Cloning of a cDNA encoding *Schistosoma mansoni* cathepsin C and expression in the yeast *Pichia pastoris*

A thesis presented for the degree of M.Sc.

By

Charlene Carty B.Sc.

Under the supervision of
Professor John P. Dalton

School of Biotechnology
Dublin City University

February 2003

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of M.Sc. is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Charlene Carty.

ID No: 51160447

Date: 10/02/2003

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor, Professor John Dalton for giving me the opportunity to work on this project and for all his support, guidance, advice and enthusiasm over the past eighteen months.

Secondly, I would like to thank everyone in the Molecular Parasitology group in particular Sheila Donnelly who helped me a great deal with my work on this project. I would also like to thank my fellow post-grads, Elaine, Alan and Peter for all their help. I wish you all every success with your careers and hope that everything works out well for you. Many thanks also to Mary Cunneen, Mary Sekiya, Colin, Sandra, Eimear, Patricia, Melanie and Rokni and any one else I may have forgotten to mention. I would also like to mention Sarah Farnon, who worked on cathepsin C for her fourth year project and Canadian students Helene and Jonathan who spent time working in the lab. Once again, thanks everyone!

A special word of thanks to Alan Stewart for everything over the past few years and to my friends Carolyn, Gary, Karoline and Lorna. Thanks for all the guidance, support, advice, fun and good times. I would also like to mention my friends from my Biotechnology class in Galway: Ray, Caoimhin, Carina, Sinead and Marie. Thanks for everything.

Many thanks also to Sligo County Council for providing me with financial assistance during my college years. It was very much appreciated.

Lastly, I would like to thank my family, in particular my mother, who is probably glad at this stage that I am finished studying. Thanks for all the support and help over the years.

TABLE OF CONTENTS

ABSTRACT

The cDNA encoding the proenzyme of cathepsin C from the parasitic helminth *Schistosoma mansoni* was amplified from a cDNA library using 5' and 3' specific primers containing sites for *SnaBI* and *AvrII* digestion. The full-length cathepsin C cDNA was cloned into the *Pichia pastoris* expression vector, pPIC9K, in frame with the initiation codon of the yeast α-factor signal sequence. Sequencing of the 1.3 kb cDNA-vector construct revealed a 94.5% identity to the *S. mansoni* cathepsin C sequence in the public database at amino acid level.

A smaller fragment was also amplified from the library and was cloned into the pGEM vector and sequenced to characterise its identity. This fragment proved to be a truncated version of the cathepsin C cDNA.

Prior to transformation of *P. pastoris* strain GS115, the pPIC9K-cathepsin C construct was linearised with *BglII*. Transformants were selected on the basis of resistance to the antibiotic G418. PCR revealed that the cathepsin C cDNA had been successfully integrated into the yeast genome of transformants.

Protein expression was induced by addition of methanol to a final concentration of 1.0-1.5%, once or twice daily for six days. Recombinant protein was detected by SDS-PAGE. Proteins of 58, 55, 47 and 25 kDa were produced but a western blot and ELISA did not detect the presence of the hexahistidine tag fused to the COOH terminus of the recombinant enzyme. Protease assays were performed to determine if the enzyme was active against the fluorogenic cathepsin C substrate H-Gly-Arg-NHMec. Activity proved low in all instances relative to positive controls.

Biochemical characterisation of *Fasciola hepatica* cathepsin C was undertaken. The enzyme was demonstrated to exhibit similar properties to its *S. mansoni* counterpart such as enhancement of activity by the reducing agent DTT and halide ions. A peak of activity was seen at pH 5.5 but also at pH 8.5, which proved to be unique to *F. hepatica* cathepsin C.

ABBREVIATIONS

Amp	Ampicillin
AOX1	Alcohol Oxidase Promoter
Asp	Asparagine
cDNA	Complementary DNA
Cys	Cysteine
DMF	Dimethylformamide
DPPI	Dipeptidyl peptidase I
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetra Acetic Acid
E-64	Trans-epoxysuccinyl-L-leucylamide (4-guanidinol)-butane
FhCC	*Fasciola hepatica* cathepsin C
GAP	Glyceraldehyde-3-Phosphate Dehydrogenase
GRAS	Generally Regarded As Safe
GST	Glutathione-S-Transferase
Hb	Haemoglobin
HBsAg	Hepatitis B Surface Antigen
H-Gly-Arg-NHMec	H-Glycine-Arginine-NHMec
His	Histidine
Hr	Hour(s)
IL	Interleukin
IPTG	Isopropyl beta-D-thiogalactopyranoside
kb	Kilobase
kDa	Kilodalton
LB	Luria-Bertani Medium
mA	Milliamps
Min	Minute(s)
Mut^+	Methanol Utilisation Plus
Mut^S	Methanol Utilisation Slow
Mw	Molecular Weight
NEM	N-ethyl-malelimide

NHMec (AMC)	7-amino-4-methyl-coumarin
Ni-NTA	Nickel Nitrilo Acetic Acid
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
Pmy	Paramyosin
PMSF	Phenylmethylsulfonylfluorid
rSmCC	Recombinant *Schistosoma mansoni* cathepsin C
RBCs	Red Blood Cells
RT	Room Temperature
Sec	Second(s)
SDS	Sodium Dodecyl Sulfate
SjCC	*Schistosoma japonicum* cathepsin C
Sj pmy	*Schistosoma japonicum* paramyosin
SmCB	*Schistosoma mansoni* cathepsin B
SmCC	*Schistosoma mansoni* cathepsin C
SmCL1	*Schistosoma mansoni* cathepsin L1
SmCL2	*Schistosoma mansoni* cathepsin L2
Sm28 GST	*Schistosoma mansoni* Glutathione-S-Transferase
WHO	World Health Organisation
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Z-Phe-Arg-NHMec	H-Phenylalanine-Arginine-NHMec
Z-Phe-Ala-NH_2	Z-Phenylalanine-Alanine-NH_2
Z-Phe-Phe-OH	Z-Phenylalanine-Phenylalanine-OH
V	Volts

1: INTRODUCTION

1: INTRODUCTION

1.1: PARASITES

Parasitism is defined as an intimate and obligatory relationship between two organisms during which the parasite, usually the smaller of the two partners, is metabolically dependent on the host (Cheng *et al.*, 1986). Parasitism is described as being "the most common life style on Earth" and parasites by far outnumber the free-living species (Windsor, 1998). Parasitism is a way of life that transcends all phylogenetic boundaries. Parasites are considered ubiquitous among all plant and animal groups and it has been estimated that at least 50% of all plant and animals are parasitic at some stage during their life cycles (Bush *et al.*, 2001). The World Health Organisation (WHO) has stated that of the six major unconquered human diseases, five of them, namely filariasis, leishmaniasis, malaria, schistosomiasis and African trypanosomiasis are parasitic in the traditional sense. At least 500 million people on Earth - one living person in ten - is afflicted by one or more of the eight major tropical diseases in which parasites are implicated (Godal, 1996).

Typically parasites can be endoparasites that reside within their hosts, in the gut, body cavity, lungs or other tissues. Endoparasites nearly always live a completely parasitic existence. Conversely, ectoparasites are organisms such as fleas, lice and ticks, which live outside their hosts usually attached to skin, feathers or hair. They can never attain a completely parasitic existence as they utilise oxygen from outside the host (Smyth, 1976). A parasite that has an absolute dependence on the host during some or all of its life cycle is referred to as an obligatory parasite. Alternatively a parasite can be a facultative parasite, which does not solely depend on the parasitic way of life but can adapt to it if placed in such a relationship (Smyth, 1976; Cheng, 1986). A parasite can have many different types of host during its life cycle. A host in which a

parasite produces progeny by sexual means is referred to as the definitive host. An intermediate host is needed by the parasite to complete its life cycle and the parasite will undergo physiological and morphological change in this host. However, despite the fact that asexual reproduction may occur in the intermediate host, sexual reproduction will not occur within this host (Fernandez and Esch, 2001).

Parasitic infections can be classified into two groups, those that are caused by protozoa and those attributed to the presence of helminths. Three hundred and forty-two helminth parasite species have been detected in humans (Bush *et al.*, 2001). Helminths infections result in morbidity as opposed to mortality with disease severity related to worm burden (Maizels *et al.*, 1993). The phylum Platyhelminthes is a large and diverse group. The class Trematoda (flukes) to which the genus *Schistosoma* belongs, also referred to as digeneans or digenetic trematodes, generally possess a complex life cycle. Most flukes are hermaphrodites, the schistosomes being the exception. Those that infect man are found typically in the bile ducts, lungs, intestine and blood depending on the species (Seaton, 1979).

1.1.1: Parasite Vaccines

To date no vaccines against any eukaryotic parasite of man are available although several highly effective vaccines against parasites of veterinary importance exist (Vercruysse *et al.*, 2001). Vaccines to protect against human parasitic infections are not yet available for a number of reasons. The use of attenuated organisms, which where viral infections are concerned have proven to be the most successful approach, is not feasible for parasites owing to the impracticability of culturing large numbers of parasites (Ada, 1993). Parasites cannot be cultured *in vitro* as easily as prokaryotic organisms can. Most of the successful veterinary vaccines currently available contain live parasites whose virulence has been attenuated by irradiation, *in vitro* culture or *in vivo* passage.

They are difficult to produce, have a short shelf life and must be kept refrigerated, which makes large-scale application difficult to install (Vercruysse *et al.*, 2001). By virtue of the fact that parasites possess a multi-stage life cycle producing distinct antigens at each stage, in order for a vaccine to be effective it would have to evoke protection against more than one infectious stage. Many parasites are capable of manipulating immune responses to their favour (Mulcahy *et al.*, 1999) and possess an efficient means for evading host immune responses. Some of the ways in which this is accomplished include:

- Molecular mimicry: The possession of endogenously produced host-like or host acquired molecules, which confers "self" status upon the parasite.

- Genetic complexity: Parasites have larger genomes than their eukaryotic counterparts and this is expressed in the variety of complex methods by which parasitic organisms are able to avoid or subvert the host's immune defences (Vercruysse *et al.*, 2001).

- Antigenic variation: Stage specific antigen presentation. For example, the tegumental membrane is periodically sloughed off in schistosomes and so too are any bound immunoglobulins.

- Innate lack of immunogenicity of relevant epitopes.

- Subversion of immune responses e.g., by cleaving bound immunoglobulins (Newport and Colley, 1993).

In the case of any infectious disease, if it can be demonstrated that immunity to reinfection occurs after a natural infection then hope exists that a suitable vaccine can be developed. Schistosomes can live in humans for up to thirty years (Bush *et al.*, 2001) and establish a long-term relationship with their host. Chronic parasitic infections, in which the immune responses are incapable of

effecting long lasting immunity to continuing infection or re-infection are the norm rather than the exception in parasitic diseases (Vercruysse *et al.*, 2001). Protective immunity may occur only after many years or decades of exposure to the parasite (Maziels *et al.*, 1993). Irrefutable evidence exists which shows that people living in endemic areas do develop various degrees of immunological protection against re-infection (Hagan *et al.*, 2000). The nature of immunity in humans is thought to involve an IgE production mediated mechanism which could prove problematic since a vaccine designed to promote IgE production could elicit undesirable side-effects such as exacerbation of allergy (Wilson and Coulson, 1998). In the case of some helminth infections "concomitant immunity" may occur where an initial infection is not eliminated but becomes established and the host then acquires resistance to invasion by new worms of the same species (Bradley, 2001). Another obstacle to the successful development of a parasite vaccine is the fact that animal models of disease do not correlate with human diseases. Cell mediated immunity involves helper T cells, which are phenotypically divided into Th1 and Th2 subsets that differ in the cytokines they produce. An over simplification is that Th1 responses mediate in the killing of intracellular parasites and that Th2 responses eliminate extracellular ones. Th1 and Th2 subsets have contrasting and cross-regulating cytokine profiles (Bradley, 2001). It is generally accepted for example that in humans, primates and rats protective immunity to schistosomes is associated with Th2 responses involving the production of interleukin (IL) 4, 5, IgE and eosinophils. In murine schistosomiasis immunity is associated with a Th1 response involving IL 2 and 12 and interferon-gamma (IFN-γ). However, different responses have been observed within the same host species, which is indicative of the complex interaction between the parasite and the host (Capron, 2001). It has been argued that the Th2 subset (IgE) has evolved to help combat helminth infections. Therefore, perhaps helminths have evolved mechanisms that allow them to counteract Th2 responses (Maizels *et al.*, 1993).

Aided by recombinant DNA technology, many novel parasite proteins have been and will continue to be isolated and cloned and this will undoubtedly contribute enormously to the pool of information that already exists about these parasites. It is only as a result of much research into the basic cell biology and immunology of parasites that advances can be made in attaining the goal of producing an efficacious parasite vaccine.

1.2: SCHISTOSOMIASIS

Schistosomiasis (bilharzia) is one of the world's most prevalent parasitic infections and has significant economic and public health consequences. It is caused by digenetic blood trematodes that belong to the family Schistosomatidae (Smyth, 1976). First identified in 1851 by Theodor Bilharz to be the cause of African schistosomiasis, schistosomiasis is considered to be an ancient disease of man (Jordan, 2000; Ross, 2002). Old papyrus records from Egypt depict the incidence of the disease around 2,000 B.C. and calcified schistosome ova from two mummies dating from 1,250-1,000 B.C have also been discovered. Around 600 million people are at risk in 74 countries particularly in urban areas, refugee camps and where water resources are being developed (Godal, 1996; Ross, 2002). It is estimated that 200 million people are infected, 120 million are symptomatic and 20 million have severe disease (Chitsulo, 2000).

1.2.1: CAUSATIVE SPECIES AND GEOGRAPHICAL LOCATION

The major forms of human schistosomiasis are attributable to five species of blood fluke:

***Schistosoma mansoni*:** Causes intestinal schistosomiasis in 53 countries in Africa, the East Mediterranean, the Caribbean and South America.

***Schistosoma japonicum*:** (including *Schistosoma mekongi* in the Mekongi river basin) causes oriental or Asiatic intestinal schistosomiasis. This species is endemic in seven countries in South East Asia and in the Western Pacific region.

***Schistosoma intercalatum*:** Also causes intestinal schistosomiasis and has been linked to ten central African countries.

***Schistosoma haematobium*:** Gives rise to urinary schistosomiasis endemic in 54 countries in Africa and the Eastern Mediterranean.

(Adapted from WHO fact sheet)

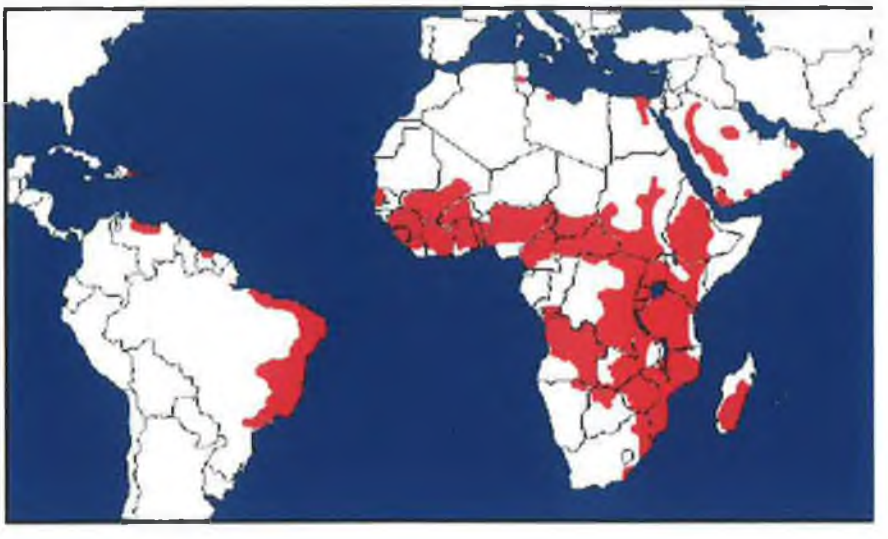


Figure 1: Geographical Distribution of *S. mansoni*

Figure 2: Geographical Distribution of *S. japonicum*

1.2.2: SCHISTOSOME LIFE CYCLE

An alternation of generations occurs during the schistosome life cycle, with asexual reproduction taking place in a susceptible freshwater snail host and sexual reproduction occurring in a warm blooded vertebrate (Newport and Colley, 1993).

- All schistosoma infections occur as a result of direct contact with fresh water that harbours free-living larval forms of the parasite known as cercaria (Ross, 2002).

- A muscular tail mediates locomotion of cercaria. When a suitable host is contacted the cercaria penetrates the skin and looses its tail. It is thought that cercariae possess a number of external sensory structures which regulate a number of tropisms that attracts them to areas where the definitive host is likely to be located (Newport and Colley, 1993).

- Cercariae, now called schistosomulae, reach the venous circulation either directly or via the lymph vessels. In the venous system schistosomulae

are carried to the lungs within 4 days and reach their highest concentration at 7-9 days.

- By the 15th day the schistosomules can be found in the liver where they feed on portal blood and undergo rapid growth (Smyth, 1976; Cheng, 1986).

