proC2) was linearised with SacI restriction enzyme for efficient recombination into the *P. pastoris* genome. The linearised plasmid DNA was concentrated using the ZymoResearch™ DNA Clean and Concentrator kit. *P. pastoris* strain GS 115 yeast cells were made competent by using the protocol described by Wu and Letchworth (2004). Briefly, cells were incubated in lithium acetate/DTT buffer (100 mM lithium acetate buffer pH 7.5, 10 mM DTT, 0.6 mM sorbitol) for 30 min, followed by several washes of the cell pellet in ice cold 1 M sorbitol. At the end, the cells (10^{10} cells/ml) were resuspended in ice cold 1 M sorbitol and frozen at -80ºC until required for use.

Competent *P. pastoris* strain GS 115 cells (200 µl) were incubated with SacI-linearised plasmid DNA (2.5 µl) on ice for 2 min. Plasmid DNA was electroporated into yeast cells using a

BioRad® electroporator (Hercules, CA, USA) set at 1.5 kV, 25 µF and 186 Ω in 2 mm gap cuvettes. Immediately after electroporation, the cells were diluted with ice cold 1 M sorbitol (1 ml). The cells were spread on minimal media plates [1.34% (w/v) yeast nitrogen base without amino acids (YNB), 0.0004% (w/v) biotin, 2% (w/v) glucose, 15 g/l bacteriological agar] containing ampicillin (50 µg/ml). The plates were incubates at 30ºC until the colonies were visible (generally 2-3 days).

Colony PCR using AOX primers (Table 4.3) was performed on *P. pastoris* transformants using the adapted Ayra-Pardo *et al*. (1998) protocol as described in Table 4.2. Positive colonies were grown in liquid YPD media [1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v dextrose] containing antibiotics (generally ampicillin, 50 µg/ml).

BiP-C2 chimeras expression in Pichia pastoris

Liquid YPD (100 ml) containing antibiotics (generally ampicillin, 50 µg/ml) was inoculated with a single colony of recombinant *P. pastoris* and grown at 30ºC in a baffled flask with agitation for two to three days. This culture was used to inoculate 1 L BMGY media [1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate buffer, pH 6.5] and incubated for a further two days to increase biomass. The cells were pelleted (6000×g, 10 min, RT) and resuspended in an equal volume of buffered minimal medium [BMM, 100 mM potassium phosphate buffer, pH 6.5, 1.34% (w/v) YNB, 0.0004% (w/v) biotin, 5% (v/v) methanol]. The baffled flasks were covered with cheesecloth to facilitate aeration of the cultures. The cultures were supplemented with methanol [0.5% (v/v) methanol] daily for the entire duration of the expression period to induce protein expression.

4.2.2.2 Cloning of BiP-C2 chimeras for bacterial expression

Amplification and subcloning of inserts into pGEM-T vector

T. congolense full-length BiP (R69) had been cloned into pMal-cRI expression vector (Boulangé *et al*., 2002). Thus it was convenient, simple and cost effective to clone congopain catalytic domain (C2) into the pMal-R69 expression vector. The truncated catalytic domain of congopain (C2) and C2 containing the proregion (proC2) were amplified with specific primers (Table 4.3) from the pPic-proC2 vector used as a template. PCR was performed by denaturing plasmid DNA at 95ºC for 5 min, followed by 25 cycles of denaturation at 95ºC for 45 s, primer annealing at

50ºC for 30 s, elongation at 72ºC for 1 min and final elongation at 72ºC for 7 min. The proC2 amplicon contained EcoRI cloning sites on both ends whereas the C2 amplicon contained one XhoI site on the 5' end and one XbaI site plus the engineered stop codon on the 3' end. The cloning sites determined whether congopain was to be cloned upstream or downstream of the R69 open reading frame in the pMal-R69 expression vector. Two constructs, proC2-R69 (full-length BiP) and R60-C2 (truncated BiP) were generated (Fig. 4.5)

Table 4.3. Primers used to amplify proC2 and C2 from pPic9-proC2 template DNA. Plasmid DNA was isolated using the GeneJet™ plasmid miniprep kit from an overnight culture of cells. The restriction sites are underlined.

Primer	Sequence
proCDEcoRIFw	5'-GCCGAATTCGCGTGCTTTGTTCCCGTGGCG-3'
CDEcoRIRv	5'-GTGGAATTCCGGAGTCGGGGGTGGAGGCGG-3'
CDXhoIFw	5'-CCACCCTCGAGGCACCTGAGGCAGTTGACTGGCG-3
CDXbaIRv	5'-CTGTCTAGATTACGGAGTCGGGGGTGGAGGCGG-3'
AOXFw	5'-ACTGGTTCCAATTGACAAGC-3'
AOXRv	5'-GCAAATGGCATTCTGACATCC-3'

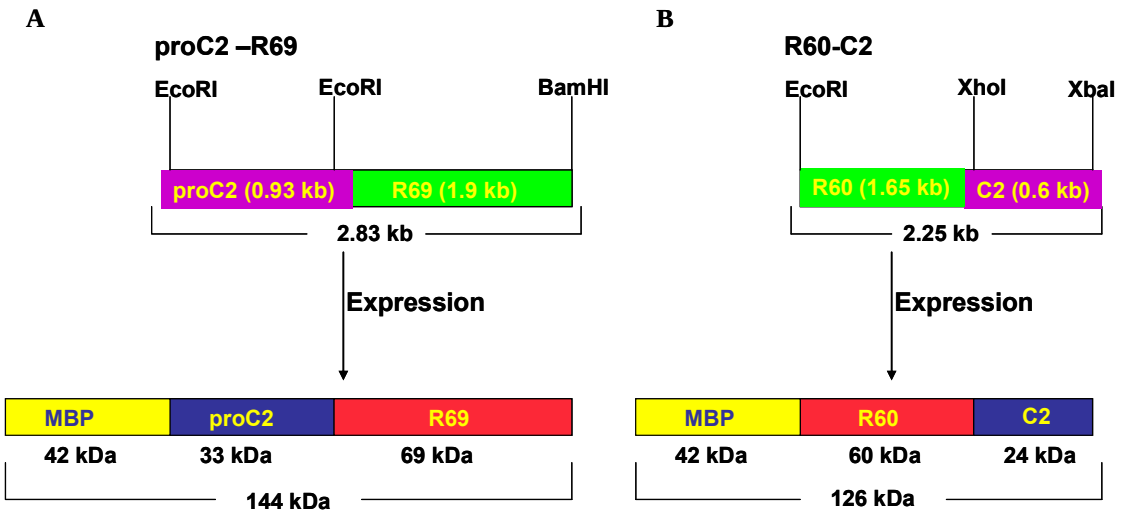

Figure 4.5. Design of proC2-R69 (A) and R60-C2 (B) for expression in bacteria. The constructs were generated using different cloning sites. ProC2-R69 was generated using EcoRI cloning sites and R60-C2 was generated using XhoI and XbaI cloning sites. The MBP-proC2-R69 fusion chimera is expected to be 144 kDa and MBP-R60-C2 is expected to be 126 kDa.

The amplicons were purified and ligated into the pGEM-T vector. Recombinant colonies were screened by colony PCR using SP6 and T7 vector specific primers. Plasmid DNA was isolated

from positive colonies and digested with either EcoRI or XhoI/XbaI to drop out proC2 and C2 respectively. The inserts were purified using the peQLab gel extraction kit.

Subcloning of inserts into pMal expression vector

Recombinant pMal-R69 was digested with either EcoRI or XhoI/XbaI to generate linear plasmid DNA. Linear plasmid DNA was dephosphorylated with SAP and concentrated using the ZymoResearch® DNA Clean and Concentrator kit. The proC2 (EcoRI/EcoRI) and C2 (XhoI/XbaI) inserts were ligated into linear pMal-R69 plasmid DNA. The ligation mix (3 µl) was transformed into competent *E. coli* (JM 109) cells (50 µl) using the TransformAid™ bacterial transformation kit. The transformed cells were plated onto 2× YT plates containing ampicillin (50 µg/ml). Colony PCR was performed with pMal and insert specific primer pairs on transformants to screen for recombinant colonies. Positive colonies were used to make glycerol stocks.

Expression of the BiP-C2 chimeras in bacteria.

Expression of the BiP-C2 chimeras was carried out as described in the pMal manual (NEB pMal instruction manual). An overnight culture was prepared by inoculating terrific broth, supplemented with glucose, with a single colony of bacteria containing the respective recombinant plasmid DNA. Glucose represses the expression of amylase from the maltose genes, thus, prevents degradation of amylose in the resin in subsequent purification steps (NEB pMal instruction manual). In addition, glucose exerts additional repression of the lac Z promoter in the absence of the inducer IPTG (NEB pMal instruction manual). This pre-culture was transferred into fresh terrific broth (500 ml) containing ampicillin (50 µg/ml) and incubated at 37ºC in a shaking incubator until the culture reached an OD_{600} of 0.6. The culture was induced with 0.1 mM IPTG and incubated at 37ºC for 4 h for protein expression in an orbital shaking incubator. The cells were collected by centrifugation (5000×g, 10 min, RT) and resuspended in TE buffer (100 mM Tris-Cl buffer, pH 7.4, 10 mM EDTA). Lysozyme was added to the lysate to facilitate degradation of the bacterial cell wall. The lysate was frozen overnight at -80ºC.

4.2.2.3 Solubilisation and renaturation of bacterial-expressed chimeras

The inclusion bodies were solubilised and renatured using the protocol described by Sijwali *et al*., (2001). Firstly, the inclusion bodies were purified from the crude lysate by washing the pellet

with 2 M urea, 20 mM Tris-Cl, 2.5% Triton X-100, pH 8, followed by subsequent washes with 20% (w/v) sucrose, 20 mM Tris-Cl, pH 8. The inclusion bodies were pelleted (17000×g, 30 min, 4ºC) and solubilised with 8 M urea. The crude insoluble material was separated by centrifugation (27000×g, 30 min, 4ºC) from the soluble protein.

The fusion chimeras (MBP-R69-C2 and MBP-proC2-R69) were reduced with 10 mM DTT to ensure complete activation of the cysteine protease. Renaturation was performed by drop-wise dilution of the fusion chimera with refolding buffer (100 mM Tris-Cl buffer, pH 9.0, 1 mM EDTA, 30% (v/v) glycerol, 250 mM L-arginine, 1 mM GSH, 1 mM GSSG) at 4ºC. The fusion chimera was diluted a 100 fold in refolding buffer. The renatured chimera was concentrated by ultrafiltration using an Amicon PM 10 filter membrane (10 kDa cut-off).

4.2.2.4 Purification of BiP-C2 chimeras by three phase partitioning (TPP)

Three phase partitioning (Pike and Dennison, 1989) was used to purify the BiP-C2 chimeras from expression supernatants. The supernatant was filtered through Whatman No 1 filter paper. The pH of the supernatant was lowered to pH 4.2 with phosphoric acid for autocatalytic processing of the cysteine protease to the mature form. The fusion chimeras were precipitated with 40% ammonium sulfate (40% of total volume) in the presence of tertiary butanol (1/3 final volume). The salt was dissolved by gentle stirring with a magnetic stirrer. The precipitated protein mixture was centrifuged (6000×g, 10 min, 4ºC) using a pre-chilled swing-out rotor. The protein layer at the interface between the aqueous and tertiary butanol phases was collected and dialysed against PBS, pH 7.2 using a 10 kDa cut-off dialysis membrane to remove traces of salt and tertiary butanol. The diluted protein was concentrated using polyethylene glycol M_r 20 000 until 1-2 ml remained in the dialysis bag.

4.2.2.5 Purification of BiP-C2 chimeras by amylose affinity chromatography

Maltose binding protein (MBP) affinity for amylose was exploited to purify the fusion chimeras from the crude bacterial lysate. Amylose resin (NEB pMal instruction manual) was packed into a column (10 ml). The column was washed with 8 column volumes of column buffer (20 mM Tris-Cl buffer, pH 7.4, 0.2 M NaCl, 10 mM 2-mercaptoethanol) before the renatured protein sample was loaded onto the column. The column was washed with 2 column volumes of column buffer.

The fusion chimera was eluted with 10 mM maltose in column buffer and protein elution was monitored by measuring A_{280nm} of each 1.5 ml fraction. The fractions containing proteins were pooled and concentrated by ultrafiltration using an Amicon PM 10 filter membrane. The resin was regenerated by sequential washes with distilled water, 0.1% (w/v) SDS, distilled water and column buffer.

4.2.2.6 Protein Visualisation

Laemmli SDS-PAGE

Proteins were visualised by discontinuous SDS-PAGE (Laemmli, 1970). The proteins were resolved through a large pore size stacking gel (pH 6.8) and a small pore size running gel (pH 8.8). For accurate determination of protein molecular weight, the fusion chimeras were treated with reducing treatment buffer [125 mM Tris-Cl buffer, pH 6.8, 4% (w/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercaptoethanol] in a 1:1 ratio. SDS and 2-mercaptoethanol present in the treatment buffer facilitate denaturation and reduction of the proteins and give the protein a net negative charge. Consequently, when an electric field is applied, the proteins separate based solely on molecular weight. For our purpose, proteins were run at 20 mA/gel for 2 h in tank buffer [250 mM Tris-Cl buffer, 192 mM glycine, 0.1% (w/v) SDS, pH 8.3]. Coomassie blue R-250 [0.125 (w/v) Coomassie blue R-250, 50% (v/v) methanol, 10% (v/v) glacial acetic acid] was used to stain proteins.

Silver staining

Silver staining was performed when more sensitive staining was required (Blum *et al*., 1987). The gel was immersed in fixing solution [50% (v/v) methanol, 12% (v/v) acetic acid, 0.5% (v/v) formaldehyde] for one hour and subsequently washed (3 × 5 min) in 50% (v/v) ethanol. The gel was incubated in pre-treatment solution (4 mg $Na_2S_2O_3.5H_2O$ dissolved in 200 ml of water) for 1 min, rinsed with distilled water (3 × 5 min) and soaked in impregnation solution [0.2% (w/v) $AgNO_3$, 0.28% (v/v) formaldehyde] for 20 min. The gel was rinsed with distilled water (2 × 20 s) and incubated in developing solution [0.019% (v/v) formaldehyde, 0.004% $Na_2S_2O_3.5H2O$, 60 g/l Na_2CO_3] until the first bands became visible. The gel was then placed in distilled water to allow for complete development of the bands. When the desired intensity of the bands was obtained, the gel was immersed in stopping solution [50% (v/v) methanol, 12% (v/v) acetic acid].

Western blotting

The proteins separated by Laemmli SDS-PAGE were electroblotted to a nitrocellulose membrane using a semi-dry blotter, run at 7 V overnight (Towbin *et al*., 1979). The unoccupied sites on the membrane were blocked with low fat milk [5% (w/v) in TBS]. The membrane was washed with TBS (3 × 5 min) and probed with bovine anti-congopain and rabbit anti-full-length BiP antibodies diluted in 0.5% (w/v) BSA in TBS for 2 h. The membrane was washed with TBS (3 × 5 min) and subsequently incubated with rabbit anti-bovine IgG and goat anti-rabbit IgG conjugated to horse radish peroxidase (HRPO) for 1 hour. Finally, the membrane was washed with TBS (3 × 5 min) and incubated in substrate solution [0.06% (w/v) 4-chloro-1-naphthol, 10% (v/v) methanol, 0.0015% (v/v) H_2O_2 in TBS] until bands appeared.

4.3 RESULTS

4.3.1 Cloning, expression and purification of *P. pastoris* expressed BiP-C2 chimeras

Truncated BiP (R60) lacking the C-terminal domain was amplified from pMal-R69 DNA using specific primers. The two amplicons were observed as prominent bands of 1.65 kb (Fig. 4.6. lanes 2 and 3). This is consistent with the predicted size of the amplicons. The products were excised from the agarose gel and purified using the peQLab gel extraction kit.

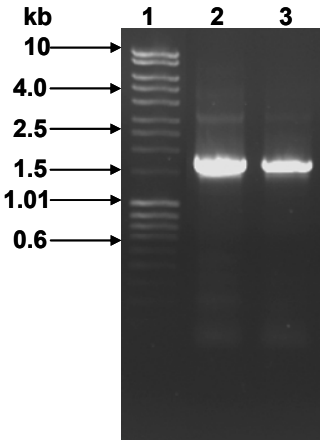

Figure 4.6. Amplification of truncated BiP (R60) from the pMal-R69 template DNA. The PCR product was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, MassRuler™ DNA ladder mix; lane 2, R60NotI amplicon and lane 3, R60EcoRI amplicon.

The purified PCR product (R60) was ligated into the pGEM-T vector and the ligation mix was used to transform competent *E. coli* cells. Colony PCR screening with T7/SP6 primers was

attempted to identify recombinant colonies. However, no recombinants were observed using this method. Two colonies per transformation were grown in selective media overnight and plasmid DNA isolated. The isolated plasmid DNA appeared as two bands above 10 kb (Fig. 4.7A, lanes 2-5). This means that the plasmid exists in two conformations. Miniprep DNA was digested with either EcoRI or NotI to confirm the presence of the inserted fragment (Fig. 4.7B). All the plasmids were recombinant as indicated by the drop out of a 1.65 kb product after digestion. Digestion of plasmid DNA with NotI was incomplete as indicated by the presence of bands slightly below and above 5 kb (Fig. 4.7B, lanes 4 and 5). The dropped out insert was purified using the peQLab gel extraction kit.

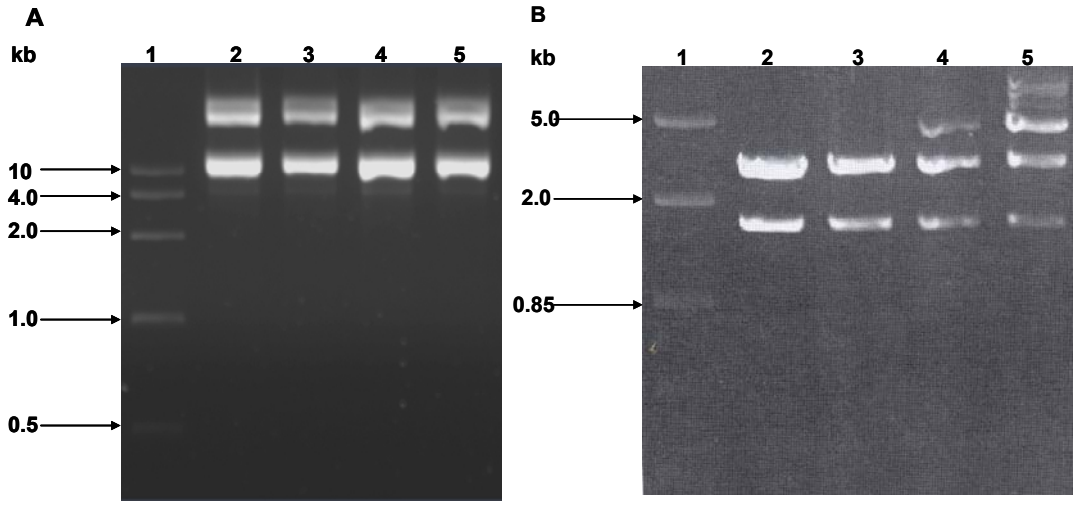

Figure 4.7. Screening of pGEM-T vector recombinants. Panel A, isolation of recombinant pGEM-T-R60EcoRI or pGEM-T-R60NotI plasmid DNA. Plasmid DNA was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, FastDNA™ Ruler High Range; lane 2, pGEM-T-R60EcoRI clone 1; lane 3, pGEM-T-R60EcoRI clone 2; lane 4, pGEM-T-R60NotI clone 1 and lane 5, pGEM-T-R60NotI clone 2. Panel B, restriction digest of pGEM-T vector recombinants with either EcoRI or NotI. Lane 1, FastDNA™ Ruler High Range; lane 2, EcoRI restricted pGEM-T-R60EcoRI clone 1; lane 3, EcoRI restricted pGEM-T-R60EcoRI clone 2; lane 4, NotI restricted pGEM-T-R60NotI clone 1 and lane 5, NotI restricted pGEM-T-R60NotI clone 2.

The catalytic domain of congopain (C2) had been cloned into pPic9 expression vector (Boulangé *et al*., submitted). The pPic9-proC2 expression vector was prepared by miniprep and used to clone the truncated BiP open reading frame upstream or downstream from the proC2 open reading frame. Therefore, pPic9-proC2 was digested with either EcoRI or NotI. Linear vector DNA was observed as a single band around 9 kb (Fig. 4.8, lanes 2 and 3). Linear plasmid DNA was dephosphorylated with shrimp alkaline phosphatase to prevent recircularisation of the vector during ligation.

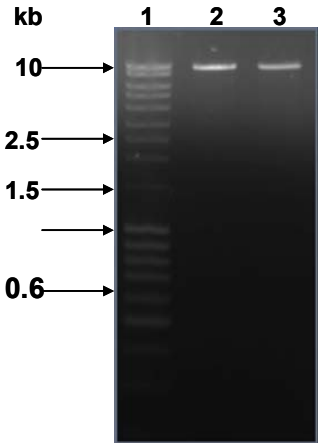

Figure 4.8. Purification of the pPic9-proC2 expression vector DNA after linearisation with either EcoRI or NotI. The digestion mixture was analysed by agarose gel electrophoresis. Lane 1, MassRuler™ DNA ladder mix; lane 2, EcoRI restricted pPic9-proC2 and lane 3, NotI restricted pPic9-proC2.

Purified R60EcoRI and R60NotI were ligated into the pPic9-proC2 expression vector. The ligation mix was used to transform competent *E. coli* cells. Colony PCR using three sets of primers was performed to screen for recombinant colonies per cloning experiment. To screen for pPic9-R60-proC2 recombinants, colony PCR was performed using an AOX primer pair, the insert-specific primers (R60EcoRIFw and CDEcoRIRv) and a combination of R60EcoRIFw and AOXRv primers (Fig. 4.9A). Amplification with a combination of insert-specific and vector-specific primers was performed to confirm the ligation of the insert into the vector. Screening of colonies with AOX primers produced a 3 kb PCR product (Fig. 4.9A, lanes 8 and 11). The PCR product (R60-proC2) was bigger than the expected 2.8 kb due to the 200 bp vector sequence amplified by the AOX primers on both ends. Screening of pPic9-R60-proC2 with insert-specific primers produced an expected 2.8 kb product (Fig. 4.9A, lanes 9 and 12). However, a combination of R60EcoRIFw and AOXRv primers amplified a 2.85 kb PCR product (Fig. 4.9A, lanes 10 and 13). Differences in the size of PCR product amplified by insert-specific primers and a combination of R60EcoRIFw and AOXRv primers was expected due to the capacity of the AOX primers to amplify some of the vector sequence. Two positive clones were obtained for pPic9-R60-proC2.

A screening process similar to the above was applied to screen for pPic9-proC2-R60 recombinants. The colonies were screened with the AOX primer pair, insert-specific primers (proCDEcoRIFw and R60NotIRv) and a combination of proCDEcoRIFw and AOXRv primers (Fig. 4.9B). As shown for R60-proC2, screening with AOX primers amplified a 3 kb PCR

product (proC2-R60) in recombinant colonies (Fig. 4.9B, lanes 8 and 11). Screening with insert-specific primers produced an expected 2.8 kb PCR product (Fig. 4.9B, lanes 9 and 12). However, screening with a combination of proCDEcoRIFw and AOXRv primers amplified a 2.85 kb product (Fig. 4.9B, lanes 10 and 13). This product was slightly bigger due to the 200 bp vector sequence amplified by AOX primers. Two positive clones were obtained for pPic9-proC2-R60.

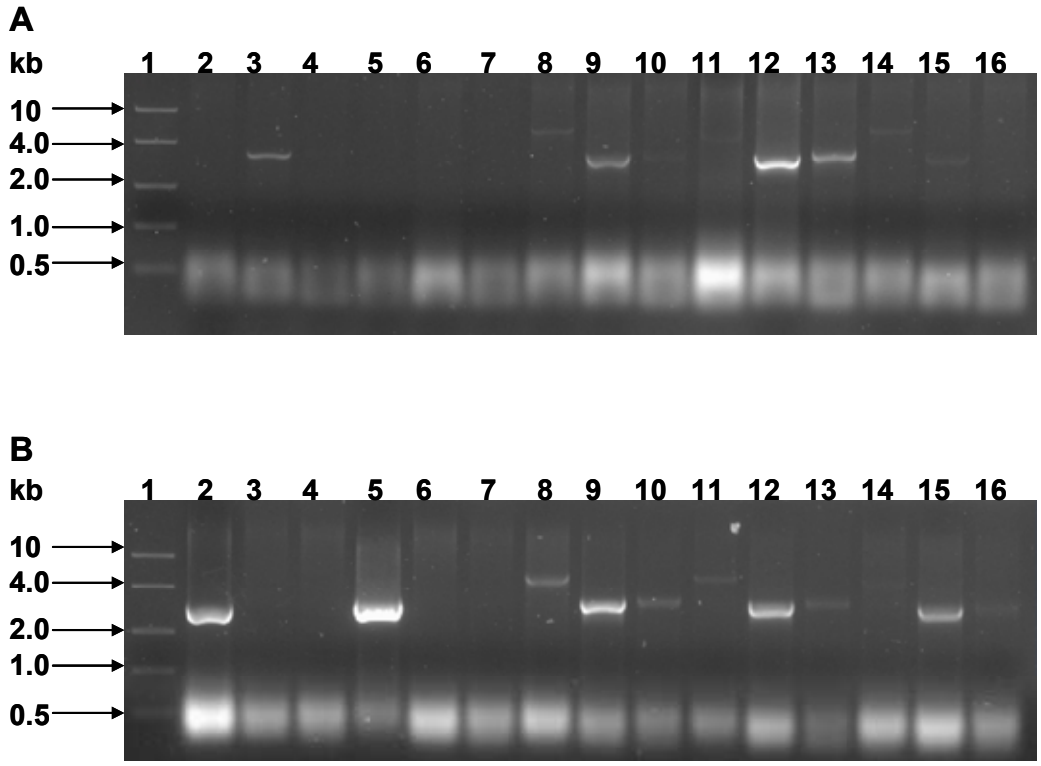

Figure 4.9. Colony PCR screening of pPic9-proC2 recombinants. Panel A, colony PCR screening of pPic9-R60-proC2 recombinant colonies with AOX primers, R60EcoRIFw-CDEcoRIRv specific primers and a combination of R60EcoRIFw and AOXRv primers. Lane 1, FastDNA™ Ruler High Range; lanes 2, 5, 8, 11 and 14, AOX screening of colonies 1-5; lanes 3, 6, 9, 12, and 15, R60EcoRIFw-CDEcoRIRv screening of colonies 1-5 and lanes 4, 7, 10, 13 and 16, R60EcoRIFw-AOXRv screening of colonies 1-5. Panel B, colony PCR screening of pPic9-proC2-R60 recombinant colonies with AOX primers, proCDEcoRIFw-R60NotIRv specific primers and a combination of proCDEcoRIFw and AOXRv primers. PCR products were analysed by agarose gel electrophoresis. Lane 1, FastDNA™ Ruler High Range; lanes 2, 5, 8, 11 and 14, AOX screening of colonies 1-5; lanes 3, 6, 9, 12 and 15, proCDEcoRIFw-R60NotIRv screening of colonies 1-5 and lanes 4, 7, 10, 13 and 16, proCDEcoRIFw-AOXRv screening of colonies 1-5.

Recombinant expression vector DNA was isolated by miniprep and digested with the SacI restriction enzyme (Fig. 4.10). Miniprep plasmid DNA was observed as two bands above 10 kb (Fig. 4.10, lanes 2 and 4). Linear plasmid DNA migrated as a single band slightly above 10 kb (Fig. 4.10, lanes 3 and 5). Linear plasmid DNA migrated faster than miniprep DNA (Fig. 4.10). This is not customary as linear vectors are expected to migrate slower than uncut vectors

(Sambrook and Russell, 2001). Plasmid DNA was linearised and used to transform *P. pastoris* strain GS 115 cells. Colony PCR was performed on transformants to confirm successful integration of recombinant vector DNA into the *P. pastoris* genome.

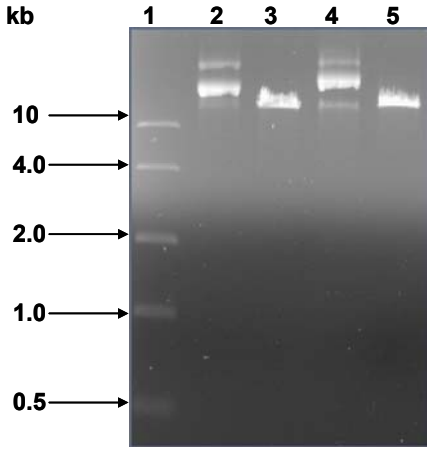

Figure 4.10. Linearisation of recombinant plasmid DNA with the SacI restriction enzyme. The digestion mixture was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, FastDNA™ Ruler High Range; lane 2, uncut pPic9-R60-proC2; lane 3, SacI cut pPic9-R60-proC2; lane 4, uncut pPicp-proC2-R60 and lane 5, SacI cut pPic9-proC2-R60.

Recombinant *P. pastoris* transformants were used to inoculate minimal media for protein expression. Protein expression was induced by daily supplementation of the media with methanol for the entire duration of the expression period. Analysis of the expression of the proteins by SDS-PAGE revealed a poor level of expression by the yeast cells (Fig. 4.11). The bands corresponding to the expected chimeras were not visible on the gel (Fig. 4.11, lanes 2 and 3). Both chimeras (proC2-R60 and R60-proC2) were expected to migrate at 84 kDa on a SDS-PAGE gel. However, the presence of the 27 kDa band corresponding to the processed mature protease (C2) in the proC2-R60 lane suggested some degradation of the chimera (Fig. 4.11, lane 2). The cleavage of proC2 from R60 was unexpected because cleavage was expected to occur at the proregion maturation site (Fig. 4.11).

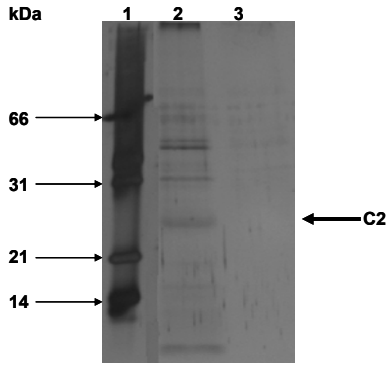

Figure 4.11. Silver staining of reducing SDS-PAGE (12%) gel analysing expression supernatants of the BiP-C2 by *P. pastoris*. The samples were boiled and reduced before loading. Lane 1, BioRad LMWM; lane 2, proC2-R60 expression supernatant and lane 3, R60-proc2 expression supernatant.