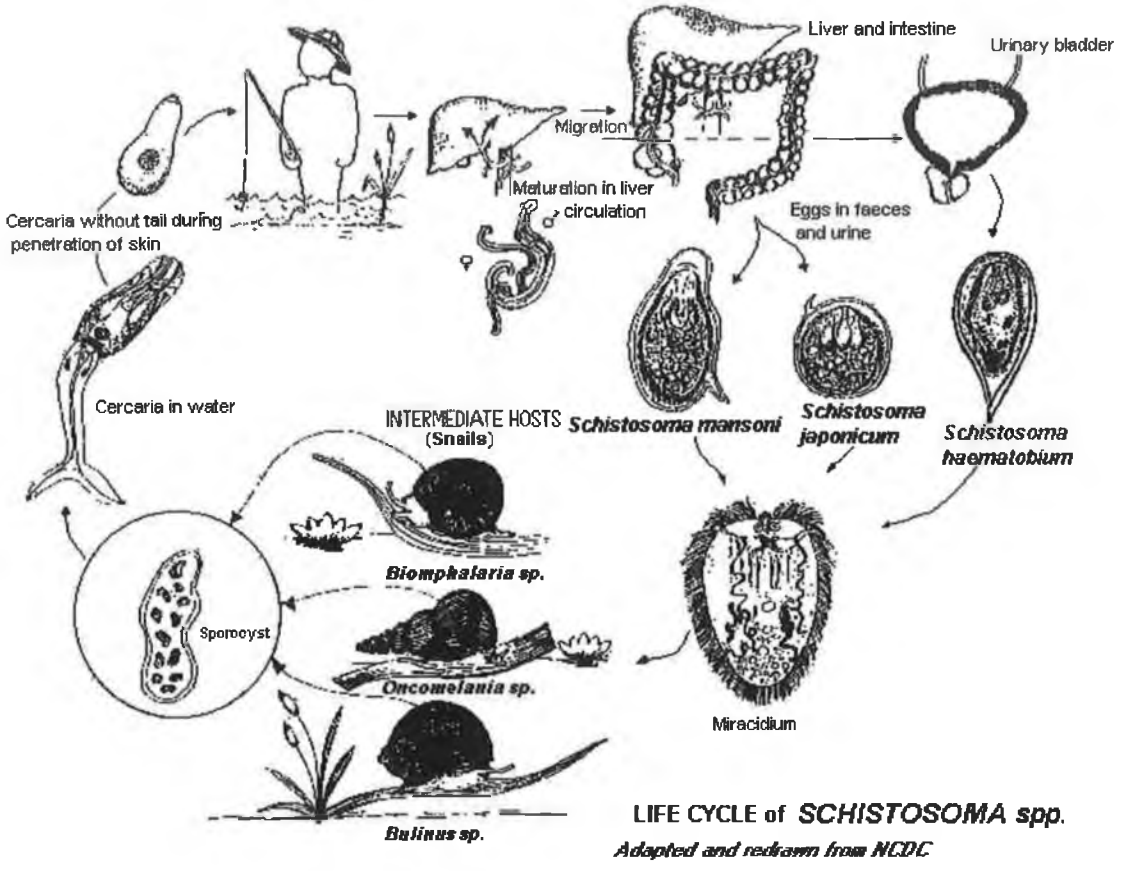


Figure 3: Schistosome life cycle

- Depending on the species, the worms then migrate to the superior mesenteric veins (*S. mansoni*), the inferior mesenteric and superior haemorrhoidal veins (*S. japonicum*) or the vesical plexus and veins draining the ureters (*S. haematobium*). Egg production commences 4-6 weeks after infection and will continue for the life of the worm which is usually 3-5 years although some worms can live for up to 30 years (Bush

et al., 2001; Ross, 2002). The mature female will lay between 200-300 eggs per day depending on the species (Newport and Colley, 1993). Fertilisation will occur when spermatozoa from the male are introduced into the female. The male is broader than the female and its body is curved to form a groove, called the gynecophoral canal, in which the female is snugly held (Cheng, 1986).

- Eggs will pass from the lumen of the blood vessels into adjacent tissues. Many will exit the body through the intestinal or bladder mucosa and are shed in the faces (*S. mansoni* and *S. japonicum*) or urine (*S. haematobium*) (Ross, 2002).

- The eggs will hatch upon contact with water releasing free-living ciliated miracidia. Successful miracidia penetrate the soft tissues of the intermediate snail host. *S. mansoni* will infect *Biomphalaria* species, *S. japonicum* infects *Oncomelania* species and *S. haematobium* infects *Bulinus* species of freshwater snails. Inside the host asexual reproduction will occur resulting in the generation of mother sporocysts, which in turn give rise to several generations of daughter sporocysts (Cheng, 1986; Newport and Colley, 1993; Ross, 2002).

Daughter sporocysts will eventually produce large numbers of cercaria which will exit the snail and the cycle will be repeated when the cercaria come into contact with the skin of a suitable human host (Cheng, 1986).

1.2.3: Disease Symptoms and Pathology

Schistosomiasis brings about intermittent morbidity and chronic infections, which may be potentially fatal. After initial skin penetration by cercaria, "swimmer's itch" a characteristic rash or dermatitis develops. This condition is also attributable to skin penetration by cercaria of blood flukes that usually

infect only avian species. The cercaria will be eliminated by the human host's immune system in a matter of days (Seaton, 1979; Bush *et al.*, 2001).

The onset of early schistosomiasis is characterised by an explosive rise in temperature known as katayama fever. The most common manifestations are fever, chills, weakness, weight loss, headache, nausea, vomiting, diarrhoea, hepatomegaly, splenomegaly and marked eosinophilia (Lambertucci, 2000). However, in some cases the acute phase of infection may prove to be asymptomatic. Chronic schistosomiasis results from the immune response of the host to schistosome eggs and the granulomatous reaction evoked by the antigens they secrete (Ross, 2002). As a direct result of frequent re-infection repeated penetration of the intestinal and bladder walls results in extensive scar formation which prevents migration of the eggs through these structures. Because of this many eggs will be transported to the liver and occasionally to other sites such as the brain and spinal cord (Cheng, 1986). Wherever the eggs become lodged local inflammation, fibrosis and necrosis will occur. Hepatosplenic schistosomiasis is the most usual manifestation of chronic disease, which is characterised by enlargement of the liver and spleen (Lambertucci, 2000). As the disease progresses the liver recedes and becomes contracted and cirrhotic but the spleen will continue to enlarge, resulting in a bloated abdomen. If eggs become lodged in the blood vessels congestive heart failure and aneurysms may result (Smyth, 1976).

1.2.4: TREATMENT

Due to the fact that no effective schistosomiasis vaccine is available, to date the most significant step towards control has been the availability of a safe and efficacious drug, praziquantel. Discovered in the mid 1970's praziquantel's precise mode of action on adult worms is unknown although it is believed to cause tetanic contractions and tegumental vacuoles causing the worm to detach from the wall of the vein and die (Ross, 2002). Praziquantel is active against

all five schistosome species that can infect humans. Another drug, artemeter, commonly used in the treatment of malaria has also been shown to possess anti-schistosomal properties by acting on the juvenile stages of the parasite. It has been reported that combined therapy using high doses of both praziquantel and artemeter showed very high worm reduction rates of 90% and above (Utzinger, 2001). From 1976-1979 the Brazilian government implemented an ambitious programme of schistosomiasis control known as PECE (Special Programme for Schistosomiasis Control), which recommended the use of integrated control measures such as education, construction of sanitary facilities and mass chemotherapy. Mass chemotherapy prevailed as the most successful control strategy. The morbidity of schistosomiasis decreased since the wide spread use of chemotherapy and a decrease of 47% in mortality rate between 1977-1994 was observed (Lambertucci, 2000).

Evidence is now emerging of resistance to praziquantel. There are already reports of low efficacy of the drug in Egypt and Senegal (Wilson and Coulson, 1998). Similarly in regions of Kenya where there has been heavy exposure to praziquantel there are reports of *S. mansoni* and *S. haematobium* infections that are not responsive to multiple courses of treatment (Ismail *et al.*, 1999; William *et al.*, 2001). Another drug, oxamniquine has also proved efficacious in the treatment of schistosomiasis, but as is the case for artemeter, its limited availability means that there is essentially only one drug widely available for the treatment of schistosomiasis. Therefore resistance to praziquantel would have serious implications for schistosomiasis control (Hagan *et al.*, 2000). Rapid re-infection often follows chemotherapy and drug delivery itself requires a substantial infrastructure to cover all parts of an endemic area regularly (Capron, 1998). Another problem with chemotherapy as a control measure is that in a minority of individuals with a high worm burden, transition to severe and irreversible pathology is usually diagnosed at too late a stage for chemotherapy to affect the outcome (Wilson and Coulson, 1998).

By virtue of the fact that chemotherapy only provides a short-term solution to the problems posed by schistosomiasis it is vital to explore all other avenues in attempts to find a suitable means of controlling the disease such as vaccination. A vaccination programme would be cheaper in the long term and would produce less strain on services than drugs, which have to be constantly re-distributed in endemic areas (Hagan *et al.*, 2000). However developing a schistosomiasis vaccine is by no means a simple task. Efforts are hindered by the long-term relationship helminths establish with their hosts. Mature schistosomes inhabit the bloodstream, potentially the most hostile environment in the body yet they are long lived parasites primarily due to their highly evolved evasion mechanisms (refer to 1.1.1). To date at least 100 different schistosome antigens have been partially or fully characterised. Among the antigens selected on the basis of their immunogenicity, a significant proportion are muscle proteins or enzymes (Capron, 1998).

1.2.5: Potential Schistosomiasis Vaccine Candidates

Progress in identifying and synthesising potential immunogens by recombinant DNA technology together with advances in immunology and cell biology have prompted a feeling of optimism among parasitologists that effective vaccines for parasites of medical and clinical importance, including schistosomiasis are within reach. With regards to schistosomiasis, in the mid 1990's six vaccine candidate molecules were selected by the WHO Special Programme for Research and Training in Tropical Diseases (WHO/TDR) for independent testing. These were:

- 28 kDa glutathione S-transferase
- 97 kDa paramyosin
- 62 kDa Irv-5, an irradiated larva associated vaccine derived from a 200 kDa molecule with extensive sequence homology with human myosin.
- 28 kDa triose phosphate isomerase
- 23 kDa integral membrane antigen

- 14 kDa fatty acid binding protein

Two of the most promising of the above candidates have proven to be glutathione S-transferase and paramyosin (Doenhoff, 1998; Bergquist, 2002).

Glutathione-S-Transferases (GSTs) are multi-functional enzymes that play an important role in the detoxification of xenobiotics and function as intracellular binding proteins (Scott and McManus, 2000). In parasitic helminths it has been proposed that GSTs help protect the parasites from host immune attack and could play a key role in the passive detoxification of antihelminthic compounds and hematin, the end product of haemoglobin digestion by blood flukes (Howell *et al.*, 1988; Brophy and Barrett, 1990). Early reports of a significant level of protection of rats and mice immunised with an *S. mansoni* antigen of molecular mass 28 kDa against a natural challenge infection with live cercaria stimulated much in depth study of this antigen, later shown to be a GST (Balloul 1987 a and b). In the case of Sm28 GST a reduction in worm burden of 38% was observed in baboons on receiving three vaccinations of rSm28 GST. When two vaccinations with Sm28 GST in aluminium hydroxide and *Bordtella pertussis* were administered, female schistosoma fecundity was reduced by 33% with a more pronounced effect (66%) on faecal egg output (Boulanger, 1991). Protection was attributed to specific IgA antibody responses to rSm28 GST which lead to impaired schistosome fecundity by limiting both the egg laying of mature worms and the hatching capacity of schistosome eggs into viable miracidia (Grzych *et al.*, 1993). Phase I clinical trials in normal human volunteers in endemic areas have been undertaken on a *S. haematobium* GST called Bilhvax (Capron, 1998; Hagan *et al.*, 2000; Ross, 2002). The vaccine was deemed safe and proved to be an excellent immunogen (Hagan *et al.*, 2000). *S. haematobium* was the species of choice for this trial because it has been proven that resistance to *S. haematobium* is associated strongly with immune mediated inhibition of fecundity. Secondly, the methods of qualitative evaluation of eggs in urine are reliable and easier to handle than stool examination and, finally, disease progression in urinary schistosomiasis can be easily followed by non-

invasive methods such as bladder and urinary tract ultrasound tomography (Capron, 1998).

Paramyosin is a 97 kDa myofibrillar protein located in the muscle layers and tegument of schistosomes (Ramirez *et al.*, 1996). Immunisation of mice with native paramyosin from *S. mansoni* or a partial recombinant fragment of paramyosin with *Bacille Calmette Guerin* (BCG) reduced worm burden by 39% and 26% respectively (Pearse *et al.*, 1988). Also, it has been shown that mice immunised with paramyosin without adjuvant were protected by 24-53% against cercarial challenge (Flanigan *et al.*, 1989). A murine monoclonal IgE antibody that recognised *S. japonicum* paramyosin (Sj pmy) was shown to confer 19-58% protection against *S. japonicum* cercarial challenge following passive transfer into mice (Kojima *et al.*, 1987). This prompted further evaluation of Sj pmy as a vaccine candidate. Unlike the African schistosomes, of which humans are the sole definitive hosts, Asian schistosomes can also infect buffaloes, cattle, dogs, pigs and sheep, which poses an additional problem for the implementation of control measures (Kalinna *et al.*, 1997; McManus *et al.*, 2002). However, a unique control opportunity exists, by immunising reservoir farm animals, in particular the water buffalo, a known culprit in the transmission of Asian schistosomiasis. Subsequently, it was illustrated by the use of a bacterially expressed Sjpmy that high levels of specific anti-pmy IgG antibodies, significant reduction in worm burden and liver eggs could be obtained (McManus *et al.*, 2002). Limitations to the widespread use of rSj pmy are costs of scale-up and the fact that the low expression levels of bacterially produced Sj pmy make large-scale trials impossible. Nevertheless, solutions to these problems are under investigation and if these obstacles can be surmounted then it is feasible that a rSj pmy vaccine could control Asian schistosomiasis.

Nucleic acid vaccines have been tested as potential vaccine candidates in the fight against schistosomiasis. A nucleic acid vaccine has been used to induce protective immune responses to *S. mansoni* integral membrane protein Sm23

(Da'dara *et al.*, 2002). Sm23 is a member of the "tetraspanin" family of proteins, which possess four hydrophobic trans membrane domains. It is expressed in all stages of the schistosome life cycle. In a recent trial, on cercarial challenge of mice immunised with a plasmid carrying cDNA for Sm23, a 21-44% reduction in worm burden was observed (Da'dara *et al.*, 2002). Another study employing a cocktail DNA vaccine comprising four plasmids encoding four different *S. japonicum* antigens, namely Sj62, Sj28, Sj23 and Sj14-3-3 showed significant resistance (34-45%) against *S. japonicum* cercarial challenge infection (Zhang, 2002). A DNA vaccine containing the cDNA coding for *S. mansoni* legumain, asparaginyl endopeptidase (Sm32) cloned in a mammalian expression vector under the control of the CMV promoter/enhancer was shown to have an anti-fecundity effect on immunisation of mice. Female worms of a challenge infection produced 37% less eggs than those grown in naïve or non-vaccinated mice (Chlichlia, 2002).

Vaccines derived from radiation attenuated cercaria have to date been the most effective of all schistosomiasis vaccine candidates tested. In rodents and mice vaccination with radiation attenuated cercaria resulted in a 50-80% reduction in challenge worm burden (Coulson, 1997). In a small scale study using non-human primates (chimpanzees), the protective effect of vaccination with radiation attenuated *S. mansoni* cercaria was manifested as an amelioration of acute disease and overall morbidity and a 38% reduction in faecal egg output upon challenge infection (Eberl, 2001). Radiation attenuated cercarial vaccines could form the basis for the development of a recombinant vaccine if the relevant immune mechanisms and protective antigens were known.

The aforementioned schistosomiasis vaccine candidates together with the haemoglobin degrading proteinases, also potential vaccine candidates (refer to 1.2.6), illustrate the concentration of international effort that has been invested into the development of a vaccine against schistosomiasis. Given the encouraging results, along with advances in immunology and parasitology, the

prospect of a future schistosomiasis vaccine appears to be both feasible and attainable (Wilson and Coulson, 1998; Ross, 2002).

1.2.6: Digestion of Haemoglobin by Schistosomes

Developing and mature schistosomes are obligate blood feeders. It has been estimated that male and female *S. mansoni* ingest 39,000 and 330,000 red blood cells (RBCs) per hour respectively (Dalton *et al.*, 1995). The host haemoglobin (Hb) is degraded into usable amino acids by a number of proteases present in the alimentary tract of the worm and is used by the parasite for its growth, development and reproduction. Evidence for this comes from the discovery of haematin, a dark pigment from the end product of Hb digestion in the digestive tract of adult worms (Brindley *et al.*, 1997).

- In the oesophagus of the schistosome, ingested RBCs are lysed by the action of haemolysin.

- Hb released from the RBCs flows into the caecum and gastrodermis of the schistosomes where it is degraded further by endopeptidases such as cathepsin B, cathepsin L1 and cathepsin D, which cleave the protein into polypeptides.

- Next, these polypeptides are acted on by exopeptidases such as cathepsin C, which break them into dipeptides or free amino acids that are metabolised by the gastrodermal cells.

- Haematin, the end product of Hb digestion will be regurgitated by the parasite and will accumulate in the liver of infected hosts.

(Brindley *et al.*, 1997)

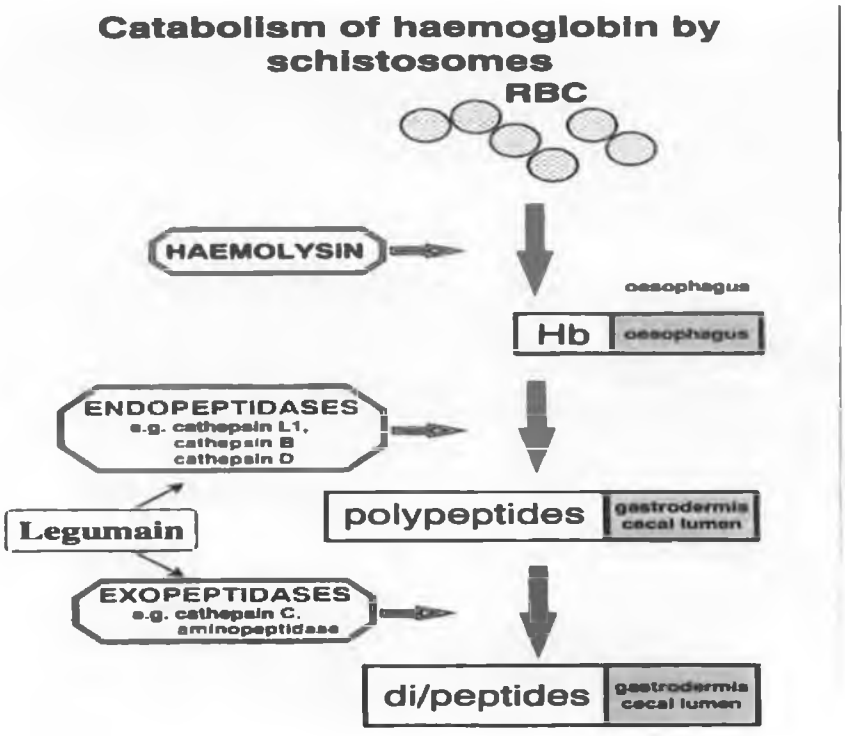


Figure 4: The digestion of haemoglobin by schistosomes (taken from Tort *et al.*, 1999)

Schistosome legumain also plays an indirect role in the digestion process. It is thought that it may play a role in the activation of the proteolytic enzymes pivotal to the digestion of Hb such as cathepsin C, D and L.

It is believed that immune responses directed against any of these digestive enzymes could block the digestive process to the detriment of the parasite thus providing an effective vaccine against schistosomiasis. Such a vaccine could succeed where others have failed because it would be targeting one of the most, if not the most, vital process in the life cycle of the parasite. By preventing the fluke from feeding, its life support will effectively be cut off and it will die before eggs can be produced and before it can cause any damage to the host.

1.3: PARASITE PROTEINASES

Proteolytic enzymes or proteinases are enzymes that catalyse the cleavage of peptide bonds in other proteins (Neurath, 1984). It is accepted that proteolytic enzymes have their origins very early in evolution as they are required by all organisms even the most primitive for digestion and metabolism of their own proteins. Although proteinases of all the four major groups of peptidases: serine, aspartic, cysteine and metalloproteinases, have been characterised in parasitic helminths, by far the largest number reported to date belong to the papain superfamily of cysteine proteinases (Tort *et al.*, 1999).