Three phase partitioning was used to concentrate the chimeras from the large volumes of expression supernatant. The proteins were precipitated with a kosmotropic ammonium sulfate salt in the presence of tertiary butanol. The predicted 84 kDa band corresponding to the full-length chimeras (proC2-R60 and R60-proC2) became visible after TPP (Fig. 4.12A, lanes 2 and 3). The identity of the band was confirmed by rabbit anti-BiP antibodies that weakly recognised the 84 kDa band (Fig. 4.12B, lanes 2 and 3). The bovine anti-congopain antibodies failed to recognise the chimeras. However, these particular antibodies specifically recognised bands at 27 kDa and 33 kDa corresponding to mature and immature congopain respectively (Fig. 4.12B, lanes 2 to 4). This indicates cleavage of the chimeras into their constituent proteins. The poor yields and low purity of both chimeras did not warrant immunisation experiments in mice.

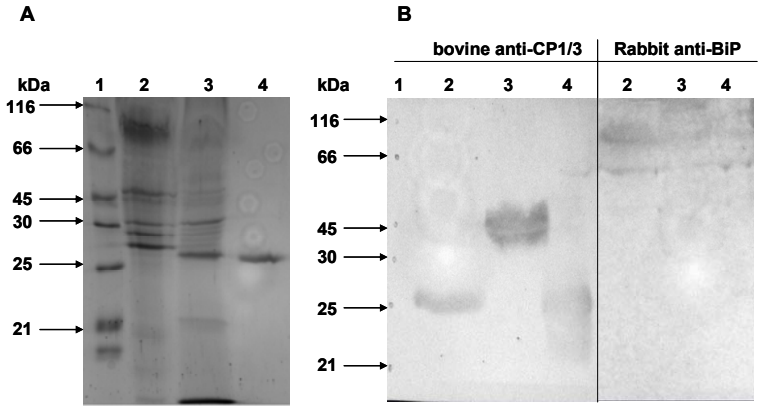

Figure 4.12. Analysis of BiP-C2 chimera purification by reducing SDS-PAGE and western blot. Panel A, reducing SDS-PAGE (12%) gel showing the purification of BiP-C2 chimeras. The BiP-C2 chimeras were boiled and reduced prior to loading. Lane 1, Fermentas LMWM; lane 2, proC2-R60; lane 3, R60-proC2 and lane 4, C2**.** Panel B, detection of the BiP-C2 chimeras with bovine anti-congopain and rabbit anti-BiP antibodies following blotting of the proteins to a nitrocellulose membrane. The antigen-antibody reaction was visualised using anti-species antibodies labelled with HRPO, followed by incubation with 4-chloro-1-naphthol/H_2O_2. Lane 1, Fermentas LMWM; lane 2, proC2-R60; lane 3, R60-proC2 and lane 4, C2.

4.3.2 Cloning, expression and purification of bacterial expressed C2-BiP chimeras

Truncated catalytic domain of congopain containing or lacking the proregion (proC2 or C2) was amplified by PCR using *Taq* DNA polymerase. The amplicons migrated at 0.6 kb and 0.93 kb, as expected for C2 and proC2 open reading frames respectively (Fig. 4.13). The proregion is essential for correct folding of the protease, hence, its inclusion in the proC2 construct (Vernet *et al*., 1995). The proregion was omitted in the R60-C2 constructs because during maturation of the protease, the proregion will be cleaved leading to the loss of the antigen C2. The PCR products were purified by gel extraction to remove the PCR buffers that contains high salt concentrations.

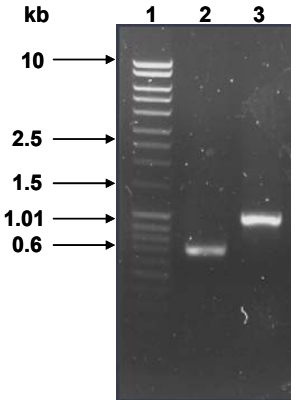

Figure 4.13. PCR amplification of C2 and proC2 from pPic9-proC2 template DNA. The amplicons were analysed by electrophoresis on a 1% (w/v) agarose gel. Lane 1, MassRuler™ DNA ladder mix; lane 2, C2 amplicon and lane 3, proC2 amplicon.

The purified PCR products (proC2 and C2) were ligated into the pGEM-T vector and the ligation mix was used to transform competent *E. coli* cells. White colonies were screened by colony PCR using T7/SP6 primers. Colony PCR was unsuccessful for pGEM-T-proC2 screening. Therefore, two random colonies were used to inoculate selective media and grown overnight to isolate plasmid DNA. Plasmid DNA migrated as three clearly visible bands at 2.8, 5.6 and 10 kb (Fig. 4.14A, lanes 2 and 3). Plasmid DNA was digested with EcoRI to confirm successful ligation of the insert into the vector (Fig. 4.14B, lanes 4 and 7). The plasmids dropped out a 1.2 kb product corresponding to the cloned fragment. The dropped out product migrated higher than the expected 0.93 kb for the proC2 open reading frame. This might be due to the high salt concentration present in restriction enzyme buffers. Plasmid DNA linearised with XbaI migrated as a single band around 5 kb (Fig. 4.14B, lanes 3 and 6). The band corresponding to the insert was gel purified and prepared for ligation into the expression vector.

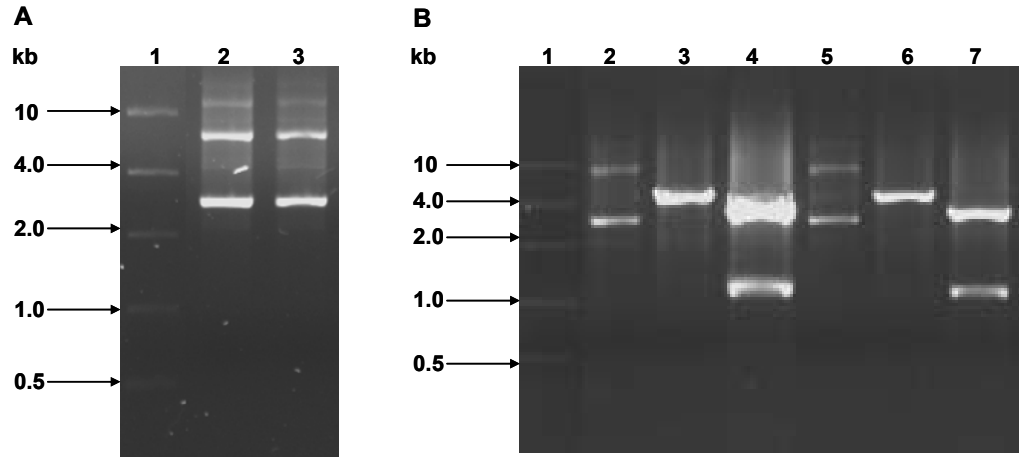

Figure 4.14. Screening of pGEM-T-proC2 recombinants by miniprep and restriction digestion. Panel A, isolation of recombinant pGEM-T-proC2 plasmid DNA using the GeneJet™ plasmid miniprep kit. Plasmid DNA was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, FastDNA™ Ruler High Range; lane 2, pGEM-T-proC2 clone 1 and lane 3, pGEM-T-proC2 clone 2. Panel B, digestion analysis of the pGEM T- vector recombinants with EcoRI or XbaI restriction endonucleases. The digested plasmid DNA was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, FastDNA™ Ruler High Range; lane 2, unrestricted pGEM-T-proC2 clone 1; lane 3, XbaI restricted pGEM-T-proC2 clone 1; lane 4, EcoRI restricted pGEM-T-proC2 clone 1; lane 5, unrestricted pGEM-T-proC2 clone 2; lane 6, XbaI restricted pGEM-T-proC2 clone 2 and lane 7, EcoRI restricted pGEM-T-proC2 clone 2.

Colony PCR screening of pGEM-T-C2 transformants with T7/SP6 primers produced a 0.7 kb PCR product (Fig. 4.15, lanes 6, 7, 8, 10 and11) corresponding to the cloned C2 fragment. Five recombinant clones were obtained for pGEM-T-C2, two colonies (7 and 11) were chosen to inoculate selective media for subsequent plasmid DNA isolation by miniprep. Plasmid DNA was digested with XhoI/XbaI and the dropped out insert was purified by gel extraction. (Fig. 4.15B, lanes 2 and 3).

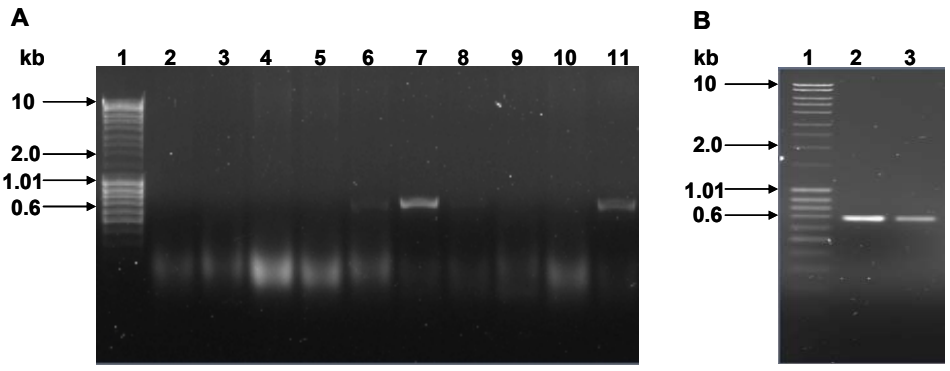

Figure 4.15. Screening of pGEM-T-C2 recombinants by colony PCR. Panel A, colony PCR screening of pGEM-T-C2 recombinants using T7 and SP6 universal primers. The PCR products were analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, MassRuler™ DNA ladder mix; lanes 2-11, pGEM-T-C2 clones 1-10. Panel B, purified C2 inserts after digestion of recombinant plasmid DNA with XhoI/XbaI. The inserts were purified by gel extraction and analysed by electrophoresis on a 1% (w/v) agarose gel. Lane 1, MassRuler™ DNA ladder mix; lane 2, C2 insert clone 7 and lane 3, C2 insert clone 11.

The pMal-R69 vector DNA was prepared by miniprep (Fig. 4.16, lane 2). Plasmid DNA migrated as two bands above 10 kb (Fig. 4.16, lane 2). Plasmid DNA was digested with either EcoRI or XhoI/XbaI to produce linear plasmid DNA (Fig. 4.16, lanes 3 and 4). Linear plasmid DNA migrated slightly above 10 kb as expected (Fig. 4.16, lanes 3 and 4). Linear plasmid DNA migrated slower than closed circular uncut DNA.

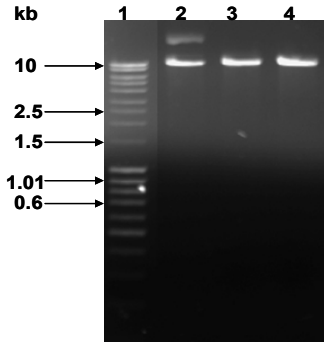

Figure 4.16. Preparation of the pMal-R69 vector by restriction with either EcoRI or XhoI/XbaI. The linear vector was analysed by agarose gel electrophoresis. Lane 1, MassRuler™ DNA ladder mix; lane 2, uncut pMal-R69; lane 3, EcoRI cut pMal- R69 and lane 4, XhoI/XbaI cut pMal-R69.

The purified inserts (proC2 and C2) were ligated into linear pMal-R69 or pMal-R60 DNA. Colony PCR screening for recombinants was attempted using the vector specific primers. Colony PCR was unsuccessful, hence, two random colonies per transformation were grown overnight in selective media. Plasmid DNA was isolated using the GeneJet™ plasmid miniprep kit. The pMal-proC2-R69 DNA migrated as a band at approximately 10 kb (Fig. 4.17, lanes 2 and 3). The yield was very low for plasmid DNA isolated using a commercial kit.

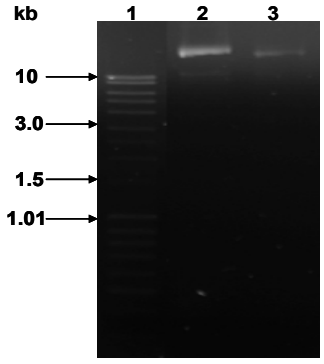

Figure 4.17. Analysis of isolated recombinant plasmid DNA of the pMal-proC2-R69 construct. Plasmid DNA was analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, MassRuler DNA™ ladder mix; lane 2, pMal-proC2-R69 clone 1 and lane 3, pMal-proC2-R69 clone 2.

A single band above 10 kb was observed for pMal-R60-C2 plasmid DNA (Fig. 4.18A, lanes 2 and 3). The yields of plasmid DNA were better than those obtained for pMal-proC2-R69. The isolated plasmid DNA was used as a template for PCR amplification of R60-C2 using pMal primers, a combination of R60EcoRIFw/pMalRv primers and CDEcoRIFw/CDXhoIRv specific primers (Table 4.1 and Table 4.3). PCR amplification of R60-C2 with pMal primers was unsuccessful (Fig. 4.18B, lanes 2 and 5). This may be due to the fact that the primers were too old, thus could not anneal properly to the template DNA. A combination of R60EcoRIFw/pMalRv primers (Table 4.1 and Table 4.3) amplified a 2.2 kb product corresponding to the expected full-length R60-C2 chimera (Fig. 4.18B, lanes 3 and 6) whereas CDEcoRIFw/CDXhoIRv specific primer pair amplified a 0.7 kb fragment corresponding to the cloned C2 fragment (Fig. 4.18B, lanes 4 and 7). Therefore, this confirmed that the transformants were recombinants.

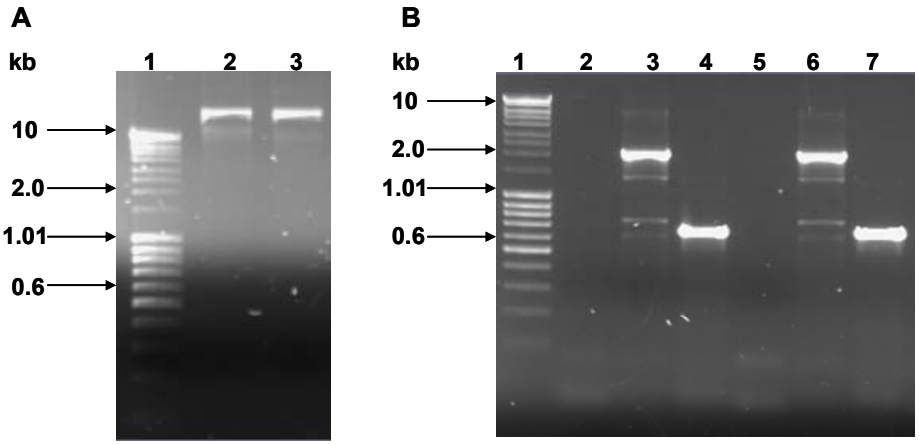

Figure 4.18. Screening of pMal-R60-C2 recombinant colonies by miniprep and PCR. Panel A, miniprep of recombinant pMa-R60-C2 plasmid DNA. The DNA was analysed by agarose gel electrophoresis. Lane 1, MassRuler™ DNA ladder mix; lane 2, pMal-R60-C2 clone 1 and lane 3, pMal-R60-C2 clone 2. Panel B, PCR amplification of insert fragments in pMal-R60 vector with pMal primers, a combination of R60EcoRIFw/pMalRv primers and CDEcoRIFw/CDXhoIRv specific primers. The PCR products were analysed by electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. Lane 1, MassRuler™ DNA ladder mix; lanes 2 and 5, amplification of pMal-60-C2 with pMal primers; lanes 3 and 6, amplification of pMal-60-C2 with R60EcoRIFw/pMalRv primers and lanes 4 and 7, amplification of pMal-R60-C2 with CDEcoRIFw/CDXhoIRv specific primer pair.