1.3.1: CYSTEINE PROTEINASES

Cysteine proteinases are widely distributed among living organisms. Cysteine proteinases play numerous indispensable roles in the biology of parasitic organisms. Thought to play a key role in parasite immunoevasion, excystment/encystment, digestion, exsheating and cell and tissue evasion, proteinases comprise approximately 2% of all expressed genes with little variance between organisms (Sajid and McKerrow, 2002). Cysteine proteinases may be subdivided into groups on the basis of their primary structure. The two major superfamilies have diverged from different ancestral genes: the papain superfamily consisting of papain, cathepsin B, C, H, L and calpain, and the superfamily that includes viral cysteine proteinases (Ishidoh *et al.*, 1991). The majority of cysteine proteinases belong to the papain family (clan CA, family C1). The papain superfamily consists of papain and related plant proteinases such as chymopapain, caricain, bromelain, cruzipain and related parasite proteases and of lysosomal cathepsins (Turk *et al.*, 2000). The papain like proteases consists of 2 domains, the L- (left) and R- (right) domains. The central helix of about 30 residues in length is the most prominent feature of the L-domain whereas the fold of the R-domain is based on the β-barrel motif. The two-domain interface opens on top to form a "V" shaped active site cleft. The

two active site residues Cys 25 and His 159 are located in the middle of the cleft, Cys being positioned at the N-terminus of the central helix from the L-domain and His 159 as part of the R-domain β-barrel structure (Sajid and McKerrow, 2002).

Mode of Action

The mechanisms by which cysteine proteinases hydrolyse peptide bonds are as follows (Sajid and McKerrow, 2002):

- Prior to hydrolysis of a peptide bond, the proteinase binds the protein or peptide substrate in its active site.

- An essential cystidine Cys 25 is found at the active site. The thiol group is enhanced as a nucleophile because of its close proximity to the active site histidine, which acts as a proton donor/general base. Sequences that directly flank the active site cystidine and histidine are highly conserved to maintain catalytic register.

- The sulfhydryl (SH) of the cystidine side chain and the imidazole of histidine give rise to a thiolate-imidazolium charge relay diad. This is stabilised by a highly conserved asparagine residue.

- A highly conserved glutamine forms the oxyanion hole, which is crucial to the formation of an electrophilic centre to stabilise the tetrahedral intermediate during hydrolysis.

The evolutionary history of proteinase families may be regarded as the evolution from a single, general-purpose ancestral proteinase to multiple, increasingly specific paralogous enzymes through the process of repeated gene

duplications. This process has resulted in at least 12 paralogous cysteine proteinases and possibly many more (Neurath, 1984; Berti and Storer, 1995).

1.3.2: Lysosomal Cysteine Proteinases

Lysosomal cysteine proteinases, also known as cathepsins were discovered in the first half of the 20th century. There are 11 human cathepsins (B, C, F, H, K, L, O, S, V, W, X) (Sajid and McKerrow, 2002). Cathepsin C (dipeptidyl peptidase I) was the first pure lysosomal cathepsin to be isolated in the 1940's (Gutman and Fruiton, 1948). It wasn't until the 1970's that more cathepsins (B, H and L) were isolated and the first amino acid sequences of mammalian cathepsins did not appear until the 1980's when the sequences of rat cathepsin C and H were published (Barrett *et al.*, 1998). The principal function of the lysosomal proteases is considered to be the non-selective degradation of proteins inside lysosomes. Occasionally, cathepsins have been found outside the lysosomes, most often resulting in a pathological condition (Turk *et al.*, 2000). Lysosomal cysteine proteinases are optimally active at a slightly acidic pH. Inactivation at neutral pH is common to all lysosomal cysteine proteinases an exception being cathepsin S. The enzymes are monomers of ~ 30 kDa with the exception of cathepsin C which is a tetramer of 200 kDa (Turk *et al.*, 2001 and 2002). Catalysis can be initiated either within a polypeptide chain (endopeptidases) or from amino or carboxyl ends (exopeptidase). All cathepsins are initially synthesised as inactive zymogens or preproenzymes consisting of a pro-region, which may serve as an intramolecular chaperone to assist in protein folding and as an endogenous inhibitor to regulate protease activity. Additionally a leader or signal sequence composed of 15-22 hydrophobic amino terminal sequences will target the protease to its intracellular compartment or will act as a secretion signal (Sajid and McKerrow, 2002).

1.3.2.1: CATHEPSIN C, DIPEPTIDYL PEPTIDASE I (DPP I)

In humans cathepsin C (DPP I) serves a vital role in the activation of a number of granule serine proteases such as granzyme A and B exclusively expressed in the granules of activated cytotoxic lymphocytes (Pham and Ley, 1999), in cell-mediated apoptosis, inflammation and connective tissue remodelling (Olsen *et al.*, 2001). It has been suggested that cathepsin C may be essential for establishing and maintaining the structural organisation of the epidermis of the extremities and the integrity of the tissue surrounding the teeth and may participate directly in the processing of proteins such as keratins (Nuckolls and Slavkin, 1999). Recently a mutation in the cathepsin C gene was shown to be implicated in Papillon-Lefèvre syndrome, an autosomal recessive disorder that manifests as severe peridontitis and palmoplantar keratosis. Peridontitis leads to premature tooth loss whereas keratosis affects the knees and elbows (Toomes *et al.*, 1999; Hart *et al.*, 2000). Spleen and kidney provide the richest sources of cathepsin C in humans. Serum levels of cathepsin C were highest in hepatic diseases, peripheral arterial disease, thromboembolism, myocardial infarction, diabetes mellitus and prostatic hypertrophy (Dolenc *et al.*, 1995). Cathepsin C has also been suggested to play a role in chronic airway diseases such as asthma (Turk *et al.*, 2001).

Cathepsin C functions in the removal of dipeptides from the amino terminus of peptide and protein substrates and is active in the pH range of 3.5-8.0. It is an exopeptidase with broad specificity. Only peptides with a basic residue at the N-terminus, a blocked amino group at the N-terminus, or a proline on either side of the scissile bond are not hydrolysed (Paris *et al.*, 1995; Lauritzen *et al.*, 1998 and Cigic and Pain, 1999). The enzyme's requirement for halide ions and sulfhydryl reagents such as dithiothreitol (DTT) is well known. Cathepsin C is inhibited by rat stefin A and chicken cystatin, two inhibitors of cysteine proteinases from the cystatin family and also by general cysteine protease inhibitors such as leupeptin and NEM (Dolenc *et al.*, 1995).

Cathepsin C is unusual in that it has a very long pro-peptide of ~ 200 amino acids due to extension at the N-terminus. A large part of this pro-region, called the residual propart remains attached in the mature enzyme, which is unique to cathepsin C (Cigic and Pain, 1999; Horn *et al.*, 2002). The N-terminal portion of the propeptide bears no sequence similarity to any of the other cysteine proteases. However, the C–terminal portion of the propeptide is homologous with propeptides of other cysteine proteases (Hola-Jamriska *et al.*, 1998).

Cathepsin C is the only member of the papain superfamily to exist as a tetramer of four identical subunits. During activation the structure of cathepsin C undergoes a series of transformations, from the monomeric form of the preproenzyme via a dimeric form of proenzyme to the tetrameric form of the mature enzyme (Muno *et al.*, 1993). Each subunit or monomer consists of an N-terminal fragment, a heavy chain and a light chain. The N-terminal fragment also called the residual pro-part or exclusion domain is suggested to be involved in the formation of the tetramer (Dolenc *et al.*, 1995; Turk *et al.*, 1995; Cigic *et al.*, 2000; Dahl *et al.*, 2001). The C terminal part of the propeptide functions as the activation peptide and is liberated from the enzyme during maturation (Dahl *et al.*, 2001). Procathepsin C is not capable of autoactivation despite the fact that many members of the papain superfamily such as cathepsin B and papain possess this quality. The dimeric cathepsin C precursor, which can be activated 2000 fold with papain can also be activated by cathepsin L and S (Dahl *et al.*, 2001).

The cDNA sequences encoding bovine, canine, human, murine, rat and schistosome cathepsin Cs have been determined (Ishidoh *et al.*, 1991; Butler *et al.*, 1995; Paris *et al.*, 1995; McGuire *et al.*, 1997; Hola-Jamriska *et al.*, 1998; Wolters *et al.*, 1998; Frye *et al.*, 2000). From the deduced sequence of all cathepsin Cs, a distinct tyrosine is found adjacent to the active site cysteine. This tyrosine motif appears to be unique to the papain superfamily in which tryptophan usually occupies the position adjacent to the active site residue. The

conservation of this tyrosine motif among all cathepsin Cs suggests that it plays a role in the determination of substrate specificity (McGuire *et al.*, 1997). Phylogenetic analysis of cathepsin C sequences has also revealed that the closest evolutionary relation to cathepsin C is cathepsin B (Berti and Storer, 1995; Tort *et al.*, 1999).

1.3.3: PROTEASES OF SCHISTOSOMES

Cysteine proteases such as cathepsins B, C, D and L and asparaginyl endopeptidases (legumain) are all implicated in the digestion of haemoglobin by schistosomes (refer to Section 1.2.6). The main proteases involved in this process will be discussed briefly:

1.3.3.1: CATHEPSIN B

Cathepsin B belongs to the papain superfamily of cysteine proteases and is localised in the lysosome. Cathepsin B is synthesised as a preproenzyme of 340 amino acids with the processed mature enzyme containing 250 residues (Brindley *et al.*, 1997). It has a molecular mass of 31 kDa, which is larger than its predicted molecular mass of 28 kDa due to N-linked glycosylation of the mature form (Tort *et al.*, 1999). *S. mansoni* cathepsin B (SmCB) was revealed to be 70% similar to mammalian cathepsin B (Klinkert *et al.*, 1989). SmCB, expressed in an insect system was demonstrated to be able to hydrolyse the cathepsin B specific substrate Z-Arg-Arg-NHMec (Gotz and Klinkert, 1993). Recombinant *S. mansoni* procathepsin B has also been expressed in *Saccharomyeces cerevisiae* (Lipps *et al.*, 1996). The enzyme could not be activated *in vitro*, which suggests that autoprocessing does not occur but implies that another enzyme could be responsible for activation. This theory was confirmed when it was shown that porcine pepsin could activate recombinant SmCB. However, in our laboratory *Pichia pastoris* expressed

cathepsin B was shown to autoactivate at low pH (Cunneen, M, unpublished data).

1.3.3.2: CATHEPSIN C

S. mansoni cathepsin C (SmCC), which is the focus of this thesis, shares many of the characteristics of mammalian cathepsin C referred to in Section 1.3.2. Butler *et al.*, 1995, first reported the cDNA sequence of SmCC. The deduced amino acid sequence of preprocathepsin C is 454 amino acids in length, which comprises a signal peptide of 24 amino acids, a long propeptide of 193 amino acids and a mature enzyme of 237 residues (Brindley *et al.*, 1997). Using the cathepsin C specific substrate H-Gly-Arg-NHMec, cathepsin C from schistosome extracts was observed to have a pH optimum of 5.5 (Hola-Jamriska *et al.*, 1999). The cDNA sequence of *S. japonicum* cathepsin C (SjCC) was reported by Hola-Jamriska *et al.*1999. The preproenzyme shared only 59% identity with SmCC, which suggested that they were discrete enzymes rather than species homologues. Phylogenetic analysis using the mature enzyme sequence illustrated the common ancestry shared by cathepsin C and B. The proenzyme of SjCC was expressed in a baculovirus expression system. The purified recombinant enzyme resolved in reducing SDS-PAGE gels as three forms of 55, 39 and 38 kDa all of which were reactive with antiserum raised against bacterially expressed SjCC (Hola-Jamriska *et al.*, 2000).

1.3.3.3: CATHEPSIN L

Two cathepsin L proteinases are expressed and secreted by adult schistosomes, cathepsin L1 and cathepsin L2. Analysis of the primary structure revealed that both enzymes are only 44% similar and cathepsin L2 showed more identity (52%) with human cathepsin L than with schistosome cathepsin L (Dalton *et al.*, 1996). Differences in their active sites, propeptide region and potential glycosylation sites would indicate that each enzyme has a separate function.

When the dominant protease activity in soluble extracts and ES products of Sm and Sj was examined, cathepsin L was seen to be the dominant proteases at pH 4.0-6.0 (Brindley *et al.*, 1997). The specific activity for cathepsin L substrates was observed to be 60 fold greater than that for cathepsin B specific substrates (Dalton *et al.*, 1995). *S. mansoni* cathepsin L1 (SmCL1) full-length transcript encodes a preprocathepsin L1 of 319 amino acid residues inclusive of a signal peptide. After processing the mature enzyme is 215 amino acids with an apparent molecular mass of ~24 kDa. SmCL2 preprocathepsin consists of 330 amino acids including a signal peptide of 18 residues, a propeptide of 97 and a mature enzyme of 215 amino acids. Active site labelling and immunoblotting revealed SmCL2 to have a molecular mass of 31 kDa (Brindley *et al.*, 1997).

1.3.3.4: SCHISTOSOME LEGUMAIN

Asparaginyl endoproteinases (Sm32 in *S. mansoni*) or legumains are members of a family of proteinases characterised predominately from legumes (Dalton *et al.*, 1995). These enzymes cleave peptide bonds on the carboxyl side of Asn residues except where the Asn occurs at the N-terminus or at the second position from the N-terminus of the polypeptide or where the Asn is glycosylated (Brindley, 1997). Legumains function in the post-translational modifications of storage proteins in legume seeds by cleaving asparaginyl peptide bonds between propeptides and mature proteins. As in plant legumains, schistosome legumain may process other proteins (Dalton and Brindley, 1996). The mRNA of *S. mansoni* legumain encodes a 50 kDa single chain protein of 429 amino acids. The *S. mansoni* and *S. japonicum* legumains are 73% identical and are 30-40% identical to plant legumains (Brindley *et al.*, 1997). A second gene was shown to encode an active legumain as prior to this only inactive legumain had been identified that had the Cys 197 replaced by an asparagine residue. Isolation and sequencing of a number of clones derived from Sm mRNA identified a second gene for Sm32 with the active site Cys present. This was expressed in *P. pastoris* and autocatalysis was shown to

occur at acid pH which resulted in the production of mature, active Sm32 accompanied by the removal of the C-terminal extension and possibly of the N-terminal pro domain (Caffrey *et al.*, 2000).

1.3.3.5: CATHEPSIN D

Aspartic protease or cathepsin D, an integral endopeptidase in schistosomes is optimally active at pH 3.0-4.0. It has been suggested that aspartic protease plays a more crucial, perhaps even the principal role in the process of haemoglobin (Hb) digestion by schistosomes than cysteine proteases (Brindley *et al.*, 1997). It cleaves Hb into numerous tripeptides, tetrapeptides and other fragments. Schistosome cathepsin D expressed in insect cells was demonstrated to autoactivate at pH 3.6 to a ~40 kDa form that cleaved the substrates o-aminobenzoyl-Ile-Glu-Phe-nitro-Phe-Arg-LeuNH_2 and Hb (Brindley et al., 2001). *S. mansoni* preprocathepsin D consists of a signal peptide of 14 residues, a propeptide of 37 amino acids and a mature enzyme of 377 amino acids with a predicted molecular mass of 41 kDa. Schistosome cathepsin D is expressed in the gastrodermal cells of adults and activity is blocked in the presence of pepstatin, a general inhibitor of aspartic proteases (Tort *et al.*, 1999). It has been suggested that schistosomes express at least two forms of cathepsin D like aspartic proteases, one form with a COOH extension and another without this extension (Wong *et al.*, 1997). In a recent study, mice were vaccinated with recombinant Sj cathepsin D aspartic protease expressed in both insect cells and bacteria in order to evaluate vaccine efficacy. Although vaccination did not reduce fecundity mean total worm burdens were significantly reduced in vaccinates by 21-38%. High levels of antibodies, predominantly IgG1, IgG2 and IgG2b were present in vaccinates but no correlation was seen between antibody levels and protective efficacy (Verity *et al.*, 2001).

1.4: HETEROLOGOUS PROTEIN PRODUCTION

Over the past few decades' scientists have learned how to manipulate DNA to identify, excise, move and place genes into a myriad of organisms that are genetically quite different to the original organism from which the gene was taken. Due to these biotechnological advances it is not surprising that hundreds of recombinant proteins have been produced to date in a wide variety of expression systems. Heterologous protein expression refers to the expression of recombinant proteins in cells where they do not naturally occur (Walsh and Headon, 1994).

Four steps are employed in the production of a recombinant protein (Cereghino and Cregg, 1999):

1) The DNA sequence coding for the protein of interest is cloned into a vector (e.g., a plasmid) containing a promoter and transcriptional termination sequences.
2) The next step involves the transformation and stable maintenance of this DNA fusion in the host.
3) Synthesis of the foreign protein will occur when specialised culture conditions are employed.
4) The final step involves the purification of the heterologous protein and comparison with its native counter part.

When a host is been selected for heterologous protein production the following criteria should be satisfied in order to obtain optimum expression of the recombinant protein (Nicaud *et al.*, 1986):

- The desired protein should be produced as a major proportion of the total cell protein.
- The active protein should be secreted at the level of several grams per L.

- Post-translational modifications should be carried out in the host, in particular if the protein is for human use.
- The host should be easy to culture on a large scale, preferably on cheap substrates.
- The host should be free of toxic factors (e.g., endotoxins, pyrogens) especially where the production of therapeutics are concerned.

Early successes in the commercial production of heterologous proteins were achieved using the well-characterised bacterium *Escherichia coli* as a host. However, *E. coli* proved to be inadequate where the production of complex proteins for therapeutic uses was concerned (Buckholz and Gleeson, 1991). To date the majority of therapeutic proteins have been produced in mammalian cell culture systems such as Chinese Hamster Ovary (CHO) cells (Andersen and Krummen, 2002). An extensive choice of expression systems are available each with its own limitations and successes. The choice of expression system is dependent on the final use of the recombinant protein. Some of the common expression systems currently available will now be discussed:

1.4.1: *E. COLI* EXPRESSION SYSTEMS

Despite its limitations the gram negative bacterium *E. coli* still remains one of the most attractive systems for the production of heterologous proteins because of its ability to grow rapidly to a high density on inexpensive substrates, its well characterised genetics and the availability of an increasingly large number of cloning vectors and mutant host strains (Baneyx, 1999). Although *E. coli* has proven useful for the production of recombinant proteins of modest size containing not more than two or three disulfide bridges, it has many limitations where the production of proteins with more complex structures is concerned, particularly those requiring post-translational modifications for biological activity (Hockney, 1994). By virtue of the fact that *E. coli* has no intracellular protein-processing compartment such as endoplasmic reticulum or golgi, it has

no capacity to glycosylate proteins in either N-or O-linked conformations. Glycosylation is the most extensive of all post-translational modifications and has important functions in the secretion, antigenicity and clearance of glycoproteins (Jenkins *et al.*, 1996). N- and O-linked oligosaccharides differ in the sugar residues they contain. For instance, in O-linked sugars, N-acetylgalactosamine is invariably linked to serine or threonine. In all N-linked oligosaccharides, N-acetylglucosamine is linked to asparagine. O-linked oligosaccharides are generally short, most containing only one to four sugar residues. The N-linked oligosaccharides, in contrast, have a minimum of five sugars and always contain mannose in addition to N-acetylglucosamine (Lodish *et al.*, 1995).

Many proteins expressed in *E. coli*, particularly those that are over expressed, can accumulate intracellularly in the form of insoluble, biologically inactive inclusion bodies. Biologically active protein may be recovered from these inclusion bodies by denaturation and refolding, which adds to the expense and complexity of downstream processing (Hockney, 1994). Another obstacle to the use of microbial systems such as *E. coli* for the production of therapeutic proteins is the presence of the lipopolysaccharide cell wall (endotoxin), which is highly immunogenic. Its presence in a therapeutic protein would have a detrimental effect on the recipient. Nevertheless, attempts are continually been made to circumvent many of the problems associated with the use of *E. coli* as an expression system. For example, research has illustrated that many proteins do not attain their natural conformation spontaneously but require the assistance of accessory proteins like molecular chaperones and foldases to achieve their final biologically active state (Hockney, 1994). Coexpression of these chaperones and foldases with the heterologous protein has been shown to reduce inclusion body formation in some cases (Weickert *et al.*, 1996; Baneyx, 1999). Generally inclusion bodies tend to form only at temperatures above 30^0C. It has been shown that growth of cultures at temperatures lower than 30^0C lead to the production of active, soluble protein whereas when growth

proceeded at 37^{0}C most of the produced protein was insoluble and inactive (Schein, 1989). Although the aforementioned solutions exist for some of the problems associated with the use of *E. coli*, the production of fully functional, active protein in this system still remains quite a challenge. Because of the safety issues concerning the use of gram negative bacteria for the production of a recombinant vaccine, it is necessary to use alternative expression systems that have attained GRAS (generally regarded as safe) status which are capable of performing post-translational modifications similar to those present in higher eukaryotes.

1.4.2: FUNGAL SYSTEMS

In recent years techniques for transforming many filamentous fungi have evolved which allow for heterologous proteins to be produced in established fungal systems. Many of the industrially important species currently used such as *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei* and *Penicillium chrysogenum* have GRAS status and are considered safe for use in the production of recombinant proteins for human use (Radzio and Kück, 1997). The type of vectors used to transform filamentous fungi integrate heterologous DNA into the host genome by means of homologous or heterologous recombination and the resulting transformants usually have a high mitotic stability. Secretion of the recombinant protein can occur if the appropriate secretion signal is present. One special feature of protein secretion in filamentous fungi is that proteins are secreted mainly at the tips of growing hyphae (Punt *et al.*, 1994). Post-translational modifications such as disulfide bond formation and glycosylation are all known to occur in filamentous fungi. Glycosylation patterns similar to those which occur in yeast systems like *Saccharomyeces cerevisiae* have been reported but in contrast to *S. cerevisiae* where hypermannosylation occurs, the glycosylation patterns in filamentous fungi are more similar to those found in higher eukaryotes (Jenkins *et al.*, 1996). Nevertheless, certain problems with the use of this system serve as a barrier to

its widespread acceptance. Due to the fact that filamentous fungi secrete an abundance of proteases, yields of heterologous proteins can be quite low. Despite the fact that protease deficient strains have been developed even the best strains have shown residual levels of protease activity which under certain conditions can result in substantial losses in protein yield (Punt *et al.*, 2002).

1.4.3: Insect (Baculovirus) Expression Systems

The baculovirus-infected insect cell expression system has become a popular route for recombinant protein synthesis because of its short process development time and potentially high yields (Jenkins *et al.*, 1996). The majority of evidence to date suggests that the N-glycosylation capabilities of this system are limited to the production of only simple oligomannose type oligosaccharides (Jarvis and Finn, 1995). However, this isn't always the case, a secreted alkaline phosphatase produced in *Trichoplusia ni* has been reported that has both galactose and terminal sialic acids present (Jenkins *et al.*, 1996). Over expression of the appropriate galactosyltransferases and sialyltransferases have led to some success in generating sialylated oligosaccharides on insect derived proteins (Andersen and Krummen, 2002). Proenzyme *S. japonicum* cathepsin C was expressed in a baculovirus expression system utilising *Trichoplusia ni* (High Five) strain host insect cells (Hola-Jamriska *et al.*, 2000). Approximately 1 mg of affinity purified schistosoma cathepsin C was obtained per litre of insect cell culture supernatant. One of the polypeptides produced in this system was seen to have a molecular mass of 55 kDa, which is larger than the predicted molecular mass of 51kDa. This indicated that the recombinant protein was glycosylated.

1.4.4: Transgenic Animals as Expression Vectors

The combination of large daily protein output, excellent post-translational modification capabilities, ease of access to the recombinant protein (milking)

and low capital costs of production plants (farms) as compared to high volume fermenters have made the transgenic mammary gland an excellent candidate for the production of recombinant proteins. Regulatory sequences of the major milk protein genes have been used with variable successes to direct the expression of an array of heterologous proteins in the milk of transgenic mice, rats, rabbits, pigs, sheep, goats and cows (Echelard, 1996). In some instances it has been difficult to obtain protein yields superior to 1 mg/ml with cDNA derived constructs and sometimes transgenes have proven to be transcriptionally silent. Transgenically produced human protein C (hPC) from mouse and swine is well characterised. Analysis of transgenic mouse and pig mammary gland derived hPC as compared to that derived from plasma showed reduced anticoagulant activity in addition to high proportions of single chain unprocessed precursors indicating that post-translational modifications were carried out inefficiently (Lee *et al.*, 1995; Medved *et al.*, 1995). However, at least two products, namely sheep derived human α-antitrypsin and goat derived antithrombin III have entered phase I clinical trials and the number of heterologous proteins being produced in transgenic animals continues to grow rapidly.

1.4.5: Yeast Expression Systems

Yeasts are attractive hosts for the production of heterologous proteins. Having been the basis of biotechnology for over the last 4,500 years the safety of this organism is assured (Spier, 2000). Yeast expression systems offer considerable advantages over alternative eukaryotic and prokaryotic systems. Post-translational modification can be efficiently carried out resulting in the production of authentic biologically active molecules (Buckholz and Gleeson, 1991; Cereghino and Cregg, 1999; Gellissen, 2000). Yeast offers the ease of microbial growth and genetic manipulation associated with bacterial systems along with the advantage of rapid growth to high densities on inexpensive media. In 1984, the first recombinant vaccine to be granted FDA approval, the

human hepatitis B vaccine, was produced in *Saccharomyeces cerevisiae*. A vector was constructed carrying the DNA sequence for hepatitis B surface antigen (HBsAg) and this was expressed within the yeast (McAleer *et al.*, 1984). Prior to the availability of this recombinant vaccine, HBsAg had to be purified from the blood of carrier individuals for use in the vaccine, which required the implementation of stringent processes to purify the antigen and eliminate any pathogens present. Disadvantages still exist with the use of yeasts for the expression of some heterologous proteins, which are mostly related to their inability to perform certain complex post-translational modifications such as prolyl hydroxylation (refer to 1.4.5.2) and amidation as well as some some types of phosphorylation and glycosylation (Cereghino and Cregg, 1999).

1.4.5.1: *SACCHAROMYECES CEREVISIAE*

Although a wealth of information exists detailing the genetics and cell biology of this yeast species, it is sometimes not viewed favourably as a host for recombinant protein expression because of the perception that it has a lower secretory capacity relative to other yeasts such as *Pichia pastoris* (Cereghino and Cregg, 1999). In some cases yields of heterologous proteins have been shown to only reach a maximum of 1-5% of the total protein even under the influence of a strong promoter. Plasmid stability can be poor, which results in difficulties in scale-up. Hyperglycosylation of recombinant proteins can occur. This can result in changes in immunogenicity, diminished activity and decreased serum retention of the heterologous protein product (Moir and Dumais, 1987). Many secreted proteins produced in *S. cerevisiae* are found in the periplasmic space in a cell-associated form. This leads to problems during purification and decreased yields of the purified product (Smith *et al.*, 1985). Nevertheless, *S. cerevisiae* has GRAS status, it possesses well-characterised auxotrophic markers and strong promoters and a high-copy number mitotically stable plasmid has been identified (Buckholz and Gleeson, 1991). While much work has been carried out in an attempt to modify and improve *S. cerevisiae* as

an expression system for the production of recombinant proteins, alternative yeast hosts have been developed and are being employed by many laboratories for stable high level expression of appropriately modified recombinant proteins.

1.4.5.2: *PICHIA PASTORIS*

Among yeasts a limited number of species belonging to the genera *Hansenula, Candida, Torulopsis* and *Pichia* are capable of growth on methanol as a sole carbon source. Initially these yeasts were considered ideal candidates for the production of single-cell protein (SCP) but nowadays the scope of these yeasts has been extended, culminating in their use as powerful expression systems for a growing number of biotechnological applications (Gellissen, 2000). *Pichia pastoris* is a single celled organism that can easily be grown and manipulated. However, it is a eukaryote capable of undertaking many of the post-translational modifications performed by higher eukaryotes such as proteolytic processing, folding, disulfide bond formation etc., (Higgins and Cregg, 1998; Cereghino and Cregg, 2000). Since 1984 *P. pastoris* has been used to produce over 300 foreign proteins. Its wide acceptance is attributed to many factors. The techniques used for the molecular genetic manipulation of *P. pastoris* are similar to those of *S. cerevisiae*, one of the most well characterised experimental systems in modern biology (Higgins and Cregg, 1998). Extremely high yields of both intracellular and secreted proteins have been reported using *P. pastoris* (Romanos, 1995). By virtue of the fact that *P. pastoris* secretes only very low levels of endogenous proteins, the secreted heterologous protein constitutes the vast majority of total protein in the medium which will greatly simplify any subsequent purification steps (Cereghino and Cregg, 2000). Protein expression in *P. pastoris* is under the control of the tightly regulated alcohol oxidase promoters AOX1 and AOX2. The AOX1 gene is responsible for the vast majority of alcohol oxidase activity in the cell. The first step in the metabolism of methanol is the oxidation of methanol to formaldehyde, generating hydrogen peroxide in the process, by the enzyme alcohol oxidase

(Higgins and Cregg, 1998). This has to occur within the peroxisome to avoid hydrogen peroxide toxicity. The peroxisome is a specialised organelle, which is massively induced in cells growing on methanol. Peroxisomes also function as storage compartments for the foreign protein, which would be of particular benefit if the protein of interest is highly susceptible to proteolytic degradation or if it is toxic to the host (Faber *et al.*, 1995). The AOX1 promoter is tightly repressed by glucose and most other carbon sources but is induced >1000-fold in cells shifted to methanol as a sole carbon source (Cereghino and Cregg, 1999 and 2000). *P. pastoris* vectors are designed for homologous integration into either the AOX1 locus or the HIS4 locus (Hollenberg and Gellissen, 1997). Cleavage of a *P. pastoris* vector within a sequence shared by the host genome stimulates homologous recombination events that efficiently target integration of the vector to that genome locus (Cereghino and Cregg, 2000). Homologous recombination will maximise the stability of expression strains. The vector is restricted at a unique site in the marker gene (e.g., HIS4) or the AOX1 promoter fragment and is subsequently transformed into the appropriate auxotrophic mutant, for example, in this project *P. pastoris* strain GS115 was used. The vector of choice for this project was pPIC9K supplied by Invitrogen, which contains the bacterial kanamycin gene that allows for the selection of transformants on plates containing the antibiotic G418. The level of G418 resistance can be roughly correlated to the vector copy number. If the recombination event results in the replacement of the AOX1 gene, such strains will have to rely on the transcriptionally weaker AOX2 gene for growth on methanol. These strains are referred to as Mut^S (Methanol utilisation slow), which utilise less methanol and sometimes express higher levels of foreign protein than wild type (Mut^+) strains, particularly in shake flasks. Typically, transformed strains are initially grown in a defined medium containing glycerol as its carbon source. During this time biomass will accumulate but heterologous gene expression will be fully repressed. Upon depletion of glycerol a transition phase will be initiated during which additional glycerol will be fed to the culture at a growth-limiting rate. Finally, methanol or a

mixture of glycerol and methanol will be fed to the culture to induce expression (Cereghino and Cregg, 2000).

(www.invitrogen.com)

Figure 5: A diagrammatic representation of the pPIC9K expression vector

pPIC9K contains the *S. cerevisiae* α-mating factor, which will allow the foreign protein to be secreted in *P. pastoris*.

Many factors will affect recombinant protein production in *P. pastoris* including the following:

- The copy number of the expression cassette. A strain that contains multiple integrated copies of an expression cassette can sometimes yield more heterologous protein than single copy strains (Clare *et al.*, 1991a). In a detailed optimisation study of tetanus toxin fragment C expression it was shown that protein levels increased with increasing copy number (Clare *et al.*, 1991 b, Romanos, 1995). Another such study compared the expression levels of a single-copy vector encoding HBsAg to those

obtained using increasing copies of the heterologous gene expression cassette created using an *in vitro* multimerisation approach. A systematic investigation of the resultant clones demonstrated that the increase in copy number results in a proportional elevation in the steady-state levels of the HBsAg-specific mRNA, which in turn is closely paralleled by a corresponding increase in the total levels of HBsAg protein (Vassileva *et al.*, 2001). For secreted proteins, the effects of gene dosage are not so simple. Although there are many reports showing that product yield has been improved using multiple vector copies, in some cases it has been discovered that a too-high copy number reduces yield i.e., an optimal rather than maximal copy number is required (Romanos *et al.*, 1998). Optimisation studies should always be carried out using single and multi-copy transformants to determine the exact copy number required for optimum protein production before any expression project is undertaken.

- The site and mode of chromosomal integration of the expression cassette. In some instances it has been shown that strains with AOX mutations (i.e., those that occur as a result of gene transplacement) are better producers of foreign proteins than wild-type strains (Cregg *et al.*, 1987; Tschopp *et al.*, 1987).

- The fermentation parameters, including media and growth conditions (see "Recommendations", Section 4.7). Expression levels in shake flasks are generally low relative to what is obtainable in fermenter cultures. Fermenters provide a more controlled environment relative to shake flasks where factors such as oxygen concentration, pH and temperature can be closely monitored. Methanol concentration in the media also has an effect on expression levels and it has been shown that the level of transcription initiated from the AOX1 promoter can be 3-5 times greater in *P.pastoris* cells fed methanol at growth limiting rates

relative to cells grown in excess methanol (Higgins and Cregg, 1998). Optimisation of the pH medium is crucial especially for yeast-secreting protein because they can grow over a wide pH range. *P. pastoris*, for instance is capable of growth between pH 3.0-7.0 (Thiry and Cingolani, 2002). *P. pastoris* can produce proteases that could have a detrimental effect on the recombinant protein. Adjusting the pH to one that is not optimal for a problem protease will remedy that problem. Temperature is another factor, which may be controlled in a fermenter that has a marked effect on both the solubility and productivity of the expressed protein. Lowering the temperature from 30 to 25^{0}C during the methanol induction phase results in a four-fold increase in yield production of galactose oxydase cloned in *P. pastoris* (Whittaker, 2000).

Active *S. mansoni* cathepsin B has recently been produced in a *P. pastoris* expression system using the pPICZα vector (Caffrey *et al.*, 2002). Induction was carried out for 48 hours in 1% methanol and yields of the purified enzyme were in the range of 40-60 mg/L of culture medium. An earlier study by Caffrey *et al.*, from 2001 used pPICZα to produce rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*, which causes African sleeping sickness, to a final yield of 20-40 mg/L.

In a high-density fermentation an expression level of 1.5g/L of human insulin precursor was achieved using the pPIC9K vector in *P. pastoris*. High-copy number transformants were generated and *BglII* was used to linearise the plasmid because it is considered more efficient in the generation of high-copy number transformants than *SacI* or *SalI* (Wang *et al.*, 2001).

Human cystatin C, a cysteine proteinase inhibitor was produced in a fed-batch fermentation system. Mut^s transformants were used in this fermentation. During the induction phase it was found that addition of glycerol together with methanol increased the growth rate of the cells, which increased the production

rate of the recombinant protein. Typically, using a methanol-only feed results in long incubation times and low productivity. However, it was observed that the maximum concentration of expressed protein was much lower with the higher glycerol feed rate. Lower glycerol levels lead to higher expression levels of cystatin (Files *et al.*, 2001).

An example of the limitations of yeast systems was seen in a comparative study, using the methylotroph *Hansenula polymorpha* and *P. pastoris* to produce human α1(I) procollagen. *H. polymorpha* was capable of secreting recombinant human α1(I) procollagen at a level of 0.6g/L. However, pPICZB was used as the vector in *P. pastoris*, which does not contain the *S. cerevisiae* α-mating factor secretion signal and it was shown that the human procollagen preprosegment could not be used to obtain secretion of recombinant collagen in *P. pastoris*. Amino acid analysis illustrated that proline residues of the collagen product were not hydroxylated. This is essential for production of triple helical collagen. Yeasts appeared to lack the necessary prolyl 4-hydroxylase required to produce recombinant triple helical collagen (de Bruin *et al.*, 2000).

Generally, when a recombinant protein is being expressed in *P. pastoris* expression is under the control of the inducible AOX promoter. However, constitutive promoters, namely the glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter have been developed for the *P. pastoris* expression system. These vectors allow for the continuous production of the desired recombinant protein without methanol induction. This system is more desirable for large-scale production because the hazard and cost associated with the storage and delivery of large volumes of methanol are eliminated. Using the GAP promoter continuous production at levels of 300 mg/L of human chitinase was observed in a fed-batch fermentation using either glucose or glycerol as the carbon source (Goodrick *et al.*, 2001).

In conclusion, *P. pastoris* has been widely employed over the past decade to produce a myriad of recombinant proteins. *P. pastoris* has intrinsic features that make it pre-eminently suitable for specific commercial and scientific applications (Faber *et al.*, 1995). However, there are certain proteins, which simply cannot be expressed by this system (Cereghino and Cregg, 2000). Nevertheless, many of the problems encountered in protein expression can be overcome by due consideration of the factors that influence protein expression. Each and every expression project is different and many parameters of the process may need to be manipulated before optimum expression of a heterologous protein can occur. The widespread use of this system will undoubtedly expand our knowledge of the intricacies and limitations of the system.

1.5: PROJECT AIMS

The aim of this project was to isolate the cDNA encoding *Schistosoma mansoni* cathepsin C (SmCC) using primers that were designed based on the sequence that was available in the public database (Genbank accession no. Z32531). Secondly, this cDNA was cloned into a vector that could be used to transform the yeast *Pichia pastoris* with a view to obtaining functional recombinant enzyme.