High level expression of both chimeras was achieved in the bacterial expression system. The chimeras were expressed as fusion proteins with maltose binding protein (MBP). Hence, the proC2-R69 chimera was observed at 144 kDa (Fig. 4.19A, lane 3). The size is consistent with the predicted molecular mass of the full-length MBP-proC2-BiP fusion chimera. The full-length MBP-proC2-BiP fusion chimera was overexpressed as insoluble fusion protein. Western blotting

with bovine anti-congopain antibodies confirmed expression of the fusion chimera (Fig. 4.19B, lane 3). These antibodies recognised two other bands at 45 kDa and 77 kDa respectively. This suggest possible degradation of the MBP-proC2-BiP fusion chimera (144 kDa) to yield MBP-proC2 (77 kDa) and MBP (45 kDa). Recognition of MBP by the antibodies was unsurprising since the antibodies were raised against a bacterial expressed C2.

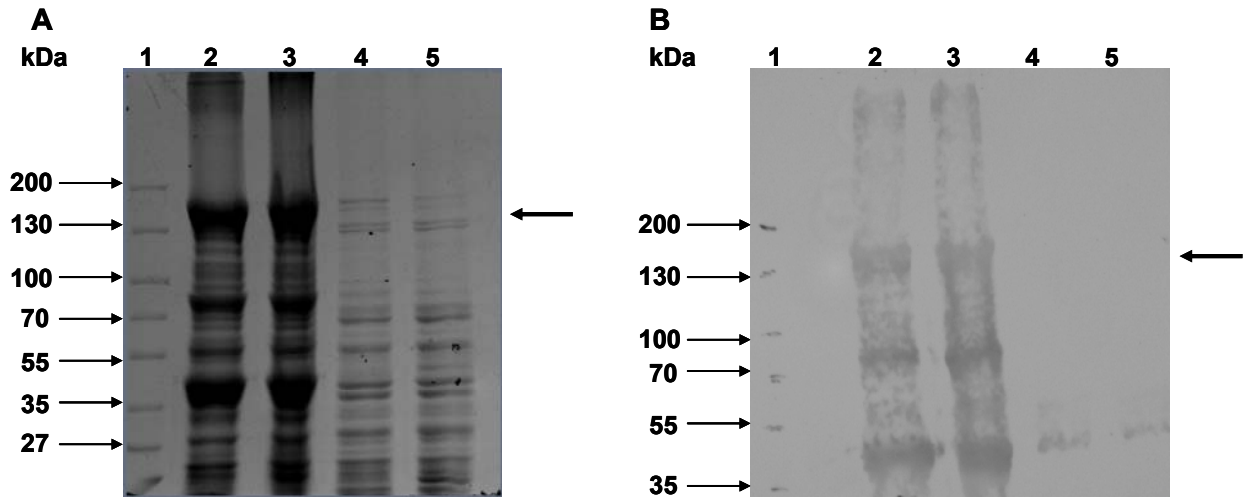

Figure 4.19. Analysis of MBP-proC2-R69 fusion chimera expression by reducing SDS-PAGE and western blotting. Panel A, reducing SDS-PAGE analysis of MBP-proC2-R69 expression by bacterial cells. The samples were boiled and reduced prior to loading. Lane 1, Fermentas protein marker; lane 2, insoluble uninduced MBP-proC2-R69; lane 3, insoluble induced MBP-proC2-R69; lane 4, soluble uninduced MBP-proC2-R69 and lane 5, soluble induced MBP-proC2-R69. Panel B, detection of the fusion chimera with bovine anti-congopain antibodies following blotting of the proteins to a nitrocellulose membrane. The antigen-antibody reaction was visualised using anti-species antibodies labelled with HRPO, followed by incubation with 4-chloro-1-napthol/H_2O_2. Lane 1, Fermentas protein marker; lane 2, insoluble uninduced MBP-proC2-R69; lane 3, insoluble induced MBP-proC2-R69; lane 4, soluble uninduced MBP-proC2-R69 and lane 5, soluble induced MBP-proC2-R69. The arrow indicates the band corresponding to MBP-proC2-R69.

The MBP-R60-C2 fusion chimera was overexpressed by *E. coli* as an insoluble protein at 126 kDa (Fig. 4.20A, lane 5). This reflects the expected size of the chimeras after fusion to MBP. The identity of the chimera band was proven by western blotting with bovine anti-congopain antibodies. These polyclonal antibodies recognised a prominent band at 126 kDa. As compared to MBP-proC2-R69, less degradation of the fusion chimera seem to occur.

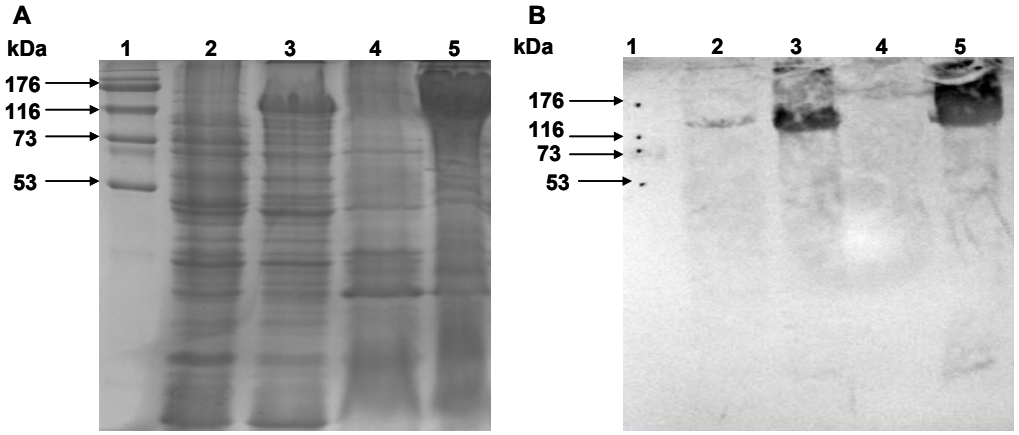

Figure 4.20. Analysis of MBP-R60-C2 fusion chimera expression by reducing SDS-PAGE and western blotting. Panel A, reducing SDS-PAGE analysis of MBP-R60-C2 expression by bacterial cells. The samples were boiled and reduced prior to loading. Lane 1, Fermentas protein marker; lane 2, soluble uninduced MBP-R60-C2; lane 3, soluble induced MBP-R60-C2; lane 4, insoluble uninduced MBP-R60-C2 and lane 5, insoluble induced MBP-R60-C2. Panel B, detection of the MBP-R60-C2 fusion chimera with bovine anti-congopain antibodies following blotting of the proteins to a nitrocellulose membrane. The antigen-antibody reaction was visualised using anti-species antibodies labelled with HRPO, followed by incubation with 4-chloro-1-napthol/H_2O_2. Lane 1, Fermentas protein marker; lane 2, soluble uninduced MBP-R60-C2; lane 3, soluble induced MBP-R60-C2; lane 4, insoluble uninduced MBP-R60-C2 and lane 5, insoluble induced MBP-R60-C2.

The inclusion bodies were successfully purified from crude cell lysates using the protocol described by Sijwali *et al*. (2001). This protocol involves successive washes of crude cell lysate with 2 M urea and 20% sucrose. The purified inclusion bodies migrated at 144 kDa corresponding to the MBP-proC2-R69 fusion chimera and 126 kDa for MBP-R60-C2 fusion chimera (Fig. 4.21, lanes 2 and 3). The MBP-proC2-R69 inclusion bodies underwent significant degradation during purification. MBP-proC2-R69 degradation products were observed at 116 kDa, 70 kDa and a faint band around 45 kDa (Fig. 4.21, lane 2).

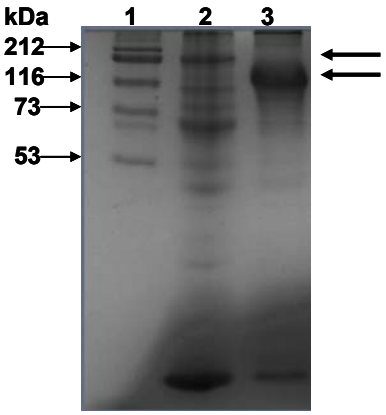

Figure 4.21. Analysis of purified inclusion bodies by reducing SDS-PAGE. The samples were boiled and reduced prior to loading. Lane 1, Pharmacia High Range protein marker; lane 2, MBP-proC2-R69 inclusion bodies and lane 3, MBP-R60-C2 inclusion bodies. The arrows indicate the bands corresponding to the fusion chimeras at 144 kDa (lane 2) and at 126 kDa (lane 3).

The inclusion bodies were solubilised with 8 M urea before renaturation. The denatured soluble fusion chimeras were refolded by drop-wise dilution of urea with the optimised renaturation buffer (Sijwali *et al*., 2001). The slow dilution of urea facilitated correct folding of the chimera to yield soluble protein. The MBP-proC2-R69 chimera was completely degraded after renaturation and concentration by ultrafiltration (Fig. 4.22, lane 3). Degradation of the chimera suggests that proC2 was properly refolded and underwent maturation to the proteolytically active C2. The MBP-R60-C2 chimera was successfully renatured. The chimera migrated as a prominent band at 116 kDa (Fig. 4.22, lane 4). This indicated that the chimera is not properly refolded. This chimera was subjected to further purification by affinity chromatography.

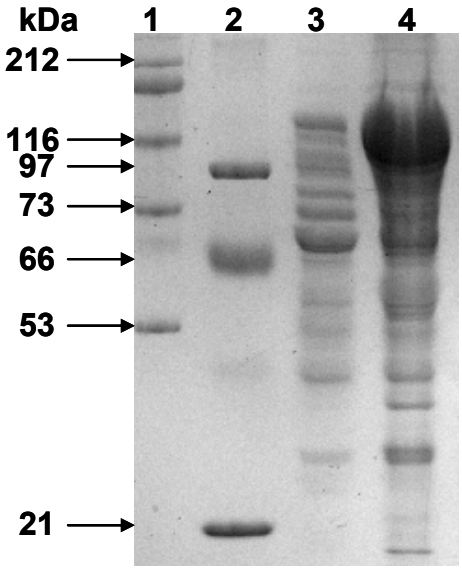

Figure 4.22. Coomassie blue R-250 staining of a reducing SDS-PAGE (12%) gel showing the renatured chimeras. The samples were boiled and reduced prior to loading. Lane 1, Pharmacia High Range protein marker; lane 2, BioRad LMWM; lane 3, renatured MBP-proC2-R69 and lane 4, MBP-R60-C2.

The renatured MBP-R60-C2 chimera was cycled through the amylose column for optimal binding to the resin. The chimera eluted with 10 mM maltose as a single peak (Fig. 4.23A). Analysis of peak fractions by SDS-PAGE revealed relatively poor purification of the chimera (Fig. 4.23B, lanes 5-10). The higher molecular weight contaminants (212 kDa) co-eluted with the chimera from the column. Other small molecular weight contaminants were still present after affinity purification.

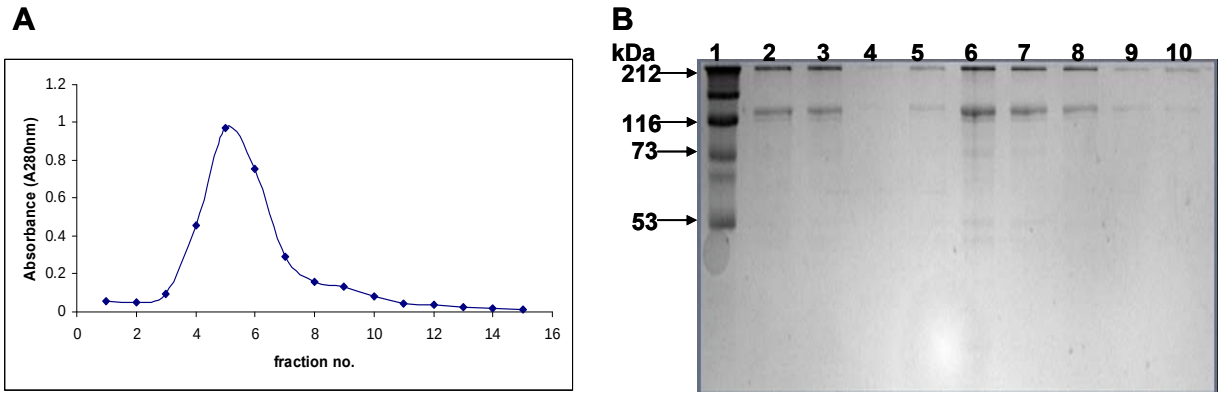

Figure 4.23. Purification of MBP-R60-C2 fusion chimera by amylose affinity chromatography. Panel A, elution profile of MBP-R60-C2 with 10 mM maltose. Elution of the fractions was monitored by measuring absorbance at 280 nm. Panel B, analysis of the elution fractions by reducing SDS-PAGE (12%) and staining with Coomassie blue R-250. Lane 1, Pharmacia High Range protein marker; lane 2, diluted renatured chimera, lane 3, unbound protein; lane 4, column wash and lanes 5-10, fractions (3-8) eluted from the amylose affinity column.

The elution fractions were pooled and concentrated using ultrafiltration. The concentrated chimera migrated at 126 kDa on a reducing SDS-PAGE gel (Fig. 4.24, lane 2). However, contaminating bands were still observed signifying failure of the purification process. The apparent impurity of the MBP-R60-C2 fusion chimera prevented the conduction of immunisation trials in mice. Thus, future work in this regards involves re-cloning of the chimeras into a pET-28a expression vector. This expression system will allow for overexpression and easy purification of the BiP-C2 chimeras using Nickel affinity chromatography. The purified BiP-C2 chimera will be used in immunisation studies to determine the adjuvant capacity of BiP.

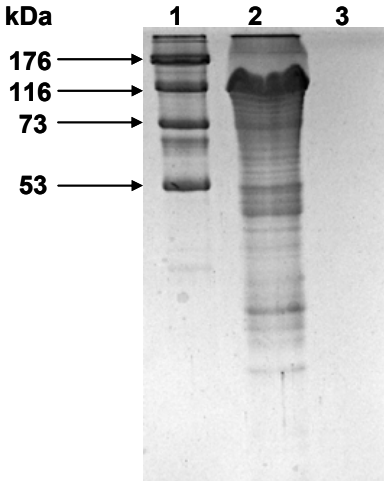

Figure 4.24. Analysis of the pooled affinity chromatography fraction of MBP-R60-C2 by reducing SDS-PAGE (12%) gel stained with Coomassie blue R-250. The samples were reduced and boiled prior to loading. Lane 1, Pharmacia High Range protein marker; lane 2, concentrated MBP-R60-C2 and lane 3, filtrate of MBP-R60-C2

4.4 DISCUSSION

Immunisation of cattle with an anti-disease vaccine containing recombinant C2 co-administered with the RWL proprietary adjuvant from SmithKline-Beecham provided limited protection against trypanosomosis (Authié *et al*., 2001). The pathological effects associated with congopain activity were less pronounced in immunised animals. Since the production of the RWL adjuvant was discontinued by SmithKline-Beecham, subsequent studies conducted in cattle using recombinant C2 expressed in *P. pastoris* co-administered with ISA206 oil-in-water adjuvant from Seppic failed to induce a protective immune response (E. Authié and A. Boulangé, personal communication). This failure is attributed to the low antigenicity of congopain, the expression system used to produce the antigen and the type of adjuvant used. Recombinant C2 expressed in *P. pastoris* has been shown to dimerise at physiological pH (Boulangé *et al*., submitted). The fact that sera from trypano-tolerant cattle recognised dimer associated epitopes had significant implications for the development of an effective anti-disease vaccine. In the present study, the ability of *T. congolense* BiP to deliver antigens to cells of immune system while maintaining the correct conformation of the antigen was investigated.

Previous studies investigating the adjuvanticity of mycobacterial HSP70 revealed the need to genetically fuse the antigen to the HSP70 molecule (Tobian *et al*., 2005; Ge *et al*., 2006; Su *et al*., 2007). In this study, different constructs were designed to contain BiP genetically fused to C2. The constructs were designed for expression in *P. pastoris* and *E. coli*. Since vectors containing the proC2 open reading frame (pPic9-proC2) (Boulangé *et al*., submitted) and full-length BiP ORF (pMal-R69) (Boulangé *et al*., 2002) were already available, it was cheap, convenient and less tedious to use these vectors to clone and express the chimeras. Two constructs were designed for expression in each expression system. ProC2-R60 and R60-proC2 were designed for expression in *P. pastoris*. In both constructs, the proregion of congopain was included because of its importance in proper folding of the protease and in maintaining the protease in the inactive zymogen state (Lalmanach *et al*., 1998; Sanderson *et al*., 2000; Lanfranco *et al*., 2008). Truncated BiP (R60), where the C-terminal domain was omitted due to its high antigenicity, was used in these constructs (Boulangé and Authié, 1994). For bacterial expression, proC2-R69 and R60-C2 constructs were designed. The construct containing full-length BiP (R69) was designed to assess the role played by the 10 kDa C-terminal domain in the stimulation of the immune response compared to truncated BiP.

The two chimeras (proC2-R60 and R60-proC2) were successfully expressed by *P. pastoris*, although the yields were very poor. Protein bands corresponding to proC2-R60 and R60-proC2 chimeras were not visible in SDS-PAGE gels analysing expression supernatants. This is possibly due to the large volumes of the expression supernatants. Faint bands corresponding to proC2-R60 and R60-proC2 chimeras (84 kDa) became visible after concentration of the expression supernatants by three phase partitioning. Bovine anti-congopain antibodies recognised bands at 27 kDa and 33 kDa corresponding to mature C2 and immature proC2. The exact mechanism for the maturation of proC2 is not clear since the cultures were maintained at neutral pH (pH 7.0) during the entire expression period whereas maturation usually occurs at low pH (Vernet *et al*., 1995; Serveau *et al*., 2003). Processing of the protease is proposed to be conducted by proteases secreted or released by dead *P. pastoris* cells in the culture. Cathepsins L and S have been shown to play a role in the activation of lysosomal proteases such as cathepsins X and C (Dahl *et al*., 2001; Turk *et al*., 2001). The rabbit anti-BiP antibodies weakly recognised the 84 kDa bands corresponding to the chimeras. Weak recognition displayed by these antibodies may be due to the fact that these antibodies recognise epitopes located in the C-terminal domain of BiP (Boulangé *et al*., 2001). Due to the insufficient quantity and impurity of the chimeras, immunisation experiments were not conducted in the murine model.

High level expression of functionally active C2 had been achieved in *P. pastoris* (Boulangé *et al*., submitted). The yeast cells produced 10-20 mg of protein per litre of culture. The poor expression level of the chimeras in *P. pastoris* is unlikely to be associated with the cysteine protease ORF. In fact, attempts to express BiP of *T. congolense* in *P. pastoris* failed (Dr A. Boulangé, University of KwaZulu-Natal, personal communication). Therefore, genetic fusion of the BiP ORF to the C2 ORF significantly altered codon usage by *P. pastoris*, resulting in poor expression of the chimeras. Similar results were obtained with the expression of mycobacterial HSP70 fused to E-protein of the Japanese encephalitis virus in *P. pastoris* (Ge *et al*., 2006). It can therefore be concluded that the *P. pastoris* expression system is not suitable for the expression of heat shock proteins including BiP.

High level expression of the chimeras was achieved in *E. coli* cells. The chimeras were expressed as fusion proteins with the carrier protein, maltose binding protein. The MBP-proC2-R69 fusion chimera migrated at 144 kDa while MBP-R60-C2 migrated at 126 kDa. The solubility test

indicated that the chimeras were expressed as insoluble inclusion bodies by the bacterial cells. Unfortunately, expression of the chimeras as insoluble inclusion bodies was a major drawback. This necessitated solubilisation and renaturation of the chimeras in order to yield soluble protein that could be used in the immunisation trials. Both chimeras were successfully solubilised with 8 M urea. The next step was to attempt to renature the chimeras using the protocol described by Sijwali *et al*. (2001). The large size of the chimeras and the presence of disulfide bridges made refolding a cumbersome task. Renaturation was performed using a buffer optimised for appropriate refolding of falcipain-2 (Sijwali *et al*., 2001). The buffer contained both reduced and oxidised glutathione in a 1:1 ratio to aid formation of correct disulfide bridges in the chimera (Sijwali *et al*., 2001). In addition, the refolding buffer contained additives such as L-arginine and glycerol to prevent incorrect folding of the protein and to maintain the stability of the folded protein during renaturation (Yasuda *et al*., 1998; Jaspard, 2000). The presence of EDTA in the refolding buffer prevents inappropriate metal-catalysed oxidation of the cysteine residues (Singh and Panda, 2005).

The chimeras were successfully renatured by slow dilution with refolding buffer and concentrated by ultrafiltration using an Amicon PM 10 membrane. SDS-PAGE analysis of the chimeras after renaturation revealed complete degradation of the MBP-proC2-R69 fusion chimera. Degradation of this chimera indicated proper folding of the protease and processing of the proenzyme to the mature forms that is possibly responsible for the degradation of the chimera. The renatured MBP-R60-C2 chimera was not degraded after renaturation. The presence of low molecular weight contaminants necessitated a further purification step. The affinity of maltose binding protein for amylose was exploited to purify the fusion chimera using amylose affinity chromatography. Although the chimera was successfully eluted from the column, several contaminants were still present after concentration of the chimera by ultrafiltration using an Amicon PM 10 membrane. The failure to attain pure chimera prevented immunisation studies in mice to determine the adjuvanticity of *T. congolense* BiP.

Although it was desirable to express the chimeras in a eukaryotic system, the low yields and high cost of the expression media were major drawbacks. Therefore, future work will involve re-cloning of the chimeras into a bacterial system that would allow for high level expression and easy purification of the chimeras. Cloning into a pET-28a expression vectors is desirable since

the expressed protein is fused with a small histidine tag. The histidine tag does not interfere with the activity of the serine protease, oligopeptidase B (Morty *et al*., 1999; Morty *et al*., 2005). The advantage of using this approach is that affinity purification can be performed using nickel-nitrilotriacetic acid (Ni-NTA) resin under denaturing conditions (Sijwali *et al*., 2001). Purification of the inclusion bodies under denaturing conditions will possibly remove associated contaminants and drastically improve refolding efficiency as was found for renaturation of egg white lysozyme (Maachupalli-Reddy *et al*., 1997). The soluble pure chimeric protein will be renatured and the endotoxins removed using the EndoTrap™ column. Finally, immunisation experiments will be conducted in mice with the purified chimera to assess the capacity of BiP to efficiently deliver the antigen to the immune system.

In the present study, the BiP-C2 chimeras were cloned and successfully expressed in both *P. pastoris* and *E. coli*. The yield and purity of the *P. pastoris* expressed chimeras was very low while technical problems associated with purification were encountered with the bacterial expressed chimeras. Time constraints did not allow for re-cloning of the chimeras into a pET-28a expression vector. Due to these reasons, the immunisation experiments were not conducted in mice to assess the capacity of *T. congolense* BiP to act as an adjuvant that stimulates an enhanced immune response that is protective against an infection with *T. congolense*.

CHAPTER 5

GENERAL DISCUSSION

African bovine trypanosomosis (nagana) is a devastating parasitic disease responsible for severe loss in livestock production in many parts of the developing world. The disease is estimated to cost US $ 5 billion per annum in Africa alone (Kristjanson *et al*., 1999; Shaw, 2004). The symptoms of the disease include fever, anaemia, loss of productivity and ultimately death if the infected animals are left untreated. Control of animal trypanosomosis critically relies on three principal strategies: vector (tsetse fly) control, chemotherapy and exploitation of trypano-tolerant cattle (McDermott and Coleman, 2001). The negative impact of insecticides on the environment has significantly limited vector control measures. Emergence of drug resistant parasites has rendered trypanocidal drugs ineffective (Anene *et al*., 2001; Geerts *et al*., 2001). Therefore, there is an urgent need to develop new strategies for controlling the disease in animals.

In this context, vaccination has been explored by researchers as an alternative for controlling trypanosomosis in cattle. Initial vaccination studies targeting the variable surface glycoprotein coat (VSG) of the parasite were proven to be ineffective (Morrison *et al*., 1982; Wells *et al*., 1982). This is due to the fact that the parasite has the capacity to change its surface coat during infection, a process known as antigenic variation (Donelson, 2003; Barry and Carrington, 2004). This necessitated a shift towards targeting invariant parasite antigens such as flagellar pocket antigen (Mkunza *et al*., 1995) or tubulin (Lubega *et al*., 2002; Li *et al*., 2007). Although the structural components of the parasite provided partial protection against the disease, attempts to reproduce the results were unsuccessful and controversy lingered. A new approach is needed to develop an effective vaccine for animal trypanosomosis.

The concept of an “anti-disease” vaccine was proposed based on targeting the pathogenic factors released by the parasites, rather than targeting the parasite itself (Playfair *et al*., 1990; Authié, 1994; Schofield, 2007). Cysteine proteases have been implicated in the development of the disease in animals. Congopain, the major cysteine protease of *T. congolense* is a circulating antigen found in the bloodstream of infected animals (Serveau *et al*., 2003). Antibodies inhibiting the activity of congopain are associated with resistance to the disease (Authié *et al*., 2001). Therefore, congopain is an attractive vaccine candidate for cattle trypanosomosis.

Congopain is a cathepsin L-like cysteine protease belonging to Clan CA of the cysteine protease family. Maturation of congopain takes place through autocatalytic removal of the propeptide at acidic pH (Serveau *et al*., 2003). The resulting catalytic domain contains the classical papain-like catalytic triad of C^{25}, H^{159} and N^{175} (papain numbering) linked to the C-terminal domain by a polyproline hinge region (Fish *et al*., 1995; Chagas *et al*., 1997). The C-terminal domain is unique to *Trypanosomatidae* cysteine proteases, being absent in mammalian homologues such as cathepsins L and B (Mottram *et al*., 1989; Hellman *et al*., 1991). Besides its high immunogenicity, hence a potential "smoke-screen" effect, the function of the C-terminal domain is yet to be elucidated.

The vaccine potential of congopain had been assessed in cattle immunised with the baculovirus expressed C2, administered in conjunction with RWL, a saponin-based proprietary adjuvant from SmithKline-Beecham. The immunised and control animals were challenged with *T. congolense*. Although immunisation of animals did not prevent the onset of an infection, the immunised animals were partially protected against the disease (Authié *et al*., 2001). Disease protection was characterised by a prominent IgG response to congopain and recovery of leukocyte counts 2-3 months post infection. Subsequent immunisation studies conducted in cattle immunised with recombinant C2 expressed in *P. pastoris* and administered with ISA206 oil-in-water adjuvant from Seppic failed to provide protection against the disease (E. Authié and A. Boulangé, personal communication). It was suggested that the failure of the study could be due to the poor antigenicity of congopain, the type of adjuvant used and the expression system used to produce the antigen, all boiling down to an incorrect presentation of protective epitopes. Protective epitopes of congopain seem to be associated with the dimeric conformation of the protease (Boulangé *et al*., submitted). Therefore, identification of the dimer-associated protective epitopes, and improving the delivery of congopain to the immune system, while maintaining the native conformation of the protease, are crucial for the development of an effective anti-disease vaccine.

Understanding the dimerisation mechanism of congopain is necessary for the identification of the dimer-associated protective epitopes. Prof. J. Hoebeke (IBMC, Strasbourg, France, personal communication) proposed a dimerisation model for congopain based on the 3-D structure of cruzipain, a monomer (McGrath *et al*., 1995) and a theoretical model of cathepsin W that is

dimeric (Brinkworth *et al*., 2000). The model identified a stretch of charged and uncharged amino acid residues that form salt bridges depending on the pH. This stretch of amino acid residues is composed of $W^{38}K^{39}V^{40}A^{41}G^{42}H^{43}E^{44}L^{45}$. In the present study, the proposed dimerisation model was tested by altering certain amino acid residues located within the proposed dimerisation motif by site-directed mutagenesis to mimick proteases of differing dimerisation capabilities. To mimick cruzipain, K^{39} and E^{44} were replaced with F^{39} and P^{44} whereas H^{43} was substituted with W^{43} to mimick cathepsin W.

The functionally active C2 mutants were heterologously expressed in *P. pastoris* alongside wild-type C2. The capacity of the C2 mutants and wild-type C2 to dimerise was assessed by PhastGel® SDS-PAGE. The electrophoretic mobility of the C2 mutants on a PhastGel® was different from that of wild-type C2. The mutant mimicking cathepsin W [C2 (H43W)] was expected to be a dimer regardless of pH, thus exhibiting electrophoretic mobility similar to wild-type C2. Analysis of this mutant under non-denaturing conditions revealed a mobility shift at 60 kDa while wild-type C2 migrated at 42 kDa. The mutant mimicking cruzipain [C2 (K39F; E44P)] was expected to be a monomer regardless of pH. However, under non-denaturing conditions the mutant was observed at 50 kDa. The effects of pH on the multimerisation of the C2 mutants as well as wild-type C2 were assessed. All the congopain variants were observed as multimers at physiological pH. This includes the C2 (K39F; E44P) mutant that was expected to be a monomer regardless of pH. Incubation of the congopain variants at acidic pH prevented multimerisation of the proteases. Therefore, the proposed dimerisation model of congopain is partially valid. The results obtained in this study suggest that the dimerisation mechanism of congopain involves more amino acid residues than postulated in the model. Brinkworth *et al*. (2000) showed that the dimerisation mechanism of cathepsin W involves amino acid residues 1-10, 40-80 and 170-200. Due to the fact that the dimerisation mechanism of congopain was not completely elucidated, the dimer associated protective epitopes were not identified.

Mutagenesis of amino acid residues located on the surface of congopain altered the enzymatic characteristics of the C2 mutants compared to wild-type C2. Significant differences were observed in the specificity constants (k_{cat}/K_m) of wild-type C2 and the C2 mutants. The introduced mutations decreased the catalytic specificity of C2, even though the mutated amino acid residues are far from the substrate binding pocket. These results were similar to those

obtained by Chen *et al*. (2008) whereby mutagenesis of amino acid residues located in the dimer interface of the SARS 3C-like protease altered the catalytic activity by the protease and changed the dimer-monomer equilibria. An inhibitor targeting the N-terminal finger of the SARS 3C-like protease prevented the catalytic activity of the protease (Ding *et al*., 2005). Therefore, development of an inhibitor preventing the dimerisation of congopain could significantly limit the catalytic activity of the protease. Inhibition of congopain activity could limit the pathological effects associated with the activity of the protease *in vivo*.

Crystallisation of congopain is necessary to map protective epitopes located at the surface of the protease and provide information regarding the dimerisation mechanism of congopain. In this second part of the present study, high level expression of wild-type C2 as well as the mutant forms of C2 was conducted in *P. pastoris*. The proteases were successfully purified to crystallisation standards using a combination of three phase partitioning, molecular exclusion chromatography and cation exchange chromatography. The proteases were sent for crystallisation at the Institut de Biologie Structurale (Grenoble, France) as part of a collaborative agreement with Prof. F. Vellieux. At the time of writing this dissertation, crystallisation of the proteases was still underway. The information obtained from the crystallisation of C2 as well as the mutant forms of C2 will be crucial for the design and construction of peptidomimotopes that could be used in vaccination studies in cattle. While awaiting the crystallographic data, *in silico* homology modelling was used as an alternative method for determining the 3-D structure of congopain. The 3-D model of congopain was constructed based on the crystal structure of cruzipain (McGrath *et al*., 1995). The 3-D structure of congopain is similar to that of papain-like cysteine proteases. The structure consisted of α-helices and β-sheets that folded to form the two domains of the protease delineated L and R domains (Lecaille *et al*., 2001).