2: Materials and Methods

2.1: MATERIALS

The following materials were obtained from the suppliers listed below:

Amresco

Ethidium bromide was obtained from Amresco Inc., 30175 Solon Industrial Parkway, Solon, Ohio 44139, USA.

Bachem

Synthetic peptides H-Gly-Arg-NHMec and Z-Phe-Arg-NHMec and inhibitors Z-Phe-Ala-NH_2 and Z-Phe-Phe-OH were supplied by Bachem (UK) Ltd., 69 High Street, Saffron Walden, Essex CB10 IAA UK.

BDH

Silicone anti-foam Silcorel AFP20 and sucrose were supplied by BDH Laboratory Supplies, Poole, UK.

Becton Dickinson

Difco Bacto Agar, Bacto Peptone, Bacto Yeast Extract, Yeast Nitrogen Base without amino acids and 3 ml Luer-Lok syringes were all obtained from Becton Dickinson Microbiology Systems, Sparks, MD, USA.

Bio-Rad

30% Acrylamide/Bis solution, Goat Anti-Mouse IgG (H + L)-HRP conjugate and Trans-Blot SD Semi-Dry Transfer Cell and Power Pac 200 were provided by Bio-Rad Laboratories, 2000 Alfred Nobel Drive, Hercules, CA 94547, USA.

Biotrin

TMB was supplied by Biotrin International, 93 The Rise, Mount Merrion, Co. Dublin, Ireland.

Chivers

Dried skimmed milk (Marvel) was supplied by Chivers, Ireland Ltd., Coolock, Dublin.

Cornell University

A *Schistosoma mansoni* cDNA library cloned in pB42AD (BD Biosciences) was obtained from Dr. Sharon McGonigle, Cornell University, Ithaca, New York.

Fisher Scientific

100% Methanol was supplied by Fisher Scientific (UK) Ltd., Bishop Meadow Road, Loughborough, Leics, LE11 5RG, UK.

Genosys Biotechnologies (Europe)

All oligonucleotide primers were obtained from Genosys Biotechnologies (Europe), London Road, Pampisford, Cambridge, CB2 4EF, UK.

Invitrogen

E. coli Top F' cells, Multi-Copy *Pichia* Expression Kit (Version E) and Pro Bond Resin 50% slurry in 20% ethanol were all supplied by Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, CA 92008, USA.

New England Biolabs

The restriction enzymes *AvrII*, *PstI* and *SnaBI* and prestained protein marker, broad range were obtained from New England Biolabs, 32 Tozer Road, Beverly, MA 01915-5599, USA.

Oswel

DNA sequencing was performed by Oswel Research Products Ltd, Lab 5005, Medical and Biological Services Building, University of Southampton, Boldrewood, Bassett Crescent East, Southampton, SO167 PX, UK.

Pall Gelman

Acrodisc 32mm syringe filters with 0.2 μm Supor Membrane were supplied by the Pall Corporation, 600 South Wagner Road, Ann Arbour, MI 48103-9019 USA.

Pierce

The BCA protein assay kit was acquired from Pierce Biotechnology Inc., Rockford, Il 61105, USA.

Promega

Promega, 2800 Woods Hollow Road, Madison, WI 57311-5399, USA supplied the following materials: Agarose, LE analytical grade and low melting point, blue/orange loading dye, 1 kb DNA ladder, 87% glycerol, IPTG (isopropyl-beta-D-thiogalactopyranoside), PCR Master mix, pGEM -T Easy Vector System, pGEM -T Vector System, PMSF (phenylmethylsulfonylfluorid) and restriction enzymes *AvaII* and *VspI*.

Roche

Anti-His_6 antibody was provided by Roche Diagnostics Ltd., Bell Lane, Lewes, East Sussex, BN7 1LG.

Schleicher & Schuell

Protran nitrocellulose transfer membrane was supplied by Schleicher & Schuell Inc., 10 Optical Avenue, Keene, MH 03431, USA.

Sigma-Aldrich

Sigma-Aldrich PO Box 14508, St. Louis, Missouri, 63178, USA, also the supplier of Fluka and Riedel-de-Häen materials supplied the materials detailed in Table 1:

Table 1: Materials supplied by Sigma-Aldrich

Sigma	Fluka	Riedel-de-Häen
-Acetic acid	-Agar	-Calcium Chloride
-Bovine Serum Albumen	-Ammonium persulfate	-Glycine
-Brilliant blue R250	-Buffer standard solution,	-Natrium-acetat-3-hydrat
-Chloroform	colour coded pH 4.0, 7.0 &	-2-propanol
-DMF	10.0	-Sodium Hydroxide Pellets
-EDTA	-Phenol	-Tween 20
-Sigma Fast ™ 3,3'-	-Sodium Carbonate	
Diamino benzidine tablet	-Sodium Chloride	
set (DAB peroxidase	-Tryptone media	
substrate)	-Yeast extract	
-Ethyl alcohol		
-Glass beads		
-X-Gal		
-5U/μL Lyticase		
-Lauryl Sulfate -(Sodium		
Dodecyl Sulfate solution)		
-Mineral oil		
-Nuclease free H_20		
-Red Taq ™ DNA		
Polymerase & 10 X		
reaction buffer, 25 mM		
$MgCl_2$ & 12.5 M dNTPs		
-Sodium Phosphate		
-TEMED		
-Tetracycline		
-Trizma base –(Tris		
[hydroxymethyl] amino		
methane)		
disodium salt		

Tulane University

Anti-cathepsin C antibody, purified baculovirus-expressed *S. mansoni* cathepsin C and *Schistosoma japonicum* worms were provided by Professor Paul Brindley, Department of Tropical Medicine, SL29A, Tulane University, Health Sciences Centre, 1430 Tulane Avenue, New Orleans, Louisiana 70112-2699, USA.

Qiagen

Nickel-nitriloacetic acid agarose (Ni-NTA) in 50% ethanol, polypropylene 1 ml columns and the QIAexpressionist ™ kit were supplied by Qiagen Inc., 28159 Avenue Stanford, Valencia, CA 91355, USA.

Equipment

- Anthos 2001 plate reader
- Beckman centrifuge J2-MC
- Consort power supply pac
- Eppendorf centrifuge 5417C
- Hybraid electrophoresis apparatus
- Hybraid omni gene thermal cycler
- Image master ®-VDS transilluminator
- Janke and Kunkel Homogeniser

- Jeio-Tech Shaker SK-71
- Kodak polaroid film
- Neubauer hemocytometer
- Olympus CK2 microscope
- Perkin Elmer fluorimeter
- Sonics and Materials, Inc., Vibra Cell ™ Sonicator
- Thermohybraid PCR express thermal cycler
- Thermo orion pH meter
- Tomy autoclave
- Unicam Helios Spectrophotometer

2.2: METHODS

2.2.1: CLONING OF *SCHISTOSOMA MANSONI* CATHEPSIN C (SMCC) CDNA

Primer Design

The full-length cDNA sequence of *Schistosoma mansoni* cathepsin C (SmCC) was retrieved from the Genbank (accession no: Z32531). This sequence was used to design primers to amplify the pro and mature segments of SmCC. *SnaBI* and *AvrII* restriction enzyme recognition sequences were incorporated into the 5' and 3' ends respectively of the primers that were initially designed (SmCCF and SmCCR). A hexahistidine tail was included in the design of SmCCR for purification of recombinant SmCC (rSmCC) on a Ni-NTA column (see below)

SmCCF: 5'- GCGGCTACGTAGCTGATACTCCTGCTAACTG – 3'

SmCCR: 5'-GCGCCTAGGTTAGTGG*TGGTGGTGGTGG*
*TGGTGG*GCCCTAACACCGGATCAAAACG– 3'

SnaBI and *AvrII* recognition sequences are underlined and the hexahistidine tail is in italics.

Several sets of primers were designed before the 1.3 kilobase (kb) SmCC cDNA was successfully amplified (refer to 3.1.1). Tables 2 and 3 depict the primers that were used to amplify the SmCC cDNA, their sequence and the regions of the SmCC cDNA they amplified.

Table 2: Forward Primers

Primer name	Sequence	Location on SmCC
SmCCF	GCGGCTACGTAGC TGATACTCCTGCT AACTG	base 82-101
SmCCF2	GCTGATACTCCTG CTAACTGTACTTA TG	base 82-101
SmCCF4	GCATTGGGTGTTC CACTG	base 27-44
SmCCMF	CATTAACTGGAAA TCTTCCTTTGG	base 662-685
SmCCMF2	GTTTCGGTAGAAC TCTCTAC	base 461-480
SmCCMF3	GGTTATTATGGCG CTACTAATG	base 1042-1063

Table 3: Reverse Primers

Primer name	Sequence	Location on SmCC
SmCCR	GCGCCTAGGTTAGTGGTGGTGGT GGTGGTGGGGCCCTAACACCGG ATCAAAACG	base 1369-1389
SmCCR2	TTATAACACCGGATCAAAACGTA CACC	base 1369-1389
SmCCR3	GATTTTATAACACCGGATCA	base 1374-1393
SmCCR4	CCAATATCCGGCTCTATTC	base 1400-1420
SmCCR5	CGGATCAAAACGTACACCTAAA CTTTCTAC	base 1351-1380
SmCCR6	TAACACCGGATCAAAACGTACA CCTAAACTTTCTAC	base 1351-1386
SmCCMR	CCAAAGGAAGATTTCCAGTTAAT G	base 662-685
SmCCMR3	CCATAACCAACCAATAAAAC	base 1222-1241
SmCCMR4	CATCAATAGCTATATCCGGG	base 210-230
SmCCMR5	TCTTCATATACTTCAAAACCC	base 1110-1130

A 500 pmol stock of each primer was prepared and from this a working stock of 10pmol/μL was made.

For example, in the case of SmCCF: Mw = 9487 g

nmoles = 33.5 = 33,500 pmoles

stock = 500 pmol = $\frac{33,500}{500}$ = 67 μL RNAse free water

Therefore to make a 500pmol stock 67 µL of RNAse free water was added to the primer. The 10 pmol/µL working stock was made by performing a 1:50 dilution of this.

Template DNA

When SmCC was amplified by PCR the template DNA used was *S. mansoni* cDNA library in pB42AD, unless stated otherwise.

Polymerase Chain Reaction (PCR)

A ready-made PCR master mix (Promega) was employed for the majority of PCR reactions carried out.

Table 4: PCR Reaction (Promega Master Mix)

Component	**Volume (25 µL rxn.)**	**Final concentration**
PCR master mix, 2X	12.5 µL	1X
5' primer	1 µL	10 pmole/µL
3' primer	1 µL	10 pmole/µL
DNA template	2.5 µL	< 250 ng
Nuclease free H_20	8 µL	N/A

2X master mix was composed of 50 units/ml of Taq DNA polymerase in a proprietary reaction buffer (pH 8.5), 3 mM $MgCl_2$, 400 µM dATP, 400 µM dCTP, 400 µM dGTP and 400 µM dTTP.

Where ready-made master mix was not used it was necessary to make up a mix by adding all the individual components of the PCR reaction.

Table 5: PCR Reaction

Component	Volume (25 µL rxn)	Final concentration
10X buffer	2.5 µL	1X
25 mM $MgCl_2$	2.5 µL	2.5 mM
12.5 mM dNTPs	0.8 µL	0.2 mM
Taq (1U/µL)	2.5 µL	2.5 U
RNAse free H_20	12.2 µL	N/A

1 µL of SmCCF and SmCCR primers and 2.5 µL of template DNA were added to each tube.

By virtue of the fact that some primers failed to successfully amplify the SmCC cDNA different PCR cycling conditions were implemented. Other conditions in the PCR reaction such as $MgCl_2$ concentration were altered in an attempt to amplify the cDNA.

A standard PCR cycle used is detailed below:

Denaturing	94^0C/5 min (1st cycle)	
Denaturing	94^0C/1 min 30 sec	40 cycles
Annealing	55^0C/1 min 30 sec	
Extension	72^0C/1 min 30 sec	
Extension	72^0C/10 min (last cycle)	

Analysis of PCR Product

Amplified fragments were electrophoretically separated in a 1% agarose-Tris-acetate gel. The gel was stained with ethidium bromide and visualised under ultra-violet light.

DNA Purification

The 1.3 kb band was excised using a sterile blade. For DNA purification, a 1% agarose low melting point gel was used. DNA was purified from the gel using the Wizard DNA Purification System (Promega) according to the protocol provided by the manufacturer. The purified band was concentrated by ethanol precipitation. 0.1 volumes of 3 M sodium acetate pH 5.0-6.0 were added to the DNA sample, followed by mixing. Then 2-3 volumes of 100% ethanol were added. Brief vortexing ensued and the sample was placed at -80^0C for 15 min Afterwards centrifugation at 1,807 x g for 10 min was carried out. Excess ethanol was allowed to evaporate and the pellet was resuspended in 20 µL RNAse free H_20.

Determination of DNA concentration

The concentration of SmCC cDNA was determined on a spectrophotometer by measuring absorbance at 260 nm.

The concentration of DNA was determined by implementation of the following formula:

$$\text{ng/µL DNA} = \frac{50\ \text{ng/µL}}{1} \times (OD_{260}\ \text{measured}) \times (\text{dilution factor})$$

2.2.2: CLONING OF SMCC CDNA INTO THE PGEM VECTOR

The SmCC cDNA was successfully amplified with SmCCF2 and SmCCR6 primers. It was then decided to clone the purified SmCC fragment into the pGEM vector in order to increase the concentration of SmCC template present for the re-amplification of SmCC with SmCCF and SmCCR primers containing the restriction sites for cloning into pPIC9K. By virtue of the fact that Taq polymerases add a single deoxyadenosine to the 3' ends of the amplified fragments in PCR and because the pGEM vector contains single 3'-T overhangs at the insertion site ligation of the gene to the vector can, in theory, be swiftly performed.

Ligation into pGEM

For ligation of DNA to pGEM the molar ratio of vector: insert DNA was 3:1. The following equation was used to determine the quantities of vector and insert to be used:

$$\text{Molar ratio: } \frac{\text{ng vector x kb size of insert}}{\text{kb size of vector}} \times \frac{3}{1}$$

The vector was supplied at a concentration of 50 ng/μL

A sample ligation set-up is outlined in Table 6 in accordance with the guidelines provided by the manufacturer (Promega).

Table 6: Sample Ligation Procedure

Component	Sample	Background control	Positive control
10X buffer	1 µL	1 µL	1 µL
vector	3 µL	1 µL	1 µL
insert	1 µL	N/A	N/A
control insert	N/A	N/A	2 µL
ligase	1 µL	1 µL	1 µL
H_20	4 µL	7 µL	5 µL
Final volume	10 µL	10 µL	10 µL

All ligation components were provided with the kit. Ligation was performed at 4^0C overnight.

Preparation of Competent Cells

In order for the subsequent transformation steps to be carried out a batch of chemically treated *E. coli* Top F' cells were prepared.

It was necessary to prepare the following:

- Luria-Bertani broth (LB) (bacto-tryptone 10g/L, bacto-yeast extract 5g/L, NaCl 10g/L, pH 7.0, autoclaved for 20 min)
- 0.1 M calcium chloride
- 80% filter-sterilised glycerol
- 10 mg/ml tetracycline stock prepared in ethanol

An overnight culture of *E. coli* Top F' cells were grown at 37^0C in 2 ml of LB broth (10μg/ml tetracycline). Four hundred ml of LB broth (10 μg/ml tetracycline) was inoculated with this overnight culture and cells were grown to an OD_{600} of 0.4-0.5, which took 2-2.5 hrs. Cells were then chilled on ice for 1 hr, followed by centrifugation at 200 x g for 10 min. Resuspension of the cells followed in 10 ml per 50 ml starting culture of 0.1 M $CaCl_2$. Cells were once again chilled on ice, this time for 40 min. The second centrifugation step was implemented as for the first except the resultant pellets were resuspended in 2 ml per 50 ml starting culture 0.1 M $CaCl_2$. 80% glycerol was added to the cells to a final concentration of 10%. Cells were stored at -80^0C until use.

Transformation

Once the overnight ligation step was performed, transformation of competent *E. coli* cells with the vector containing the ligated SmCC cDNA could proceed (Figure 6).

The following media was needed for the transformation process (Table 7).

Table 7: Media for Transformation Process

Luria-Bertani broth/XIA plates	• *15g/L agar • 5g/L NaCl • 10g/L tryptone • 5g/L yeast pH 7.0, autoclave 20 min * agar for XIA plates only

The relevant antibiotics and supplements listed in Table 8 were added when the autoclaved media had cooled to room temperature. Only ampicillin was added to the LB media.

The transformation procedure is outlined in Figure 6

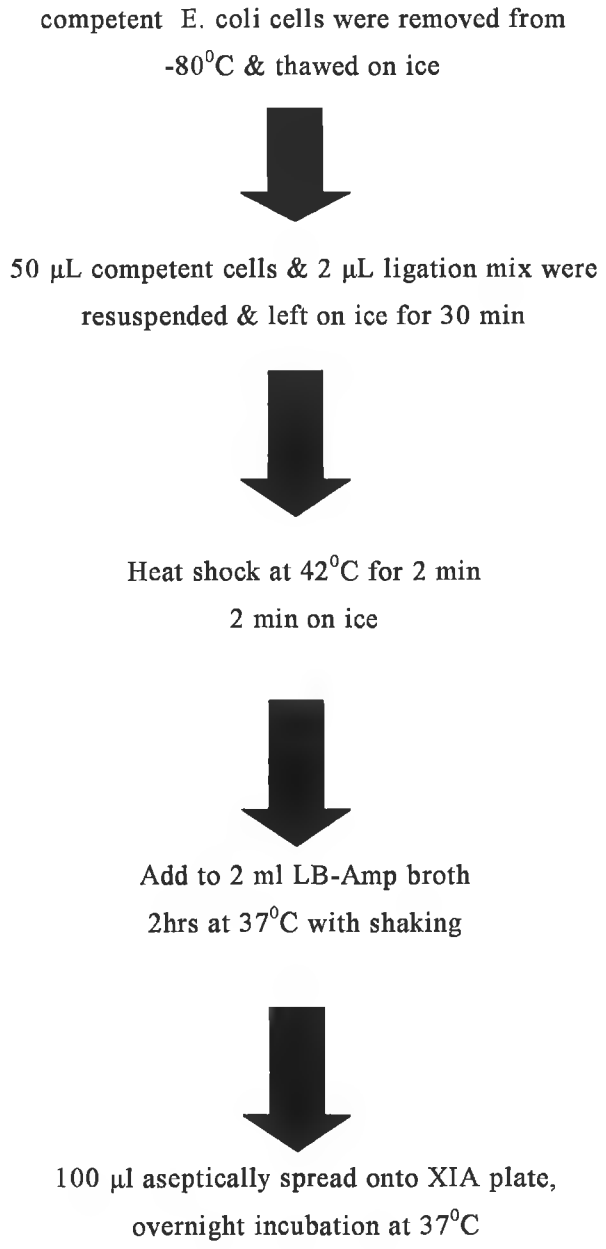


Figure 6: Overview of the transformation procedure

Table 8: Supplements to be added to media in Table 7

Ampicillin	Stock 100 mg/ml Want 60 μg/ml → 600 μL/L added
IPTG	Stock 0.1 M (0.6 g IPTG + 25 ml dH_20, filter sterilised) Want 0.2 μM → 2 ml IPTG/L added
X-Gal	Stock 50 mg/ml DMF Want 40 μg/ml → 800 μL X-Gal/L added

Only ampicillin was added to LB media

Screening of Colonies

After 24 hrs transformed colonies appeared on the plates. An advantage of using the pGEM system is that blue/white colour screening can be employed. If transformed colonies do not contain the fragment of interest blue colonies will appear on the plates because a functional β-galactosidase will be produced by the plasmid resulting in metabolism of the X-Gal substrate present in the media giving the colonies a blue colour. If the insert is successfully incorporated into the pGEM vector, the gene encoding β-galactosidase will be interrupted and a functional enzyme will not be produced. The X-Gal substrate will not be broken down and colonies on the plate will be white. White colonies were selected and grown up overnight in LB-ampicillin (LB-Amp) broth.

Plasmid Purification

Plasmids were purified from an overnight bacterial culture by employing either the Wizard Plus SV Minipreps DNA Purification System using the protocol outlined by the manufacturer (Promega) or the alkaline lysis method.

Alkaline Lysis Method

The following reagents detailed in Table 9 were prepared before the procedure was carried out:

Table 9: Alkaline Lysis Reagents

Lysis Buffer	9 ml H_20 0.25 ml Tris-HCl 0.20 ml 0.5 M EDTA 0.50 ml 20% glucose
SDS/NaOH (10X)	8.8 ml H_20 1 ml 10% SDS 200 µL NaOH (10 N)
Ammonium acetate 7.5 M	Filter sterilised
Phenol: Chloroform (1:1)	4 ml phenol 4 ml chloroform

- 1.5 ml of the overnight bacterial culture was centrifuged at 1,807 x g for 1 min. The supernatant was discarded and the pellet was resuspended in 200 µL of lysis buffer. This was left at RT for 5 min.

- 400 µL of SDS/NaOH was added, tubes were inverted five times and were left on ice for 5 min.

- 300 µL of 7.5 M ammonium acetate was added, tubes were inverted five times and left on ice for 10 min followed by centrifugation for 5 min at 1,807 x g

- The supernatant was kept and the same volume of 1:1 phenol:chloroform was addded. Vortexing for two min and centrifugation for 2 min at 1,807 x g ensued. The clear top phase was retained and transferred into a clean tube.

- 0.6 volumes of 100% isopropanol were added and tubes were left at RT for 10 min. Centrifugation at 1,807 x g for 15 min was then carried out.

- The supernatant was removed and the pellet was washed with 70% ethanol. Some of the pellets floated when ethanol was added so it was necessary to centrifuge these pellets for an additional 3 min.

- The ethanol was discarded but the pellet was not disturbed. It was left to air dry for 10 min.

- The pellet was resuspended in 40 µL of sterile H_20 and 1 µL of sterile DNAse free RNAse was added (1mg/ml).

The success of the plasmid purification process could be ascertained by running the sample on a gel in order to see the purified band under UV or also by determining the concentration of DNA in the purified sample spectrophotometrically (refer to 2.2.1)

Determination of the Presence of SmCC in pGEM

When the primers in Table 10 are used to amplify pGEM they produce a PCR product of 172 bp. If an insert is present the expected size of this PCR product will be the size of the insert + 172 bp.

Table 10: pGEM Specific Primers

Primer	Sequence	Location on pGEM
pGEM F	CGCATGCTCCC GGCCGCCAT	base 21-40
PUC/M13 R	CAGGAAACAGC TATGAC	base 176-192

pGEM vector containing the SmCC fragment was used as a template for a PCR reaction in which the primers SmCCF and SmCCR were used to amplify SmCC with restriction sites for cloning into pPIC9K vector.

2.2.3: CLONING OF SMCC CDNA INTO THE PPIC9K VECTOR

Restriction Digest

A double digest was performed on the purified SmCC cDNA and pPIC9K vector. The pPIC9K vector is supplied by Invitrogen at a concentration of 1 μg/μL. Table 11 on the next page outlines the set-up of a typical restriction enzyme digest.

Table 11: Digest Set-Up

Component	SmCC cDNA	pPIC9K
10X Buffer 4	5 µL	5 µL
BSA	5 µL	5 µL
DNA	20 µL	12.5 µL
H_20	15 µL	15 µL
AvrII	2.5 µL	2.5 µL
SnaBI	2.5 µL	2.5 µL
Final volume	50 µL	50 µL

The digest was performed for 3 hrs at 37^0C. Digested DNA was cleaned-up using the Wizard DNA Clean-Up System using the protocol supplied by the manufacturer (Promega) except when the final elution step was carried out 30 µL of H_20 was left on the column for 5 min before eluting. In order to ensure that the digest was complete and that the vector DNA had been successfully linearised the digest was run on a 1% agarose gel. Ligation and transformation protocols were carried out exactly as for pGEM (refer to section 2.2.2), the only difference being that LB-Amp plates were used when plating the transformation reaction instead of XIA plates. Plasmid DNA was isolated from an overnight bacterial culture of the transformed colonies. Due to the fact that blue/white colour screening could not be implemented when using pPIC9K, the presence of the SmCC cDNA in the vector could be determined only by PCR using SmCC specific primers, by visualisation of the plasmid DNA on a gel to see if there was a size difference between this DNA and the DNA of pPIC9K alone and by carrying out another double digest with *SnaBI* and *AvrII* to liberate the SmCC cDNA from the vector.

Preparation of Glycerol Stocks

Glycerol stocks of pPIC9K/SmCC expression cassette were made by growing up an overnight culture of a colony containing pPIC9K that was shown by PCR to contain SmCC in LB-Amp. 0.15 ml of filter sterilised 100% glycerol was added to 0.85 ml of this overnight culture, which was stored at -80^0C after adequate mixing.

2.2.4: DNA SEQUENCING

The complete DNA sequence of the selected recombinant clones was determined on a commercial basis by an automated fluorescence based method using an ABI 377 Applied Biosciences 96 lane fluorescent DNA sequencing machine by Oswel Research Products Ltd., DNA Sequencing and Molecular Biology Division, Southampton, UK.

2.2.5: TRANSFORMATION OF PICHIA PASTORIS WITH SMCC

Linearisation of the Plasmid

pPIC9K/SmCC was linearised using the restriction enzyme *BglII* prior to the transformation process to stimulate recombination when the plasmid is transformed into *Pichia pastoris*. Other enzymes conventionally used for this step such as *SalI* and *SacI* could not be used as they cut within the SmCC cDNA and *BglII* was the only enzyme that didn't.

***Pichia pastoris* host strain**

Pichia pastoris host strain GS115 was used for all transformation procedures. This strain has a mutation in histidinol dehydrogenase (his4) that prevents it from synthesising histidine. The pPIC9K expression plasmid carries the HIS4

gene, which complements his4 in the host so transformants are selected for their ability to grow on histidine deficient media. GS115 was streaked onto a Yeast Extract Peptone Dextrose (YPD) plate (Table 14) and was incubated at 28-30^{0}C for 48 hrs. A single colony from this plate was used to inoculate 10 ml of YPD in a 100 ml conical flask. Overnight incubation at 30^{0}C/225 r.p.m ensued. 200 ml of YPD in a 500 ml baffled flask were inoculated with 5, 10 and 20 µL of this overnight culture. Overnight incubation at 30^{0}C/225 r.p.m was once again carried out. Cells were harvested when an OD_{600} of 0.2-0.3 was reached. Cells were centrifuged at 1,500 x g for 10 min. The supernatant was discarded.

Preparation of Spheroplasts

- The pellet from the overnight *P. pastoris* culture was resuspended in 20 ml of sterile water and was transferred to a fresh 50 ml tube.

- The cells were pelleted by centrifugation at 1,500 x g for 5 min. The supernatant was discarded.

- The cell pellet was washed once by resuspending in 20 ml of fresh Sorbitol/EDTA/DTT (SED) and centrifugation was carried out as before.

- The cells were washed once with 20 ml of 1 M sorbitol and centrifuged.

- Cells were resuspended by swirling them in 20 ml of Sorbitol/Sodium Citrate/$CaCl_2$ (SCE) buffer and the suspension was divided into 2 x 50 ml conical tubes.

- Zymolyase was removed from –20^{0}C and was placed on ice. The optimal time of digestion with zymolyase to make spheroplasts was determined as outlined in the manufacturer's manual (Invitrogen). 7.5 µL of zymolyase was added to the cells and incubation at 30^{0}C was carried out for the pre-

determined length of time as suggested by the manual. Cells were then harvested by centrifugation at 750 x g for 10 min and the supernatant was decanted and discarded.

- Spheroplasts were washed once with 10 ml of 1 M sorbitol and were collected by centrifugation at 750 x g for 10 min.

- Spheroplasts were washed once with 10 ml of Sorbitol/Tris-HCl/$CaCl_2$ (CaS) and centrifuged as before. They were then resuspended in 0.6 ml of CaS for immediate use.

The solutions required for the spheroplast preparation and transformation are outlined in Table 12.

Table 12: Spheroplasting and Transformation Solutions

Reagent	Amount	Components
SOS medium	20ml	1 M sorbitol, 0.3 X YPD, 10 mM $CaCl_2$
Sterile Water	2 x 125 ml	Autoclaved, deionised water
SE	2 x 125 ml	1M sorbitol, 25 mM EDTA, pH 8.0
SCE	2 x 125 ml	1 M sorbitol, 10 mM sodium citrate buffer, pH 5.8, 1 mM EDTA
1 M sorbitol	2 x 125 ml	22.5 g D-sorbitol/125 ml dH_20
CaS	2 x 60 ml	1 M sorbitol, 10 mM Tris-HCL, pH 7.5, 10 mM $CaCl_2$
40% PEG	25 ml	40% (w/v) PEG 3350 in water
CaT	25 ml	20 mM Tris-HCL, pH 7.5, 20 mM $CaCl_2$
Zymolyase	2 x 20 µL	3 mg/ml zymolyase in H_20 (100,000 units/g lytic activity)
1 M DTT	10 x 1 ml	1.54 g DTT in 10 ml dH_20

Transformation of *P.pastoris*

Approximately 10 μg of the expression cassette was used incubated with 100 μL of spheroplasts for 10 min at RT. One ml of a Polyethylene Glycol (PEG)/Tris-HCL, $CaCl_2$ (CaT) solution was then added to the cells and DNA, followed by gentle mixing and incubation at RT for 10 min. Centrifugation at 750 x g ensued and the PEG/CaT solution was carefully aspirated. The pellet of the transformed cells was resuspended in 150 μL sorbitol/YPD/$CaCl_2$ (SOS) media and was incubated at RT for 20 min. The next step involved the addition of 850 μL of 1 M sorbitol. *P.pastoris* spheroplasts have to be plated in top agar to protect them from lysis prior to selection. One-three hundred μL of each spheroplast-DNA solution was mixed with 10 ml of molten Regeneration Dextrose (RD) agarose, before pouring onto RDB plates. To ensure that the spheroplasts were viable, 100 μL of the spheroplasts were mixed with 900 μL of 1 M sorbitol and 100 μL of this was mixed with 10 ml of molten RDH before pouring onto an RDBH plate. The top agarose on all the plates was allowed to harden prior to incubation of the plates at 28-30^{0}C for 4-6 days.

Media Components

Table 13: Additives for transformation and plating media

10X YNB (yeast nitrogen base with ammonium sulfate without amino acids)	134 g YNB with ammonium sulfate without amino acids/L, 34 g YNB w/o ammonium sulfate and amino acids and 100g ammonium sulfate/L
500X B (0.02% biotin)	20 mg biotin/100 ml H_20, filter sterilise
100X H (0.4% histidine)	400 mg L-histidine/100 ml H_20, heat to dissolve, filter sterilise
10X D (20% dextrose)	200 g D-glucose/L, autoclave or filter sterilise
10X M (5% methanol)	5 ml methanol in 95 ml H_20, filter sterilise
10X GY (10% glycerol)	100 ml of glycerol/900 ml H_20, autoclave or filter sterilise
100X AA (0.5% of each amino acid)	500 mg of L-glutamic acid, L-methionine, L-lysine and L-isoleucine/100 ml H_20, filter sterilise
1 M potassium phosphate buffer, pH 6.0	132 ml 1 M K_2HPO_4, 868 ml of 1 M KH_2PO_4, pH to 6.0 and autoclave
100 mg/ml G418 stock	100 mg/ml in sterile H_20, filter sterilise, store at -20^0C

Plating of Transformants onto G418

When transformants appeared onto the plates the top agar was removed and the cell density was determined for each transformation using a haemocytometer. Onto each YPD-G418 plate, with concentrations of G418 from 0.25 mg/ml-1.0 mg/ml, 10^5 cells were plated. Cells were also plated onto YPD without antibiotic as a control. Plates were incubated at 30^0C for 2-5 days. After this time, colonies that appeared on the YPD-G418 plates were isolated and streaked onto YPD. A second YPD-G418 screening was done 3 days later using the colonies on YPD, incubation as before. Table 14 outlines the components of the media needed for the transformation process and subsequent plating steps

carried out during the expression process. The solutions that must be added to the media (Table 14) are detailed in Table 13.

Table 14: Media Components

Yeast Extract Peptone Dextrose Medium (YPD)	10 g yeast extract, 20 g peptone, (20 g agar for plates)/L autoclave, cool, add 100 ml 10X D
YPD-G418 plates	YPD plates with 0.25-4.0 mg/ml G418
Regeneration Dextrose Medium + Histidine (RD and RDH) Liquid Media	186 g sorbitol/700 ml H_20, autoclave, cool and store at 45^{0}C, add pre-warmed mixture of 100 ml of 10X D, 100 ml of 10X YNB, 2 ml of 500X B, 10 ml of 100X AA to sorbitol
RDB and RDH Agar Plates	186 g sorbitol, 20 g agar/700 ml H_20, autoclave, place in 60^{0}C water bath, add same pre-warmed mixture as for RD and RDH Liquid Media, if selecting for His^+ transformants omit histidine, pour immediately
RD and RDH Top Agar	186 g sorbitol, 10 g agar/700 ml H_20, autoclave, place in 60^{0}C water bath, add the same pre-warmed mixture as for RD and RDH Liquid Media, keep at 45^{0}C and use during transformation
MD Agar Plates	15 g agar/800 ml H_20, autoclave, cool and add 100 ml 10X YNB, 2 ml of 500X B and 100 ml 10X D, cool and pour
MM Agar Plates	15 g agar/800 ml H_20, autoclave, cool, add 100 ml of 10X YNB, 2 ml of 500X B and 100 ml 10X M, cool and pour
Buffered Glycerol-complex Medium (BMGY) and Buffered Methanol-complex Medium (BMMY)	10 g yeast extract, 20 g peptone/700 ml H_20, autoclave, cool, add 100 ml 1 M potassium phosphate, pH 6.0, 100 ml 10X YNB, 2 ml 500X B, 100 ml 10X GY (BMGY only) and 100 ml 10X M (BMMY only)
Breaking Buffer	6 g monobasic sodium phosphate, 372 mg EDTA and 50 ml glycerol/900 ml H_20, pH to 7.4, store at 4^{0}C,add protease inhibitors (e.g., PMSF) before use

MM/MD Plating

A single colony was picked off the YPD-G418 plate and was dotted onto both an MM and MD plate. Extra care was taken to ensure that the MM plate was dotted first. Plates were incubated at 30^0C for 2 days. The purpose of MM/MD plating was to phenotypically differentiate between those colonies that were Mut^+ (Methanol utilisation plus) and Mut^S (Methanol utilisation slow). Mut^S phenotype exhibits a reduced ability to metabolise methanol as a result of the loss of the AOX1 gene. Mut^+ phenotype represents the wild type strain that has its full methanol utilising capabilities intact. As the expression cassette had to be linearised with the restriction enzyme *BglII* prior to transformation, all transformants were of the Mut^S phenotype as the integration event that occurred in *P. pastoris* was a gene replacement event at the AOX1 locus.

Direct PCR Screening of *Pichia pastoris* Clones

P. pastoris clones were directly tested for insertion of SmCC by PCR.

A single colony was dissolved in 10 µL H_20. This was followed by the addition of 5 µL of a 5U/µL solution of lyticase and incubation at 30^0C for 10 min. A PCR reaction was set us as follows for each tube:

10X Reaction Buffer	5 µL
25 mM $MgCl_2$	5 µL
25 mM dNTPs	1 µL
5' AOX1 primer (10 pmol/µL)	1 µL
3' AOX1 primer (10 pmol/µL)	1 µL
Sterile water	31 µL
Cell lysate	5 µL
Taq	1 µL (added after 1st denaturing step)
	50 µL

Thirty cycles were carried out

Table 15: PCR Screening of *P. pastoris* Clones

Step	Temperature	Time
Denaturation	95^0C	1 minute
Annealing	54^0C	1 minute
Extension	72^0C	1 minute
Final extension	72^0C	7 minute

A 10 µL aliquot of each PCR reaction was analysed by agarose gel electrophoresis.

2.2.6: FUNCTIONAL EXPRESSION OF SMCC IN *P. PASTORIS*

The production of rSmCC was under the control of the tightly regulated, inducible AOX1 promoter. Recombinant protein production was repressed by glucose/glycerol but was induced by growth in the presence of methanol. Only clones proven to contain SmCC integrated into the *P. pastoris* genome were used for the induction process.

Standard Methanol Induction

100 ml of BMGY in a 1 L baffled flask were inoculated with a single colony of transformed, PCR positive *P. pastoris* from a YPD plate. Incubation on a 30^0C shaker at 225 r.p.m ensued until an OD_{600} 2.0-6.0 was reached. Centrifugation at 3,000 x g for 5 min followed and the supernatant was decanted. For Mut^S phenotype which all transformants in this instance were, the pellet was resuspended in BMMY pH 6.0 using 10-20% of the original culture volume for resuspension (10-20 ml). A 1 ml sample of culture (D0) was taken prior to incubation on the 30^0C shaker at 225 r.pm. A 1 ml sample of culture was taken

every day for analysis and 100% methanol was added to a final volume of 1% every day to maintain induction. Each 1 ml sample was spun at 2,460x g for 5 min and the supernatant was removed. Each pellet and supernatant was stored at 4^0C until further analysis could be carried out. Variations on this basic induction protocol were carried out, for example 100% methanol was added to a final concentration of 1.5% twice daily for a six-day period. These variations on the standard protocol will be discussed in detail in Section 3.3.

Analysis of Protein by SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Both supernatants and pellets were analysed for rSmCC production by SDS-PAGE. A 12% separating gel and 5% stacking gel was used in all cases.