The last part of the study involved improving the delivery of C2 to the cells of the immune system while maintaining the native conformation of the protease, to improve the likelihood of producing catalytic activity-inhibiting antibodies (anti-disease vaccine). Although the chemically formulated adjuvants such as the Freund's adjuvant are potent stimulators of the immune system (Wilson-Welder *et al*., 2008), their capacity to maintain the native conformation of the antigen is unknown. In addition, Freund's adjuvant is often toxic to the immunised animals due to the inflammatory response induced by the mycobacterial antigens present in the adjuvant (Singh and

O'Hagan, 2003). In the past few years, a large amount of literature reported the adjuvant potential of HSP70. Close collaborators were already working on *T. congolense* HSP70/BiP (Boulangé and Authié, 1994; Garzòn *et al*., unpublished). Therefore, the capacity of *T. congolense* BiP to efficiently deliver C2 to the immune system and induce the production of antibodies inhibiting the activity of congopain was investigated. Genetic fusion of an antigen with the HSP70/BiP is a prerequisite for efficient delivery of the antigen to cells of the immune system (Basu and Srivastava, 2000; Srivastava, 2002; Ge *et al*., 2006). Therefore, in this part of the study, C2 or proC2 was genetically fused to full-length BiP or truncated BiP. The quantity and purity of the *P. pastoris* expressed BiP-C2 chimeras was low, hence immunisation studies were not conducted in mice to determine the adjuvanticity of BiP. The chimeras were overexpressed in a bacterial expression system. However, problems were encountered regarding the renaturation and purification of the insoluble proteins preventing again performing the immunisation studies in mice.

Future work in this regard involves re-cloning of the chimeras in an expression vector that would combine overexpression and ease of purification. The pET-28a expression vector would be a good choice since it expresses proteins as fusion proteins with a small histidine tag. The advantage of this system is that the proteins can be easily purified using nickel-chelate chromatography under denaturing conditions due to the insoluble nature of the proteins. The expressed chimeras will be passed through an EndoTrap™ column to remove lipopolysaccharides present in the bacterial lysates that are highly immunogenic. Immunisation studies using the purified chimeras will be conducted in mice to assess the adjuvanticity of *T. congolense* BiP.

In the present study, the mutant forms of congopain were successfully cloned and recombinantly expressed in yeast. The introduced mutations altered the tertiary structure of the C2 mutants compared to wild-type C2. The observed differences in the electrophoretic mobility of the mutant forms of C2 and wild-type C2 conferred partial validity to the proposed dimerisation model, confirming the involvement of the identified dimerisation motif in the dimerisation mechanism of congopain. In line with improving the protective potential of congopain, the 3-D structure of congopain was required for the identification of protective epitopes to be used to develop mimotopes. In this regard, wild-type C2 and the mutant forms of C2 were recombinantly

expressed and purified to standards required for crystallisation. Unfortunately, since crystallisation is a time consuming process, the results are not yet available. Once the crystal structure of congopain has been resolved and the mimotopes have been developed, further work will involve their testing in murine or bovine models of infection. In order to assess the adjuvant potential of *T. congolense* BiP, the BiP-C2 chimeras were successfully cloned and expressed in this study. Due to technical problems associated with purification and renaturation, immunisation studies could not be performed in mice.

Mapping of protective epitopes associated with the dimer conformation of congopain is a big step towards development of an effective anti-disease vaccine for trypanosomosis. The present study was important as it confirmed the involvement of the proposed dimerisation motif in the dimerisation mechanism of congopain. Crystallisation of congopain is essential for the design and construction of mimotopes and possibly development of a chemotherapeutic inhibitor. Genes coding for a cathepsin L-like cysteine protease in *T. congolense* exist in a multi-copy tandem array (A. Boulangé, unpublished). The presence of multiple congopain variants has raised doubts about the effectiveness of one congopain-based anti-disease vaccine against all the circulating variants. The diversity of the congopain-like gene family and enzymatic characteristics of congopain-like variants have been studied (Pillay, 2008; Kakundi, 2009). Although the congopain variants exhibited significant differences in their kinetic constants, their exact role in host-parasite interactions still needs to be determined. Further studies need to be conducted, possibly using gene disruption technology (Mottram *et al*., 1996) to elucidate the role played by the multiple congopain-like variants *in vivo*. Other potentially pathogenic factors have been identified in the secretome of *T. congolense* using immuno-proteomics technology (Holzmüller *et al*., 2008). However, more work is still necessary to determine the exact role played by these pathogenic factors in host-parasite relationships. Once the role played by these pathogenic factors has been determined, a multi-component anti-disease vaccine may be formulated to possibly provide complete protection against trypanosomosis in animals.

REFERENCES

Aguilar, C.F., Cronin, N.B., Badasso, M., Dreyer, T., Newman, M.D., Hoover, D.J., Wood, S.P., Johnson, M.J., and Blundell, J.C. (1997). The three-dimensional structure at 2.4 angstroms resolution of glycosylated proteinase A from the lysosome-like vacuole of *Saccharomyces cerevisae*. J Mol Biol *267*, 899-915.

Aksoy, S. (2003). Control of tsetse flies and trypanosomes using molecular genetics. Vet Parasitol *115*, 125-145.

Alexander, J., Coombs, G.H., and Mottram, J.C. (1998). *Leishmania mexicana* cysteine proteinase-deficient mutants have attenuated virulence for mice and potentiate a Th1 response. J Immunol *161*, 6794-6801.

Allsopp, R., and Hursey, B.H. (2004). Insecticidal control of tsetse. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 490-507.

Aloulou, A., Grandval, P., De Caro, J., De Caro, A., and Carriere, F. (2006). Constitutive expression of human pancreatic lipase-related protein 1 in *Pichia pastoris*. Protein Expr Purif *47*, 415-421.

Alvarez, V., Parussini, F., Aslund, L., and Cazzulo, J.J. (2002). Expression in insect cells of active mature cruzipain from *Trypanosoma cruzi*, containing its C-terminal domain. Protein Expr Purif *26*, 467-475.

Anene, B.M., Onah, D.N., and Nawa, Y. (2001). Drug resistance in pathogenic African trypanosomes: what hopes for the future? Vet Parasitol *96*, 83-100.

Asea, A., Kraeft, S.K., Kurt-Jones, E.A., Stevenson, M.A., Chen, L.B., Finberg, R.W., Koo, G.C., and Calderwood, S.K. (2000). HSP70 stimulates cytokine production through a CD14-dependant pathway, demonstrating its dual role as a chaperone and cytokine. Nat Med *6*, 435-442.

Authié, E., Muteti, D.K., Mbawa, Z.R., Lonsdale-Eccles, J.D., Webster, P., and Wells, C.W. (1992). Identification of a 33-kilodalton immunodominant antigen of *Trypanosoma congolense* as a cysteine protease. Mol Biochem Parasitol *56*, 103-116.

Authié, E., Duvallet, G., Robertson, C., and Williams, D.J. (1993a). Antibody responses to a 33 kDa cysteine protease of *Trypanosoma congolense*: relationship to 'trypanotolerance' in cattle. Parasite Immunol *15*, 465-474.

Authié, E., Muteti, D.K., and Williams, D.J. (1993b). Antibody responses to invariant antigens of *Trypanosoma congolense* in cattle of differing susceptibility to trypanosomiasis. Parasite Immunol *15*, 101-111.

Authié, E. (1994). Trypanosomiasis and trypanotolerance in cattle: a role for congopain? Parasitol Today *10*, 360-364.

Authié, E., Boulangé, A., Muteti, D., Lalmanach, G., Gauthier, F., and Musoke, A.J. (2001). Immunisation of cattle with cysteine proteinases of *Trypanosoma congolense*: targetting the disease rather than the parasite. Int J Parasitol *31*, 1429-1433.

Ayra-Pardo, C.C., Martinez, G., and de la Riva, G.A. (1998). A single step screening procedure for *Pichia pastoris* clones by PCR. Biotechnol Aplicada *15*, 173-175.

Bangs, J.D., Uyetake, L., Brickman, M.J., Balber, A.E., and Boothroyd, J.C. (1993). Molecular cloning and cellular localization of a BiP homologue in *Trypanosoma brucei*. Divergent ER retention signals in a lower eukaryote. J Cell Sci *105* 1101-1113.

Barrett, A.J., and Kirschke, H. (1981). Cathepsin B, Cathepsin H, and cathepsin L. Methods Enzymol *80*, 535-561.

Barrett, A.J., Kembhavi, A.A., Brown, M.A., Kirschke, H., Knight, C.G., Tamai, M., and Hanada, K. (1982). L-trans-Epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. Biochem J *201*, 189-198.

Barry, J.D., and McCulloch, R. (2001). Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. Adv Parasitol *49*, 1-70.

Barry, J.D., and Carrington, M. (2004). Antigenic variation. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 25-37.

Barry, J.D., Marcello, L., Morrison, L.J., Read, A.F., Lythgoe, K., Jones, N., Carrington, M., Blandin, G., Bohme, U., Caler, E., *et al.* (2005). What the genome sequence is revealing about trypanosome antigenic variation. Biochem Soc Trans *33*, 986-989.

Basu, S., and Srivastava, P.K. (2000). Heat shock proteins: the fountainhead of innate and adaptive immune responses. Cell Stress Chaperones *5*, 443-451.

Black, S.J., Sicard, E.L., Murphy, N., and Nolan, D. (2001). Innate and acquired control of trypanosome parasitaemia in Cape buffalo. Int J Parasitol *31*, 562-565.

Blum, H., Beier, H., and Gross, H.J. (1987). Improved silver staining of plants proteins, RNA and DNA in plyacrylamide gels. Electrophoresis *8*, 93-99.

Boulangé, A., and Authié, E. (1994). A 69 kDa immunodominant antigen of *Trypanosoma* (*Nannomonas*) *congolense* is homologous to immunoglobulin heavy chain binding protein (BiP). Parasitology *109*, 163-173.

Boulangé, A., Serveau, C., Brillard, M., Minet, C., Gauthier, F., Diallo, A., Lalmanach, G., and Authié, E. (2001). Functional expression of the catalytic domains of two cysteine proteinases from *Trypanosoma congolense*. Int J Parasitol *31*, 1435-1440.

Boulangé, A., Katende, J., and Authié, E. (2002). *Trypanosoma congolense*: expression of a heat shock protein 70 and initial evaluation as a diagnostic antigen for bovine trypanosomosis. Exp Parasitol *100*, 6-11.

Boulangé, A., Khamadi, S., Coetzer, T., and Authié, E. (submitted). Production of congopain, the major cysteine protease of *Trypanosoma* (*Nannomonas*) *congolense*, in *Pichia pastoris* reveals unexpected dimerisation at physiological pH.

Brinkworth, R.I., Tort, J.F., Brindley, P.J., and Dalton, J.P. (2000). Phylogenetic relationships and theoretical model of human cathepsin W (lymphopain), a cysteine proteinase from cytotoxic T lymphocytes. Int J Biochem Cell Biol *32*, 373-384.

Brooks, D.R., Tetley, L., Coombs, G.H., and Mottram, J.C. (2000). Processing and trafficking of cysteine proteases in *Leishmania mexicana*. J Cell Sci *113*, 4035-4041.

Bukau, B., Weissman, J., and Horwich, A. (2006). Molecular chaperones and protein quality control. Cell *125*, 443-451.

Caffrey, C.R., Hansell, E., Lucas, K.D., Brinen, L.S., Alvarez Hernandez, A., Cheng, J., Gwaltney, S.L., 2nd, Roush, W.R., Stierhof, Y.D., Bogyo, M., *et al.* (2001). Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*. Mol Biochem Parasitol *118*, 61-73.

Castellino, F., Boucher, P.E., Eichelberg, K., Mayhew, M., Rothman, J.E., Houghton, A.N., and Germain, R.N. (2000). Receptor-mediated uptake of antigen/heat shock protein complexes results in major histocompatibility complex class I antigen presentation via two distinct processing pathways. J Exp Med *191*, 1957-1964.

Cazzulo, J., Stoka, V., and Turk, V. (2001). The major cysteine proteinase of *Trypanosoma cruzi*: a valid target for chemotherapy for Chagas disease. Curr Pharm Des *7*, 1143-1156.

Chagas, J.R., Authié, E., Serveau, C., Lalmanach, G., Juliano, L., and Gauthier, F. (1997). A comparison of the enzymatic properties of the major cysteine proteinases from *Trypanosoma congolense* and *Trypanosoma cruzi*. Mol Biochem Parasitol *88*, 85-94.

Chen, S., Zhang, J., Hu, T., Chen, K., Jiang, H., and Shen, X. (2008). Residues on the dimer interface of SARS coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis. J Biochem *143*, 525-536.

Chothia, C., and Lesk, A.M. (1986). The relation between the divergence of sequence and structure in proteins. Embo J *5*, 823-826.

Chowdhury, S.F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Menard, R., Lefebvre, J., Konishi, Y., Cygler, M., *et al.* (2002). Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides. J Med Chem *45*, 5321-5329.

Coombs, G.H., and Mottram, J.C. (1997). Parasite proteinases and amino acid metabolism: possibilities for chemotherapeutic exploitation. Parasitology *114*, S61-80.

d'Ieteren, G.D., Authié, E., Wissocq, N., and Murray, M. (1998). Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis. Rev Sci Tech *17*, 154-175.

Dahl, S.W., Halkier, T., Lauritzen, C., Dolenc, I., Pedersen, J., Turk, V., and Turk, B. (2001). Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. Biochemistry *40*, 1671-1678.

Daugaard, M., Rohde, M., and Jaattela, M. (2007). The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. FEBS Lett *581*, 3702-3710.

Dehrmann, F.M., Coetzer, T.H.T., Pike, R.N., and Dennison, C. (1995). Mature cathepsin L is substantially active in the ionic milieu of the extracellular medium. Arch Biochem Biophys *324*, 93-98.

Delespaux, V., Geysen, D., Van den Bossche, P., and Geerts, S. (2008). Molecular tools for the rapid detection of drug resistance in animal trypanosomes. Trends Parasitol *24*, 236-242.

Dennison, C. (1999). A Guide to Protein Isolation. In (Dordrecht, Kluwer Publishers).

Desquesnes, M., and Davila, A.M. (2002). Applications of PCR-based tools for detection and identification of animal trypanosomes: a review and perspectives. Vet Parasitol *109*, 213-231.

Dieu, M.C., Vanbervliet, B., Vicari, A., Bridon, J.M., Oldham, E., Ait-Yahia, S., Briere, F., Zlotnik, A., Lebecque, S., and Caux, C. (1998). Selective recruitment of immature and mature dendritic cells by distinct chemokines expressed in different anatomic sites. J Exp Med *188*, 373-386.

Ding, L., Zhang, X.X., Wei, P., Fan, K., and Lai, L. (2005). The interaction between severe acute respiratory syndrome coronavirus 3C-like proteinase and a dimeric inhibitor by capillary electrophoresis. Anal Biochem *343*, 159-165.

Dolečková, K., Kašný, M., Mikeš, L., Cartwright, J., Jedelský, P., Schneider, E.L., Dvořák, J., Mountford, A.P., Craik, C.S., and Horák, P. (2009). The functional expression and characterisation of a cysteine peptidase from the invasive stage of the neuropathogenic schistosome *Trichobilharzia regenti*. Int J Parasitol *39*, 201-211.

Donelson, J.E. (2003). Antigenic variation and the African trypanosome genome. Acta Trop *85*, 391-404.

Duggan, A.J. (1977). Bruce and the African Trypanosomes. Am J Trop Med Hyg *26*, 1080-1083.

Dvořák, J., Delcroix, M., Rossi, A., Vopálenskỳ, V., Pospíšek, M., Šedinová, M., Mikeš, L., Sajid, M., Sali, A., McKerrow, J.H*., et al.* (2005). Multiple cathepsin B isoforms in schistosomula of *Trichobilharzia regenti*: identification, characterisation and putative role in migration and nutrition. Int J Parasitol *35*, 895-910.

Eakin, A.E., McGrath, M.E., McKerrow, J.H., Fletterick, R.J., and Craik, C.S. (1993). Production of crystallizable cruzain, the major cysteine protease from *Trypanosoma cruzi*. J Biol Chem *268*, 6115-6118.

Eisler, M.C., Lessard, P., Masake, R.A., Moloo, S.K., and Peregrine, A.S. (1998). Sensitivity and specificity of antigen-capture ELISAs for diagnosis of *Trypanosoma congolense* and *Trypanosoma vivax* infections in cattle. Vet Parasitol *79*, 187-201.

Eisler, M.C., Dwinger, R.H., Majiwa, P.A.O., and Picozzi, A.S. (2004). Diagnosis and epidemiology of African animal trypanosomiasis. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International).

El-Sayed, N.M., and Donelson, J.E. (1997). A survey of the *Trypanosoma brucei rhodesiense* genome using shotgun sequencing. Mol Biochem Parasitol *84*, 167-178.

El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E., and Donelson, J.E. (2000). The African trypanosome genome. Int J Parasitol *30*, 329-345.

Ellis, K.J., and Morrison, J.F. (1982). Buffers of constant ionic strength for studying pH-dependent processes. Methods Enzymol *87*, 405-426.

Ersfeld, K., Asbeck, K., and Gull, K. (1998). Direct visualisation of individual gene organisation in *Trypanosoma brucei* by high-resolution in situ hybridisation. Chromosoma *107*, 237-240.

Fearon, D.T., and Locksley, R.M. (1996). The instructive role of innate immunity in the acquired immune response. Science *272*, 50-53.

Fish, W.R., Nkhungulu, Z.M., Muriuki, C.W., Ndegwa, D.M., Lonsdale-Eccles, J.D., and Steyaert, J. (1995). Primary structure and partial characterization of a life-cycle-regulated cysteine protease from *Trypanosoma* (*Nannomonas*) *congolense*. Gene *161*, 125-128.

Franke, C.R., Greiner, M., and Mehlitz, D. (1994). Investigations on naturally occurring *Trypanosoma evansi* infections in horses, cattle, dogs and capybaras (*Hydrochaeris hydrochaeris*) in Pantanal de Pocone (Mato Grosso, Brazil). Acta Trop *58*, 159-169.

Ge, F.F., Qiu, Y.F., Gao, X.F., Yang, Y.W., and Chen, P.Y. (2006). Fusion expression of major antigenic segment of JEV E protein-hsp70 and the identification of domain acting as adjuvant in hsp70. Vet Immunol Immunopathol *113*, 288-296.

Geerts, S., Holmes, P.H., Eisler, M.C., and Diall, O. (2001). African bovine trypanosomiasis: the problem of drug resistance. Trends Parasitol *17*, 25-28.

Gérczei, T., Keserü, G.M., and Náray-Szabó, G. (2000). Construction of a 3D model of oligopeptidase B, a potential processing enzyme in prokaryotes. J Mol Graph Model *18*, 7-17, 57-18.

Germain, R.N. (1994). MHC-dependant antigen processing and peptide presentation: providing ligands for T lymphocytes activation. Cell *76*, 287-299.

Gillmor, S.A., Craik, C.S., and Fletterick, R.J. (1997). Structural determinants of specificity in the cysteine protease cruzain. Protein Sci *6*, 1603-1611.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714-2723.

Gullo, C.A., and Teoh, G. (2004). Heat shock proteins: to present or not, that is the question. Immunol Lett *94*, 1-10.

Harmala, L.A., Ingulli, E.G., Curtsinger, J.M., Lucido, M.M., Schmidt, C.S., Weigel, B.J., Blazar, B.R., Mescher, M.F., and Pennell, C.A. (2002). The adjuvant effects of *Mycobacterium tuberculosis* heat shock protein 70 result from the rapid and prolonged activation of antigen-specific CD8+ T cells in vivo. J Immunol *169*, 5622-5629.

Harrison, R.W., Chatterjee, D., and Weber, I.T. (1995). Analysis of six protein structures predicted by comparative modeling techniques. Proteins *23*, 463-471.

Harth, G., Andrews, N., Mills, A.A., Engel, J.C., Smith, R., and McKerrow, J.H. (1993). Peptide-fluoromethyl ketones arrest intracellular replication and intercellular transmission of *Trypanosoma cruzi*. Mol Biochem Parasitol *58*, 17-24.

Hartl, F.U. (1996). Molecular chaperones in cellular protein folding. Nature *381*, 571-579.

Hellman, U., Wernstedt, C., and Cazzulo, J.J. (1991). Self-proteolysis of the cysteine proteinase, cruzipain, from *Trypanosoma cruzi* gives a major fragment corresponding to its carboxy-terminal domain. Mol Biochem Parasitol *44*, 15-21.

Heussen, C., and Dowdle, E.B. (1980). Electrophoretic analysis of plasminogen activators in poly-acrylamide gels containing sodium dodecyl sulfate and copolymerised substrates. Anal Biochem *102*, 196-202.

Holmes, P.H., Eisler, M.C., and Geerts, S. (2004). Current chemotherapy of animal trypanosomiasis. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingnford, CABI International), pp. 431-451.

Holzmüller, P., Grebaut, P., Peltier, J.B., Brizard, J.P., Perrone, T., Gonzatti, M., Bengaly, Z., Rossignol, M., Aso, P.M., Vincendeau, P., *et al.* (2008). Secretome of animal trypanosomes. Ann N Y Acad Sci *1149*, 337-342.

Hoover, D.M., Schalk-Hihi, C., Chou, C., Menon, S., Wlodawer, A., and Zdenov, A. (1999). Purification of receptor complexes of interleuken-10:Stoichiometry and the importance of deglycosylation in their crystallisation. Eur J Biochem *262*, 134-141.

Horn, M., Baudys, M., Voburka, Z., Kluh, I., Vondrasek, J., and Mares, M. (2002). Free-thiol Cys331 exposed during activation process is critical for native tetramer structure of cathepsin C (dipeptidyl peptidase I). Protein Sci *11*, 933-943.

Huang, R., Que, X., Hirata, K., Brinen, L.S., Lee, J.H., Hansell, E., Engel, J., Sajid, M., and Reed, S. (2009). The cathepsin L of *Toxoplasma gondii* (TgCPL) and its endogenous macromolecular inhibitor, toxostatin. Mol Biochem Parasitol *164*, 86-94.

Jaspard, E. (2000). Role of protein-solvent interaction in refolding: Effects of cosolvents additives on the renaturation of porcine pancreatic elastase at various pHs. Biochem Biophys *375*, 220-228.

Jaye, A.B., Nantulya, V.M., Majiwa, P.A.O., Urakawa, T., Masake, R.A., Wells, C.W., and Ole-MoiYoi, O. (1994). A *Trypanosoma* (*Nannomonas*) *congolense*-specific antigen released into the circulation of infected animals is a thiol protease precursor. EMBL accession L25130.

Jonak, C., Klosner, G., and Trautinger, F. (2009). Significance of heat shock proteins in the skin upon UV exposure. Front Biosci *14*, 4758-4768.

Joshi, S., Singh, A.R., Kumar, A., Misra, P.C., Siddiqi, M.I., and Saxena, J.K. (2008). Molecular cloning and characterization of *Plasmodium falciparum* transketolase. Mol Biochem Parasitol *160*, 32-41.

Kakundi, E.M. (2009). Molecular analysis of the congopain gene family. MSc Thesis. University of KwaZulu-Natal, Pietermaritzburg.

Kang, H.K., Lee, H.Y., Lee, Y.N., Jo, E.J., Kim, J.I., Aosai, F., Yano, A., Kwak, J.Y., and Bae, Y.S. (2004). *Toxoplasma gondii*-derived heat shock protein 70 stimulates the maturation of human monocyte-derived dendritic cells. Biochem Biophys Res Commun *322*, 899-904.

Klemba, M., and Goldberg, D.E. (2002). Biological roles of proteases in parasitic protozoa. Annu Rev Biochem *71*, 275-305.

Kleywegt, G.J., and Jones, T.A. (1996). Phi/psi-chology: Ramachandran revisited. Structure *4*, 1395-1400.

Kleywegt, G.J. (1997). Validation of protein models from Calpha coordinates alone. J Mol Biol *273*, 371-376.

Kongkerd, N., Uparanukraw, P., Morakote, N., Sajid, M., and McKerrow, J.H. (2008). Identification and characterization of a cathepsin L-like cysteine protease from *Gnathostoma spinigerum*. Mol Biochem Parasitol *160*, 129-137.

Kopp, J., and Schwede, T. (2004). The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. Nucleic Acids Res *32*, D230-234.

Kristjanson, P.M., Swallow, B.M., Kruska, R.L., and Leeuw, P.N.D. (1999). Measuring the cost of African animal trypanosomosis, the potential benefits of control and returns to research. Agric Sys, 79-98.

Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A *82*, 488-492.

Kunkel, T.A., Roberts, J.D., and Zakour, R.A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol *154*, 367-382.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W. (1998). Structure of an HIV gp120 envelope in complex with the CD4 receptor and a neutralising antibody. Nature *393*, 648-659.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lalmanach, G., Lecaille, F., Chagas, J.R., Authié, E., Scharfstein, J., Juliano, M.A., and Gauthier, F. (1998). Inhibition of trypanosomal cysteine proteinases by their propeptides. J Biol Chem *273*, 25112-25116.

Lalmanach, G., Boulange´ , A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., and Authie´ , E. (2002). Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. Biol Chem *383*, 739-749.

Lanfranco, M.F., Loayza-Muro, R., Clark, D., Nunez, R., Zavaleta, A.I., Jimenez, M., Meldal, M., Coombs, G.H., Mottram, J.C., Izidoro, M., *et al.* (2008). Expression and substrate specificity of a recombinant cysteine proteinase B of *Leishmania braziliensis*. Mol Biochem Parasitol *161*, 91-100.

Lazarevic, V., Myers, A.J., Scanga, C.A., and Flynn, J.L. (2003). CD40, but not CD40L, is required for the optimal priming of T cells and control of aerosol *M. tuberculosis* infection. Immunity *19*, 823-835.

Lecaille, F., Authie, E., Moreau, T., Serveau, C., Gauthier, F., and Lalmanach, G. (2001). Subsite specificity of trypanosomal cathepsin L-like cysteine proteases. Probing the S2 pocket with phenylalanine-derived amino acids. Eur J Biochem *268*, 2733-2741.

Lehner, T., Bergmeier, L.A., Wang, Y., Tao, L., Sing, M., Spallek, R., and van der Zee, R. (2000). Heat shock proteins generate beta-chemokines which function as innate adjuvants enhancing adaptive immunity. Eur J Immunol *30*, 594-603.

Lesk, A.M. (1997). Extraction of well-fitting substructures: root-mean-square deviation and the difference distance matrix. Fold Des *2*, S12-14.

Li, S.Q., Fung, M.C., Reid, S.A., Inoue, N., and Lun, Z.R. (2007). Immunization with recombinant beta-tubulin from *Trypanosoma evansi* induced protection against *T. evansi*, *T. equiperdum* and *T. b. brucei* infection in mice. Parasite Immunol *29*, 191-199.

Li, Z., and Srivastava, P. (2004). Heat-shock proteins. Curr Protoc Immunol *Appendix 1*, Appendix 1T.

Lubega, G.W., Byarugaba, D.K., and Prichard, R.K. (2002). Immunization with a tubulin-rich preparation from *Trypanosoma brucei* confers broad protection against African trypanosomosis. Exp Parasitol *102*, 9-22.

Maachupalli-Reddy, J., Kelley, B.D., and De Bernardez Clark, E. (1997). Effect of inclusion body contaminants on the oxidative renaturation of hen egg white lysozyme. Biotechnol Prog *13*, 144-150.

Martinez, J., Campetella, O., Frasch, A.C., and Cazzulo, J.J. (1991). The major cysteine proteinase (cruzipain) from *Trypanosoma cruzi* is antigenic in human infections. Infect Immun *59*, 4275-4277.

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

Mbawa, Z.R., Gumm, I.D., Fish, W.R., and Lonsdale-Eccles, J.D. (1991a). Endopeptidase variations among different life-cycle stages of African trypanosomes. Eur J Biochem *195*, 183-190.

Mbawa, Z.R., Webster, P., and Lonsdale-Eccles, J.D. (1991b). Immunolocalization of a cysteine protease within the lysosomal system of *Trypanosoma congolense*. Eur J Cell Biol *56*, 243-250.

Mbawa, Z.R., Gumm, I.D., Shaw, E., and Lonsdale-Eccles, J.D. (1992). Characterisation of a cysteine protease from bloodstream forms of *Trypanosoma congolense*. Eur J Biochem *204*, 371-379.

McDermott, J.J., and Coleman, P.G. (2001). Comparing apples and oranges--model-based assessment of different tsetse-transmitted trypanosomosis control strategies. Int J Parasitol *31*, 603-609.

McGrath, M.E., Eakin, A.E., Engel, J.C., McKerrow, J.H., Craik, C.S., and Fletterick, R.J. (1995). The crystal structure of cruzain: a therapeutic target for Chagas' disease. J Mol Biol *247*, 251-259.

McKerrow, J.H., McGrath, M.E., and Engel, J.C. (1995). The cysteine protease of *Trypanosoma cruzi* as a model for antiparasite drug design. Parasitol Today *11*, 279-282.

McKerrow, J.H., Engel, J., and Caffrey, C.R. (1999). Cysteine protease inhibitors as chemotherapy for parasitic infections. Bioorg Med Chem *7*, 639-644.

McKerrow, J.H., Caffrey, C., Kelly, B., Loke, P., and Sajid, M. (2006). Proteases in parasitic diseases. Annu Rev Pathol *1*, 497-536.

Melo, F., and Feytmans, E. (1998). Assessing protein structures with a non-local atomic interaction energy. J Mol Biol *277*, 1141-1152.

Melville, S.E., Majiwa, P.A.O., and Tait, A. (2004). The African trypanosome genome. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 39-57.

Menard, R., Carmona, E., Takebe, S., Dufour, E., Plouffe, C., Mason, P., and Mort, J.S. (1998). Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in the processing of procathepsin L in vitro. J Biol Chem *273*, 4478-4484.

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*. Eukaryot Cell *7*, 684-697.

Mewes, H.W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S.G., *et al.* (1997). Overview of the yeast genome. Nature *387*, 7-65.

Michael, M., Gerber, S., Fetzer, J., and Folkers, G. (1997). Oligonucleotide-directed mutagenesis and subsequent expression of the corresponding recombinant proteins without changing the bacterial vector system. Pharm Acta Helv *72*, 139-143.

Michael, S.F. (1994). Mutagenesis by incorporation of a phosphorylated oligo during PCR amplification. Biotechniques *16*, 410-412.