Table 16: SDS-Gel Preparation

Components (for 4 mini-gels)	**Separating Gel (12%)**	**Stacking Gel (5%)**
30% Bis Acrylamide	12 ml	1.66 ml
1 M Tris-HCL	11.2 ml	1.25 ml
dH_20	6.7 ml	7.04 ml
10% SDS	300 µL	100 µL
10% ammonium persulfate	150 µL	100 µL
TEMED	50 µL	50 µL

Note: For the separating gel, 1 M Tris-HCL pH 8.8 was used whereas for the stacking gel 1 M Tris-HCL pH 6.8 was used.

Other reagents required for SDS-PAGE are outlined in Table 17:

Table 17: SDS-PAGE Reagents

Electrophoresis Buffer, pH 8.3	3 g/L Tris 14.4 g/L Glycine 10 ml 10% SDS, pH to 8.3
10% SDS	10 g SDS/100 ml ultra pure H_20
10% ammonium persulfate (APS)	0.2 g APS in 2 ml H_20
Sample Buffer (4X)	2.5 ml 1 M Tris-HCL pH 6.8 4 ml 100% Glycerol 0.8 ml β-mercaptoethanol 0.04 g Bromophenol Blue 0.8 g SDS Make up to 10 ml with dH_20
Coomassie Blue Stain	10 g/L 1% Brilliant Blue R 250 400 ml 40% Methanol/L 100 ml 10% Acetic Acid/L Make up to 1 L with dH_20

For analysis, 5 μL of reducing sample buffer (4X) was added to 10 μL of the sample. Boiling at 100^0C for 3 min ensued. Samples were loaded onto the gel (~10-15 μL per well) alongside a protein marker.

Disruption of Pellets

Although supernatants could be directly analysed on the gel, pellets had to be disrupted prior to analysis to liberate any protein from within the yeast cells. Pellets were first thawed quickly on ice. For each 1 ml sample 100 μL of breaking buffer was added to the cell pellet, which was resuspended. An equal volume of acid washed glass beads were added to the sample. Vortexing for 30 sec was followed by incubation of the pellets on ice for 30 sec. This step was repeated eight times. Pellets were centrifuged at 2,460 x g for 10 min and the

resulting supernatant was transferred to a clean microcentrifuge tube. Analysis by SDS-PAGE could then proceed.

All gels were run at 25 mA for 1.5 hrs. Gels were stained overnight in coomassie blue stain, followed by destaining in 10% acetic acid, 40% methanol.

Western Blot Analysis

A western blot was carried out on all samples employing anti-His_6 antibodies or anti-*Schistosoma japonicum* cathepsin C (SjCC).

Reagents required for the Western Blot protocol are outlined in the table below:

Table 18: Western Blot Reagents

10X Phosphate Buffered Saline (PBS) pH 7.0	80 g/L NaCl 2 g/L KCL 11.6 g/L Na_2HPO_4 2 g/L KH_2PO_4, adjust to pH 7.0, dilute to 1X
Transfer Buffer	3.03 g/L Tris 14.4 g/L Glycine 100 ml 100% Methanol Adjust vol. to 500 ml with dH_20, pH to 8.1-8.5

SDS/PAGE was carried out as normal. Eight pieces of filter paper, cut to the size of the gel were soaked in transfer buffer. Nitrocellulose membrane was also soaked in transfer buffer for 15 min. When migration of the protein samples on the gel was sufficient, the gel was also soaked in transfer buffer for

5-10 min. Four pieces of filter paper were first placed on the blotter followed by nitrocellulose membrane, followed by the gel and lastly by the remaining four pieces of filter paper. Transfer was carried out at 15 V for 20 min. After transfer was complete blocking was carried out with 1X PBS, 5% Marvel for 1 hr. This was followed by two five-minute washes of the membrane with 1 X PBS- 1% Tween. Between each stage of the western blot procedure three such five-minute washes were carried out on the membrane. The primary antibodies, namely the anti-His_6 antibody or the anti-SjCC antibody were added to the membrane at a dilution of 1:3000 and 1:200 respectively in 1X PBS-Tween-1% Marvel for 1 hr. Next, the secondary antibody was then added to the membrane for 1 hr. This was goat anti-mouse IgG (1:1000 dilution in PBS-Tween) when the His tag was been probed and anti-rabbit IgG (1:1000 dilution in PBS-Tween) when SjCC was been probed. Any binding that occurred was visualised with DAB peroxidase.

Colony Blot Procedure

The colony blot procedure was implemented to determine whether *P. pastoris* clones actually expressed SmCC by probing with antibodies to His_6.

The solutions outlined in Table 19 were needed for the colony blot procedure.

Table 19: Colony Blot Solutions

SDS solution	10% (w/v) sodium dodecyl sulfate
Denaturing solution	0.5 M NaOH, 1.5 M NaCl
Neutralisation solution	1.5 M NaCl; 0.5 M Tris-CL, pH 7.4 (25^0C)
20X SSC	500 ml: 87.65 g NaCl, 50.25 g trisodium citrate.$2H_20$
PBS buffer	As for western blot procedure
PBS-Tween	As for western blot procedure
YNB plates	134 g/L YNB, 15 g/L agar, 2% final vol. filter sterilised MeOH added after autoclaving

Colonies expected to express SmCC were dotted in duplicate onto a YPD plate, followed by incubation at 30^0C overnight. The next day these colonies were blotted onto sterile nitrocellulose membrane, which was then transferred to a YNB plate containing 2% methanol. The plates were incubated at 30^0C. After 4-6 days the nitrocellulose membranes were incubated colony side up on filter paper in petri dishes in the following solutions:

Table 20: Incubation Times

SDS solution	10 min
Denaturing solution	5 min
Neutralisation solution (x 2)	5 min
SSC (x 2)	15 min

Filters were washed twice in PBS-Tween for 10 min each between all stages. Blocking was carried out using PBS-Tween 5% Marvel for 1 hr. Next step was the addition of the primary antibody which was Anti-His_6 at a 1:1000 dilution in PBS-Tween–1% Marvel for 1 hr. Afterwards the secondary antibody was applied. The secondary antibody was goat-anti-mouse IgG at a 1:1000 dilution in PBS-Tween-1% Marvel and incubation with this antibody proceeded for 1 hr. DAB Peroxidase substrate was then applied to the membrane enabling expressing colonies to be visualised on the membrane.

ELISA Technique

For the ELISA technique the following solutions were needed:

Table 21: ELISA Reagents

0.1 M sodium carbonate buffer, pH 9.6	8 ml of Na_2CO_3 (21.6 g/L) 17 ml of $NaHCO_3$ (16.8 g/L) Add 25 ml dH_20
Blocking Solution	1% BSA 10% sucrose in carbonate buffer

50 µL of the sample and 150 µL of carbonate buffer were used to coat the plate, which was stored overnight at 4^0C. The plate was washed four times with 1X PBS-1% Tween, followed by application of 200 µL of blocking solution/well. The plate was then incubated at 37^0C for 2 hrs. The aforementioned washing step was carried out on completion of each stage of the ELISA protocol. The next stage was binding of the primary antibody. 100 µL of the antibody/well, diluted in PBS-Tween-Marvel was added for 1 hr. After washing, the secondary antibody was applied, using 100 µL of the diluted antibody/well as before for 1 hr. The primary and secondary antibodies were as for the western

blot procedure. Lastly, 100 µL TMB substrate was added to each well for 10 min. The reaction was stopped on addition of 100 µL of 1N H_2SO_4. Absorbance was read at 450 nm and the result was compared to a set of positive and negative controls.

Ni-NTA Beads Gel

This technique was used to detect the presence of an intact His-tag in the rSmCC samples.

- 200 µL of sample and 800 µL lysis buffer were combined
- 100 µL of Ni-chelate beads (50% suspension) for each sample were washed twice with 1 ml of lysis buffer. After each wash, centrifugation at 2,460 x g for 5 min was carried out and the lysis buffer was removed from the beads before proceeding to the next step.
- The sample from the first step was added to the beads and incubation for 1 hr with shaking followed.
- Beads were then spun at 2,460 x g for 5 min and the supernatant was removed.
- Washing of the beads with lysis buffer was carried out as before and the lysis buffer was subsequently removed from the beads.
- 10 µL of H_20 and 5 µL 4X sample buffer was added to the beads. The resuspended beads were heated to 100^0C for 3 min.

- Beads were centrifuged as before and the supernatant was transferred to a fresh tube prior to analysis of a 15 µL sample of each supernatant by SDS-PAGE.

Affinity Chromatography on Ni-NTA Agarose

Before affinity chromatography could be carried out, the following solutions had to be prepared:

Table 22: Stock Solutions for Affinity Chromatography

5X Monobasic sodium phosphate	35.5 g/L monobasic sodium phosphate (250 mM) 87.66 g/L NaCl (1.5 M)
5X Dibasic sodium phosphate	35.5 g/L dibasic sodium phosphate (250 mM), 87.66 g/L NaCl (1.5 M)
3 M Imidazole (store at 4^0C)	5.106 g/25 ml dH_20 imidazole

- The sample was centrifuged at 1,260 x g for 20 min. A 1:5 dilution of supernatant sample in lysis buffer was performed.

- 2 ml of Ni-NTA beads were poured onto the column, which was then equilibrated with 3 ml of lysis buffer.

- 150 ml of the sample was applied to the column over time and the run through was collected.

- 15 ml of wash solution was added to the column and collected.

- 5 ml of elution buffer was applied next and was collected.

- 5 ml of "flush" was then added to the column.

- The eluate was dialysed overnight in 2 L 1X PBS. The next day the PBS solution was changed twice and dialysis proceeded for a further 4 hrs

- Collected samples from all stages of the process were analysed by SDS-PAGE and the protein concentration in all samples was determined.

Table 23: Affinity Chromatography Reagents

5X Sodium Phosphate Buffer, pH 6.0	Monobasic:Dibasic 87.7:12.3 (need ~ 50 ml)
5X Sodium Phosphate Buffer, pH 7.0	Monobasic:Dibasic 39:61 (need ~ 50 ml)
5X Sodium Phosphate Buffer, pH 8.0	Monobasic:Dibasic 2:98 (need ~ 100 ml)
Lysis Buffer, pH 8.0 (200 ml)	40 ml 5X sodium phosphate pH 8.0, 667 µL of 3 M imidazole (final conc. 10 mM), adjust vol. to 190 ml with dH_20, pH to 8.0, final vol. of 200 ml
Wash Buffer, pH 8.0 (100 ml)	20 ml 5X sodium phosphate pH 8.0, 667 µL 3M imidazole (final conc. 20 mM), bring to 99 ml with dH_20, pH to 8.0, adjust vol. to 100 ml
Elution Buffer, pH 7.0 (20 ml)	4 ml 5X sodium phosphate pH 7.0, 1.667 ml imidazole (final conc. 250 mM), bring vol. to 19 ml, pH to 7.0, adjust. to 20 ml
Flush, pH 6.0 (20 ml)	As for elution buffer except use 5X sodium phosphate buffer pH 6.0
10X PBS	See western blot solutions

2.2.7: BIOCHEMICAL CHARACTERISATION OF *FASCIOLA HEPATICA* CATHEPSIN C (FHCC), *S. MANSONI* CATHEPSIN C (SMCC) AND RECOMBINANT *S. MANSONI* CATHEPSIN C (RSMCC)

Enzyme assays had to be carried out to determine whether an active SmCC was produced by *Pichia pastoris*. Prior to performing any enzyme assays a set of suitable, accurate standards had to be prepared:

Preparation of 7-amino-4-methyl-coumarin (NHMec) standards

A 500 μM stock of NHMec was prepared. The stock was prepared in 2 ml of dimethylformamide (DMF) to dissolve the fluorogenic NHMec and the volume was adjusted accordingly with dH_20. Dilutions of the stock NHMec were performed so that a set of standards were obtained with NHMec concentrations in the range of 50 μM-0.78 μM. 1 ml of each standard was incubated at 37^0C along with enzyme assays and 200 μL of the stopping reagent, 10% acetic acid was added to each standard after incubation. All standards were read on a fluorimeter at slit width 10:10 or 10:2.5 with an excitation wavelength of 370 nm and an emission wavelength of 440 nm. A standard curve was prepared plotting average fluorescence units against NHMec concentration (μM).

Preparation of *F.hepatica* somatic extract

Fifteen ml of Tris-HCL pH 7.0 buffer was added to whole, intact liver flukes. The flukes were homogenised until they had sufficiently broken up. Homogenisation was performed on ice. Homogenate was frozen at -80^0C followed by thawing on ice and brief sonication. This step was repeated three times. Centrifugation was carried out at 1,260x g for 1 hr. The supernatant was removed and this somatic extract was used in all assays as a positive control. The remaining pellet was resuspended in 0.1% Tris-Triton X and was centrifuged for 1 hr. The resultant supernatant was removed and this was the

detergent extract. The pellet was discarded. The concentration of protein in the somatic extract was estimated using the BCA protein assay procedure.

Standard Enzyme Assay

- The pH optimum of SmCC was reported as pH 5.5. Therefore, a sodium acetate buffer, pH 5.5 was employed in all assays unless stated otherwise. A series of pH profile assays were carried out to investigate the effect of pH on cathepsin C activity. A series of buffers were prepared namely, sodium citrate, pH 3.0, 4.0, 5.0, 6.0 and 7.0, sodium acetate pH 4.0, 4.5, 5.0, 5.5 and 6.0, sodium phosphate, pH 5.0, 5.5, 6.0, 6.5, 7.0 and 7.5, Tris-HCL pH 6.0, 7.0, 7.5, 8.0 and 8.5 and glycine pH 8.5, 9.0, 9.5, 10.0 and 10.5.

- DTT was known to increase the activity of cathepsin C so it was added to all assays to a final concentration of 2 mM in most assays. A 20 mM assay stock was prepared and 100 µL of this stock was used in an assay volume of 1ml for rSmCC assays. For FhCC assays 15-100 µL of 2 mM DTT was used.

- The fluorogenic cathepsin C substrate H-Gly-Arg-NHMec was used in all assays. A 10 mg/ml stock of this substrate was prepared in DMF. A 100 µmolar working stock was prepared for all assays and this was prepared by adding 1 µL of the stock in DMF to 257.5 µL dH_20. In a 1 ml tube assay 100 µL of this working stock was used so the final concentration of substrate in the assay was 10 µmolar.

Table 24: Standard cathepsin C Assay

Component	Cathepsin C assay	Positive control	Background control
Buffer	790 µL	790 µL	800 µL
DTT	100 µL	100 µL	100 µL
Enzyme	10 µL	10 µL	N/A
Substrate	100 µL	100 µL	100 µL
Final volume	**1 ml**	**1 ml**	**1 ml**

For rSmCC assays the enzyme was the supernatant sample from the induction process in *P. pastoris*. For the positive control, the enzyme refers to the *F. hepatica* somatic extract. This somatic Fh extract was also used as the source of FhCC for biochemical characterisation studies. The background control contained all assay components except the enzyme and was used to zero the fluorimeter on reading the assays.

Assays were initially carried out at 37^0C. However because rSmCC activity proved to be low it was necessary to perform overnight assays. The concentration of NHMec released during the assay could be determined from the standard curve. Enzyme activity was calculated in units (U). 1 U is defined as 1 nmol of NHMec released/min/ml of enzyme.

The enzyme reaction was stopped on addition of 200 µL of 10% acetic acid. All assays were read as for the NHMec standards.

Enzyme Inhibition Assays

Inhibition assays were carried out on all samples to determine whether the enzyme produced was a cysteine protease that could be inhibited. Unlike proteases such as aminopeptidase, cathepsin C does not possess its own specific

inhibitor so it was necessary to use a general cysteine protease inhibitor. Leupeptin was the inhibitor of choice in these inhibition assays and it was prepared at a stock concentration of 100 μM. In each assay 100 μL of inhibitor was used in a 1 ml assay volume giving a final inhibitor concentration of 10 μM in the assay.

The assay procedure was carried as before except the inhibitor was incubated with the enzyme and assay components for 10 min prior to addition of the enzyme.

Table 25: Sample Inhibition Assay

Component	**Inhibition Assay**	**Control**	**Background**
Buffer (pH 5.5)	690 μL	690 μL	700 μL
DTT	100 μL	100 μL	100 μL
Enzyme	10 μL	10 μL	N/A
Inhibitor	100 μL	100 μL	100 μL
Substrate	100 μL	100 μL	100 μL
Final volume	**1 ml**	**1 ml**	**1 ml**

For the control, the enzyme refers to the *F. hepatica* somatic extract. A regular enzyme assay without inhibitor was carried out together with the inhibition assay to enable accurate comparisons to be made. Inhibition assays were performed overnight at 37^0C unless stated otherwise.

The effect of different cysteine proteinase inhibitors on SmCC and rSmCC activity was investigated. Inhibitors used for this study included: iodoacetamide, leupeptin, N-ethyl-malelimide (NEM), trans-epoxysuccinyl-L-leucylamide (4-guanidinol)-butane (E-64).

Iodoacetamide Inhibition Assay

The assay protocol was as for leupeptin assay with the exception of the concentration of iodoacetamide used. A stock solution of 1 mM iodoacetamide was used and the final inhibitor concentration in the assay was varied. The amount of buffer used was adjusted accordingly to give a final assay volume of 1 ml.

Table 26: Iodoacetamide Inhibition Assay

Component	Assay 1	Assay 2
Buffer (pH 5.5)	690 µL	780 µL
DTT	100 µL	100 µL
Inhibitor	100 µL	10 µL
Enzyme	10 µL	10 µL
Substrate	100 µL	100 µL

In assay 1 the final concentration of inhibitor in the assay was 100 µM whereas in assay 2 it was 10 µM. Controls were prepared as for the leupeptin assay detailed in Table 25.

NEM Inhibition Assay

A 5 mM stock solution of NEM was prepared and 200 µL of this was used in the assay to give a final assay concentration of 1 mM.

E-64 Inhibition Assay

E-64 is considered a slow inhibitor of cathepsin C whereas it is an instantaneous inhibitor of cathepsin L. An E-64 time course assay was carried out. Cathepsin L1 was used as a comparison using a 100 µM stock of Z-Phe-

Arg-NHMec, the substrate for cathepsin L1. The enzyme preparation used for all assays was *F. hepatica* somatic extract. The time that 1 mM E-64 was incubated with the enzyme prior to the addition of the substrate varied.

Table 27: E-64 Assay Procedure

Sample	**Ctrl. 1**	**Ctrl. 2**	**0**	**5**	**10**	**20**	**40**	**80**
Buffer, pH 5.5 (µL)	795	790	785	785	785	785	785	785
DTT (µL)	100	100	100	100	100	100	100	100
Enzyme (µL)	N/A	10	10	10	10	10	10	10
1 mM E-64 (µL)	5	N/A	5	5	5	5	5	5
Substrate (µL)	100	100	100	100	100	100	100	100

After incubation of E-64 with the enzyme for the desired time, substrate was added and incubation at 37^0C for 30 min ensued. As always the reaction was stopped on addition of acetic acid and results were read on the fluorimeter.