Mkhize, P.P. (2003). Epitope mapping of a trypanosomal cysteine protease. MSc Thesis, University of KwaZulu-Natal, Pietermaritzburg.

Mkunza, F., Olaho, W.M., and Powell, C.N. (1995). Partial protection against natural trypanosomiasis after vaccination with a flagellar pocket antigen from *Trypanosoma brucei rhodesiense*. Vaccine *13*, 151-154.

Molgaard, A., Arnau, J., Lauritzen, C., Larsen, S., Petersen, G., and Pedersen, J. (2007). The crystal structure of human dipeptidyl peptidase I (cathepsin C) in complex with the inhibitor Gly-Phe-CHN2. Biochem J *401*, 645-650.

Momen, H. (2001). Some current problems in the systematics of Trypanosomatids. Int J Parasitol *31*, 640-642.

Morrison, W.I., Black, S.J., Paris, J., Hinson, C.A., and Wells, P.W. (1982). Protective immunity and specificity of antibody responses elicited in cattle by irradiated *Trypanosoma brucei*. Parasite Immunol *4*, 395-407.

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

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

Morty, R.E., Pelle, R., Vadasz, I., Uzcanga, G.L., Seeger, W., and Bubis, J. (2005). Oligopeptidase B from *Trypanosoma evansi*. A parasite peptidase that inactivates atrial natriuretic factor in the bloodstream of infected hosts. J Biol Chem *280*, 10925-10937.

Mottram, J.C., North, M.J., Barry, J.D., and Coombs, G.H. (1989). A cysteine proteinase cDNA from *Trypanosoma brucei* predicts an enzyme with an unusual C-terminal extension. FEBS Lett *258*, 211-215.

Mottram, J.C., Souza, A.E., Hutchison, J.E., Carter, R., Frame, M.J., and Coombs, G.H. (1996). Evidence from disruption of the lmcpb gene array of *Leishmania mexicana* that cysteine proteinases are virulence factors. Proc Natl Acad Sci U S A *93*, 6008-6013.

Mottram, J.C., Coombs, G.H., and Alexander, J. (2004). Cysteine peptidases as virulence factors of *Leishmania*. Curr Opin Microbiol *7*, 375-381.

Murray, M., d'Ieteren, G.D., and Taele, A.J. (2004). Trypanotolerance. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 461-477.

Naessens, J. (2006). Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome? Int J Parasitol *36*, 521-528.

North, M.J., Mottram, J.C., and Coombs, G.H. (1990). Cysteine proteinases of parasitic protozoa. Parasitol Today *6*, 270-275.

Otte, M.J., Abuabara, J.Y., and Wells, E.A. (1994). *Trypanosoma vivax* in Colombia: epidemiology and production losses. Trop Anim Health Prod *26*, 146-156.

Paling, R.W., Moloo, S.K., Scott, J.R., Gettinby, G., McOdimba, F.A., and Murray, M. (1991). Susceptibility of N'Dama and Boran cattle to sequential challenges with tsetse-transmitted clones of *Trypanosoma congolense*. Parasite Immunol *13*, 427-445.

Pamer, E., and Cresswell, P. (1998). Mechanism of MHC class I-restricted antigen processing. Annu Rev Immunol *16*, 323-358.

Payne, R.C., Sukanto, I.P., Djauhari, D., Partoutomo, S., Wilson, A.J., Jones, T.W., Boid, R., and Luckins, A.G. (1991). *Trypanosoma evansi* infection in cattle, buffaloes and horses in Indonesia. Vet Parasitol *38*, 109-119.

Pelham, H.R. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell *46*, 959-961.

Pike, R.N., and Dennison, C. (1989). Protein fractionation by three phase partitioning (TPP) in aqueous/t-butanol mixtures. Biotechnol Bioeng *33*, 221-228.

Pillay, D. (2008). A study of the variability of the congopain-like cysteine proteases of *Trypanosoma congolense*. . MSc Thesis. University of KwaZulu-Natal, Pietermaritzburg.

Planelles, L., Thomas, M., Pulgar, M., Maranon, C., Grabbe, S., and Lopez, M.C. (2002). *Trypanosoma cruzi* heat-shock protein-70 kDa,alone or fused to the parasite KMP11 antigen, induces functional maturation of murine dendritic cells. Immunol Cell Biol *80*, 241-247.

Playfair, J.H., Taverne, J., Bate, C.A., and de Souza, J.B. (1990). The malaria vaccine: anti-parasite or anti-disease? Immunol Today *11*, 25-27.

Playfair, J.H., Taverne, J., and Bate, C.A. (1991). Don't kill the parasite: control the disease. Acta Leiden *60*, 157-165.

Prowse, R.K., Chaplin, P., Robinson, H.C., and Spithill, T.W. (2002). *Fasciola hepatica* cathepsin L suppresses sheep lymphocyte proliferation in vitro and modulates surface CD4 expression on human and ovine T cells. Parasite Immunol *24*, 57-66.

Quintas-Granados, L.I., Orozco, E., Brieba, L.G., Arroyo, R., and Ortega-Lopez, J. (2009). Purification, refolding and autoactivation of the recombinant cysteine proteinase EhCP112 from *Entamoeba histolytica*. Protein Expr Purif *63*, 26-32.

Rawlings, N.D., and Barrett, A.J. (1999). MEROPS: the peptidase database. Nucleic Acids Res *27*, 325-331.

Rico, A.I., Del Real, G., Soto, M., Quijada, L., Martinez, A.C., Alonso, C., and Requena, J.M. (1998). Characterization of the immunostimulatory properties of *Leishmania infantum* HSP70 by fusion to the *Escherichia coli* maltose-binding protein in normal and nu/nu BALB/c mice. Infect Immun *66*, 347-352.

Robert, J. (2003). Evolution of heat shock protein and immunity. Dev Comp Immunol *27*, 449-464.

Roditi, I., and Lehane, M.J. (2008). Interactions between trypanosomes and tsetse flies. Curr Opin Microbiol *11*, 345-351.

Rogers, D.J., and Randolph, S.E. (2002). A response to the aim of eradicating tsetse from Africa. Trends Parasitol *18*, 1489-1499.

Ruszczyk, A., Forlenza, M., Savelkoul, H.F., and Wiegertjes, G.F. (2008). Molecular cloning and functional characterisation of a cathepsin L-like proteinase from the fish kinetoplastid parasite *Trypanosoma carassii*. Fish Shellfish Immunol *24*, 205-214.

Saibil, H.R. (2008). Chaperone machines in action. Curr Opin Struct Biol *18*, 35-42.

Sajid, M., and McKerrow, J.H. (2002). Cysteine proteases of parasitic organisms. Mol Biochem Parasitol *120*, 1-21.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual. In, J. Sambrook, and D.W. Russell, eds. (Cold Springer Harbour: Cold Springer Harbour Laboratory Press), p. 1.53.

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

Sarkar, G., and Sommer, S.S. (1990). The "megaprimer" method of site-directed mutagenesis. Biotechniques *8*, 404-407.

Sbicego, S., Vassella, E., Kurath, U., Blum, B., and Roditi, I. (1999). The use of transgenic *Trypanosoma brucei* to identify compounds inducing the differentiation of bloodstream forms to procyclic forms. Mol Biochem Parasitol *104*, 311-322.

Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun *27*, 157-162.

Schijns, V.E. (2000). Immunological concepts of vaccine adjuvant activity. Curr Opin Immunol *12*, 456-463.

Schmidt, G.D., and Roberts, L.S. (1989). Order *Kinetoplastida*: Trypanosomes and their kin. In Foundations of Parasitology (USA, Times Mirror College Publishing), pp. 55-80.

Schofield, C.J., and Maudlin, I. (2001). Trypanosomiasis control. Int J Parasitol *31*, 614-619.

Schofield, L. (2007). Rational approaches to developing an anti-disease vaccine against malaria. Microbes Infect *9*, 784-791.

Schwede, T., Diemand, A., Guex, N., and Peitsch, M.C. (2000). Protein structure computing in the genomic era. Res Microbiol *151*, 107-112.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res *31*, 3381-3385.

Serveau, C., Boulangé, A., Lecaille, F., Gauthier, F., Authié, E., and Lalmanach, G. (2003). Procongopain from *Trypanosoma congolense* is processed at basic pH: an unusual feature among cathepsin L-like cysteine proteases. Biol Chem *384*, 921-927.

Shaw, A.P.M. (2004). Economics of African trypanosomiasis. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 369-402.

Shi, J., Sivaraman, J., and Song, J. (2008). Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. J Virol *82*, 4620-4629.

Sijwali, P.S., Brinen, L.S., and Rosenthal, P.J. (2001). Systematic optimization of expression and refolding of the *Plasmodium falciparum* cysteine protease falcipain-2. Protein Expr Purif *22*, 128-134.

Singh, M., and O'Hagan, D. (1999). Advances in vaccine adjuvants. Nat Biotechnol *17*, 1075-1081.

Singh, M., and O'Hagan, D.T. (2003). Recent advances in veterinary vaccine adjuvants. Int J Parasitol *33*, 469-478.

Singh, S.M., and Panda, A.K. (2005). Solubilisation and refolding of bacterial inclusion body proteins. J. Bio. Bioeng. *99*, 303-310.

Srivastava, P. (2002). Interaction of heat shock proteins with peptides and antigen presenting cells: chaperoning of the innate and adaptive immune responses. Annu Rev Immunol *20*, 395-425.

Stevens, J.R., and Brisse, S. (2004). Systematics of trypanosomes of medical and veterinary importance. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 1-23.

Storni, T., Kundig, T.M., Senti, G., and Johansen, P. (2005). Immunity in response to particulate antigen-delivery systems. Adv Drug Deliv Rev *57*, 333-355.

Su, C., Duan, X., Wang, X., Wang, C., Cao, R., Zhou, B., and Chen, P. (2007). Heterologous expression of FMDV immunodominant epitopes and HSP70 in *P. pastoris* and the subsequent immune response in mice. Vet Microbiol *124*, 256-263.

Suto, R., and Srivastava, P.K. (1995). A mechanism for the specific immunogenicity of heat shock protein-chaperoned peptides. Science *269*, 1585-1588.

Suzue, K., and Young, R.A. (1996). Adjuvant-free hsp70 fusion protein system elicits humoral and cellular immune responses to HIV-1 p24. J Immunol *156*, 873-879.

Taylor, K.A. (1998). Immune responses of cattle to African trypanosomes: protective or pathogenic? Int J Parasitol *28*, 219-240.

Taylor, K.A., and Authié, E. (2004). Pathogenesis of animal trypanosomiaisis. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 331-353.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Tobian, A.A., Harding, C.V., and Canaday, D.H. (2005). *Mycobacterium tuberculosis* heat shock fusion protein enhances class I MHC cross-processing and -presentation by B lymphocytes. J Immunol *174*, 5209-5214.

Torr, S.J., Hargrove, J.W., and Vale, G.A. (2005). Towards a rational policy for dealing with tsetse. Trends Parasitol *21*, 537-541.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A *76*, 4350-4354.

Turk, B., Turk, D., and Turk, V. (2000). Lysosomal cysteine proteases: more than scavengers. Biochim Biophys Acta *1477*, 98-111.

Turk, V., Turk, B., and Turk, D. (2001). Lysosomal cysteine proteases: facts and opportunities. Embo J *20*, 4629-4633.

Uilenberg, G. (1998). A field guide for diagnosis, treatment and prevention of African animal Trypanosomosis. In (Rome, FAO Corporate document Repository).

Urban, A., Neukirchen, S., and Jaeger, K.E. (1997). A rapid and efficient method for site-directed mutagenesis using one-step overlap extension PCR. Nucleic Acids Res *25*, 2227-2228.

Vale, G.A., and Torr, S.J. (2004). Development of bait technology to control tsetse. In The Trypanosomiasis, I. Maudlin, P.H. Holmes, and M.A. Miles, eds. (Wallingford, CABI International), pp. 509-523.

Vale, G.A., and Torr, S.J. (2005). User-friendly models of the costs and efficacy of tsetse control: application to sterilizing and insecticidal techniques. Med Vet Entomol *19*, 293-305.

van Kooten, C., and Banchereau, J. (1997). Immune regulation by CD40-CD40-L interactions. Front Biosci *2*, d1-11.

Vanhamme, L., Pays, E., McCulloch, R., and Barry, J.D. (2001). An update on antigenic variation in African trypanosomes. Trends Parasitol *17*, 338-343.

Vernet, T., Berti, P.J., de Montigny, C., Musil, R., Tessier, D.C., Menard, R., Magny, M.C., Storer, A.C., and Thomas, D.Y. (1995). Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. J Biol Chem *270*, 10838-10846.

Vickerman, K. (1969). The fine structure of *Trypanosoma congolense* in its bloodstream phase. J Protozool *16*, 54-69.

Wang, S., Longo, F.M., Chen, J., Butman, M., Graham, S.H., Haglid, K.G., and Sharp, F.R. (1993). Induction of glucose regulated protein (grp78) and inducible heat shock protein (hsp70) mRNAs in rat brain after kainic acid seizures and focal ischemia. Neurochem Int *23*, 575-582.

Wang, Y., Kelly, C.G., Singh, M., McGowan, E.G., Carrara, A.S., Bergmeier, L.A., and Lehner, T. (2002). Stimulation of Th1-polarizing cytokines, C-C chemokines, maturation of dendritic cells, and adjuvant function by the peptide binding fragment of heat shock protein 70. J Immunol *169*, 2422-2429.

Wells, G.A., Birkholtz, L.M., Joubert, F., Walter, R.D., and Louw, A.I. (2006). Novel properties of malarial S-adenosylmethionine decarboxylase as revealed by structural modelling. J Mol Graph Model *24*, 307-318.

Wells, P.W., Emery, D.L., Hinson, C.A., Morrison, W.I., and Murray, M. (1982). Immunization of cattle with a variant-specific surface antigen of *Trypanosoma brucei*: influence of different adjuvants. Infect Immun *36*, 1-10.

Wendling, U., Paul, L., van der Zee, R., Prakken, B., Singh, M., and van Eden, W. (2000). A conserved mycobacterial heat shock protein (hsp) 70 sequence prevents adjuvant arthritis upon nasal administration and induces IL-10-producing T cells that cross-react with the mammalian self-hsp70 homologue. J Immunol *164*, 2711-2717.

Williams, D.J., Naessens, J., Scott, J.R., and McOdimba, F.A. (1991). Analysis of peripheral leucocyte populations in N'Dama and Boran cattle following a rechallenge infection with *Trypanosoma congolense*. Parasite Immunol *13*, 171-185.

Wilson-Welder, J.H., Torres, M.P., Kipper, M.J., Mallapragada, S.K., Wannwmuehler, M.J., and Narasimhan, B. (2008). Vaccine adjuvants: current challenges and future approaches. J Pharm Sci, 1-39.

Wu, S., and Letchworth, G.J. (2004). High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. Biotechniques *36*, 152-154.

Wyss, D.F., and Wagner, G. (1996). The structural role of sugars in glycoproteins. Curr Opin Biotechnol *7*, 409-416.

Yasuda, M., Murakami, Y., Sowa, A., Ogino, H., and Ishikawa, H. (1998). Effects of additives on refolding of a denatured protein. Biotechnol. Prog. *14*, 601-606.

APPENDIX

Nucleotide sequence alignment of the C2 mutants with wild-type full-length congopain (CP2). The sequencing of pGEM-T-C2 (H43W) and pGEM-T-C2 (K39F; E44P) was done in ILRI sequencing unit and the sequences were compared using Sequencher® 4.7 software (Gene Codes Corporation, 2006).