Halide Enhancement Assays

Cathepsin C is enhanced in the presence of halide ions. Sodium iodide (NaI), sodium bromide (NaBr) and sodium fluoride (NaF) at a stock concentration of 20 mM and sodium chloride (NaCl) at a stock concentration of 10 mM were used in the assay. A control assay that did not contain halides was set up for comparative purposes. The final assay volume was 1 ml in all instances.

Final concentrations of halides in the assays were as follows:

Table 28: Halide Enhancement Assays

Halide	Concentration
NaCl	0.1 mM, 1 mM
NaI	0.1 mM, 1 mM, 10 mM
NaBr	0.1 mM, 1 mM, 10 mM
NaF	0.1 mM, 1 mM, 10 mM

Assays were carried out as normal using 10 μL of undiluted enzyme in pH 5.5 buffer, 100 μL of 2 mM DTT, incubation as usual.

Determination of Protein Concentration-BCA Protein Assay

Prior to performing the assay, bovine serum albumen standards were prepared at the following concentrations: 2mg/ml, 1.75 mg/ml, 1.5 mg/ml, 1.25 mg/ml, 0.75 mg/ml, 0.5 mg/ml and 0.25 mg/ml. 25 μL of each standard was plated in duplicate onto rows A and B of a 96 well microtitre plate. 25 μL of the protein sample whose concentration was to be determined was plated at different dilutions (e.g., neat, 1:2, 1:5, 1:10) in duplicate onto rows C and D of the plate. BCA working reagent was prepared in accordance with the protocol outlined by the manufacturer (Pierce) and 200 μL of this working reagent was added to each well. Gentle shaking of the plate ensured adequate mixing of the assay constituents. The plate was incubated at 37^0C for 30 min after which the absorbance at 570 nm was read. A BCA standard curve was prepared plotting A_{570} against protein concentration (mg/ml). The concentration of protein in the sample could be determined from this standard curve. The specific activity of SmCC could be determined by dividing the enzyme activity (U) by the total protein concentration (mg/ml).

3: RESULTS

3.1: Isolation and amplification by polymerase chain reaction (PCR) of *Schistosoma mansoni* cathepsin C (SmCC) cDNA from a cDNA library

3.1.1: Early Attempts at PCR Amplification of the 1.3 kb SmCC cDNA and the discovery of the 450 bp fragment

A number of primers were designed for the amplification of SmCC cDNA from a cDNA library in pB42AD as outlined in Figure 7.

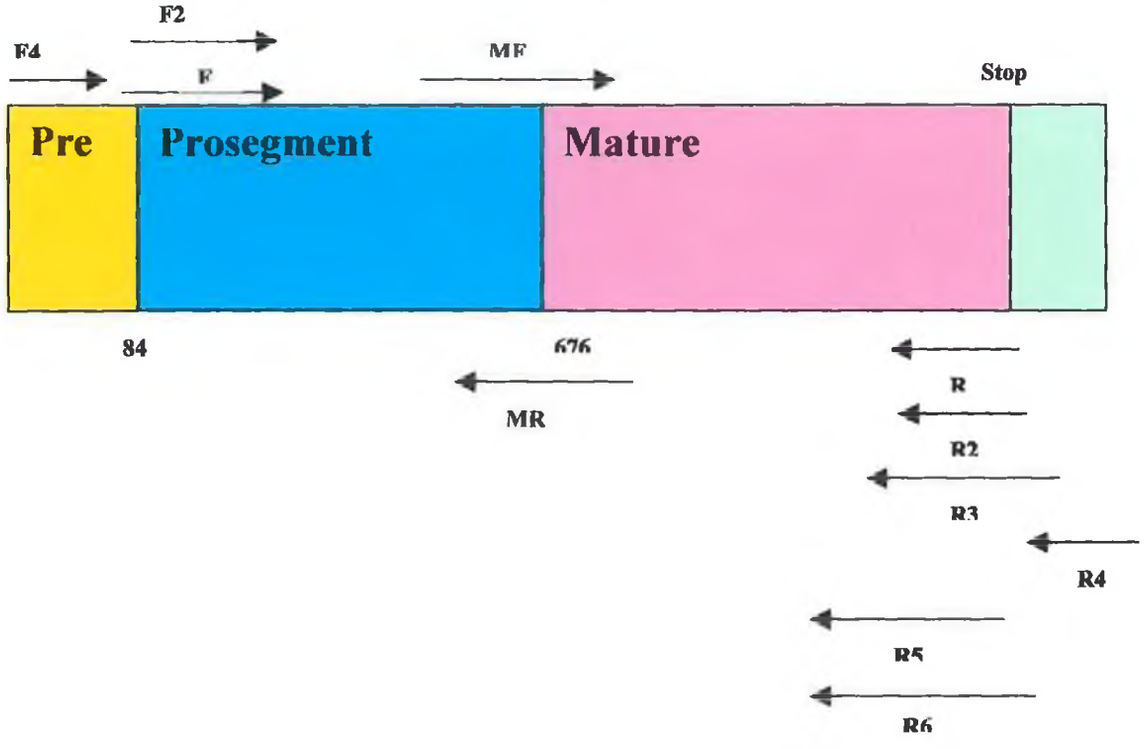


Figure 7: Diagram showing location of primers on the SmCC cDNA

For a detailed listing of all primer sequences and location on the SmCC cDNA refer to Tables 2 and 3 in the materials and methods section (section 2.2).

Table 29: The primer pairs used in the amplification process

Primers	Expected size of PCR product	Actual size of PCR product
SmCCF & SmCCR	1.353 kb	450 bp
SmCCF2 & SmCCR2	1.307 kb	450 bp
SmCCF2 & SmCCMR	601 bp	601 bp
SmCCR2 & SmCCMF	728 bp	728 bp
SmCCF2 & SmCCR3	1.311 kb	450 bp
SmCCF2 & pB42ADF	1.35 kb	none obtained
SmCCF2 & pB42ADR	1.35 kb	850 bp, 350 bp & 150 bp
SmCCMF &pB42ADF	778 bp	none obtained
SmCCMF &pB42ADR	778 bp	none obtained
SmCCMF & SmCCR3	731 bp	731 bp
SmCCF4 & SmCCR3	1.366 kb	480 bp
SmCCF4 & SmCCR4	1.393 kb	480 bp
SmCCF2 & SmCCR5	1.298 kb	1.298 kb
SmCCF4 & SmCCR5	1.353 kb	1.353 kb
SmCCF2 & SmCCR6	1.304 kb	1.304 kb
SmCCF & SmCCR6	1.315 kb	1.351 kb
SmCCF4 & SmCCR6	1.359 kb	1.359 kb
SmCCMF & SmCCR5	718 bp	718 bp

Initially when using most of the primer sets a small band of 450/480 bp was visible on the gel instead of the 1.3 kb anticipated PCR product (see Table 29 and Figure 8).

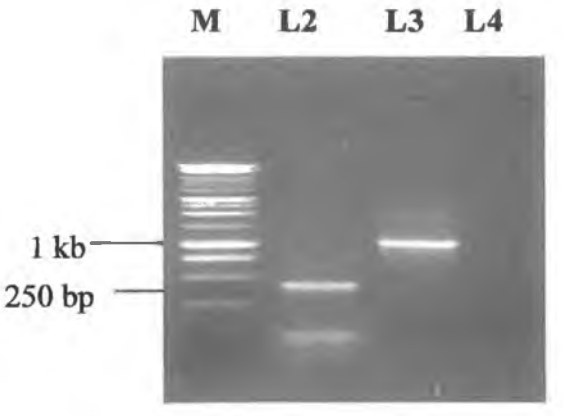


Figure 8: The 450 bp band (lane 2) as seen on a 1% agarose gel

Lane 1 contains the 1 kb molecular weight marker (M). The 450 bp band in L2 was amplified using the primers SmCCF2 and SmCCR2. The positive control (L3) was PH 1, a reverse transcriptase gene from S. mansoni in pBluescript amplified with M13 F and R primers. The negative control (L4) consisted of all reaction components except template DNA so no band is present in this lane as expected.

For the majority of PCR reactions the following cycle was used:

- Denaturing: 95^{0}C for 5 min (1st cycle only)
- Denaturing: 95^{0}C for 1 min 30 sec
- Annealing: 55^{0}C for 1 min 30 sec — 40 cycles
- Extension: 72^{0}C for 1 min 30 sec
- Extension: 72^{0}C for 10 min (last cycle only)

However, as successful amplification of the full-length SmCC cDNA with initial primers did not occur it was necessary to vary the conditions of the PCR reaction.

- Firstly the $MgCl_2$ concentration was varied in 0.5 mM increments from 0.5 mM to 3 mM.
- Secondly, the concentration of dNTPs and $MgCl_2$ were varied, $MgCl_2$ as before and dNTPs in 0.1 mM increments from 0.1 mM to 0.5 mM.
- Annealing temperature was varied in 5^0C increments from 35-65^0C.

These variations did not result in the full-length cDNA being successfully amplified.

On one occasion the full 1.3 kb SmCC cDNA was faintly visible relative to the very bright 450 bp band after amplification with SmCCF2 and SmCCR2 (Figure 9). However, re-amplification with gel purified SmCC and the same primer sets proved to be unsuccessful. Subsequent amplifications with the SmCCF2, SmCCR2 and cDNA template failed to amplify SmCC again.

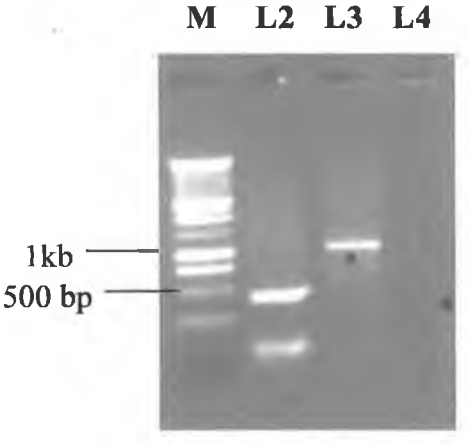


Figure 9: Faintly visible 1.3 kb SmCC cDNA relative to the 450 bp PCR product

The 1 kb molecular weight marker is depicted in L1. In L2 the PCR products at 450 bp and 1.3 kb of amplification with SmCCF2 and

SmCCR2 are visible. The positive and negative controls (L3 and L4) were as before (Figure 8).

Primers, SmCCMF and SmCCMR were designed to amplify a region in the middle of SmCC to prove that the full cDNA was present in the library. The primer sets used for this reaction are depicted in Table 29.

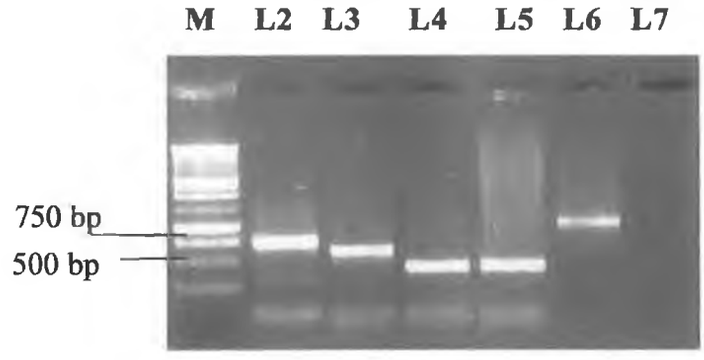


Figure 10: PCR amplification using "middle" SmCC specific primers

L2 shows the 728 bp PCR product obtained when PCR with SmCCR2 and SmCCMF was carried out. The 601 bp DNA in L3 was obtained using SmCCF2 and SmCCMR. In L4 and L5 SmCCF2 and SmCCR3 amplified the usual 450 bp product using a 1:10 diluted and neat cDNA respectively. The usual PH 1 positive control (L6) and negative control (L7) illustrated that the PCR had worked.

A PCR reaction was set up using primers specific for pB42AD, pFor and pRev to amplify the full length SmCC cDNA. By virtue of the fact that cloning of the cDNA library into pB42AD was non-directional pFor was used together with both forward and reverse SmCC specific primers.

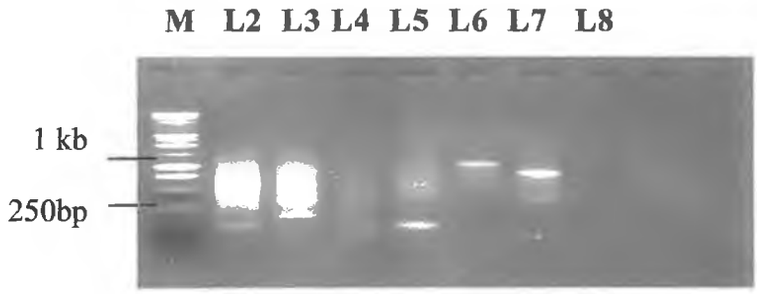


Figure 11: PCR using pFor and pRev plasmid-specific primers

The 1 kb marker is present in L1. In L2 the use of pFor and SmCCF2 failed to produce a clear product. In L3 bands at 900 bp, 350 bp and 150 bp are clear. The 900 bp band is also visible in L7. The DNA in L7 was amplified using pRev primer only and cDNA template. No band is visible in L4 using SmCCMF and pFor primers. In L5 with SmCCMF and pRev a band is present at <200 bp but this is probably primer-dimers or unused primers. The presence of the usual PH 1 positive control (L6) and absence of a product for the negative control (L8) proved that the PCR was working properly.

The 728 bp (SmCCMF and SmCCR2) and 601 bp (SmCCMR and SmCCF2) DNA was gel purified and sent as a probe to the Department of Tropical Medicine, Tulane University, New Orleans for the screening of an *S. mansoni* Bacterial Artificial Chromosome (BAC) library.

3.1.2: Successful Isolation of the 1.3 kb SmCC cDNA

The full-length SmCC cDNA was successfully amplified from the cDNA library using SmCCF2 or SmCCF4 forward primers with the reverse primer SmCCR5. Amplification with SmCCMF and SmCCR5 produced a 718 bp product as expected (Figure 12). However, for cloning into the pPIC9K vector it was necessary to re-amplify the 1.3 kb cDNA with SmCCF and SmCCR that contain the *SnaBI* and *AvrII* recognition sites. Using gel purified SmCC DNA

amplified with SmCCF2 and SmCCR5 as template, PCR was carried out using SmCCF and SmCCR but successful amplification did not occur probably as a result of low concentration and poor quality of the template DNA.

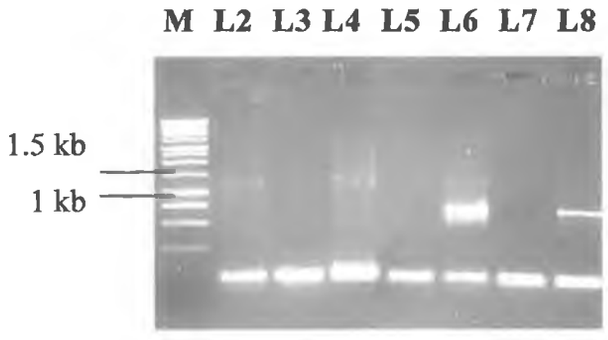


Figure 12: Successful amplification of SmCC using primers SmCCF2/F4 and SmCCR5

The 1 kb ladder (M) is clear in lane 1. The 1.3 kb band in L2 was amplified with SmCCF2 and SmCCR5 using a 1:10 dilution of cDNA template. The same reaction was carried out for the L3 sample except neat cDNA was used but no band is visible in L3. In L4 the 1.3 kb product of the PCR using SmCCF4 and SmCCR5 and a 1:10 dilution of cDNA template can be seen clearly but when neat cDNA was used as template (L5) no product was seen on the gel. In L6 a 718 bp product was successfully amplified using SmCCMF and SmCCR5 primers. The negative and positive PCR controls in L7 and L8 respectively showed that the PCR worked effectively. The PCR positive control was a 1:10 dilution of cDNA amplified with SmCCMF and SmCCR2 whereas the negative control had no template but contained all other PCR components.

It was thought that a problem existed with the 3' end of the SmCC cDNA making amplification of the full-length product difficult. PCR was undertaken to illustrate this problem.

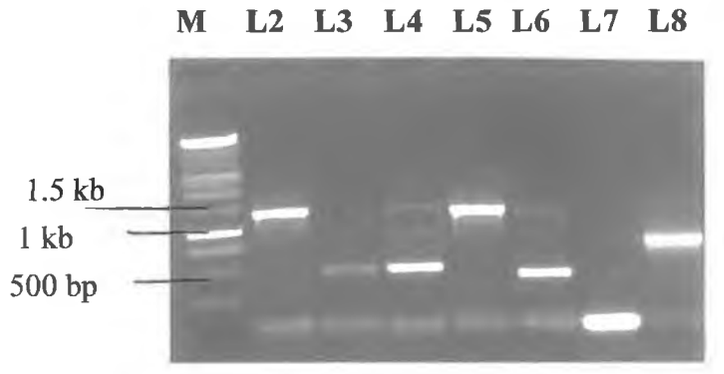


Figure 13a: PCR to prove that a problem existed with the 3' end of SmCC

A strong 1.35 kb band was amplified with SmCCF4 and SmCCR5 (L2). The usual 480 bp band appears with SmCCF4 and SmCCR2 (L3). In L4 a faint band at 1.36 kb was seen but a very bright band was also seen of 480 bp. In L5 a strong 1.3 kb band is visible with SmCCF2 and SmCCR5. The usual very bright 450 bp band and faint 1.3 kb band seen after PCR with SmCCF2 and SmCCR2 appears again in L6. PCR with SmCCF2 and SmCCR3 yielded no product (L7). In L8 the very bright 718 bp band seen was obtained with SmCCMF and SmCCR5.

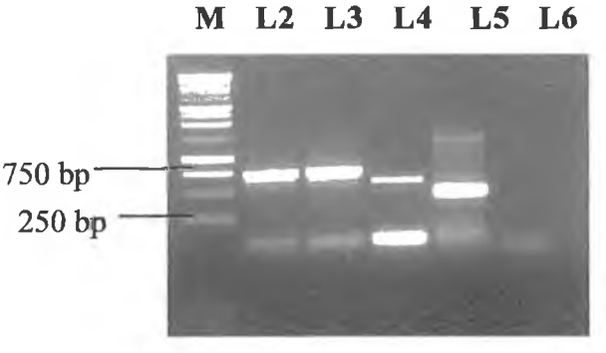


Figure 13b: PCR to prove that a problem existed with the 3' end of SmCC

PCR with SmCCMF and SmCCR2 produced the usual 728 bp product (L2). The 731 bp band seen in L3 resulted from PCR with SmCCMF and SmCCR3. The 601 bp band in L4 was the product of PCR with SmCCMR and SmCCF2. L5 depicts the very strong 450 bp band and

faint 1.3 kb band amplified with SmCCF and SmCCR. L6 contains the negative control.

As SmCCR5 did not extend beyond the stop site of SmCC, another primer SmCCR6 was designed which did (Figure 7).

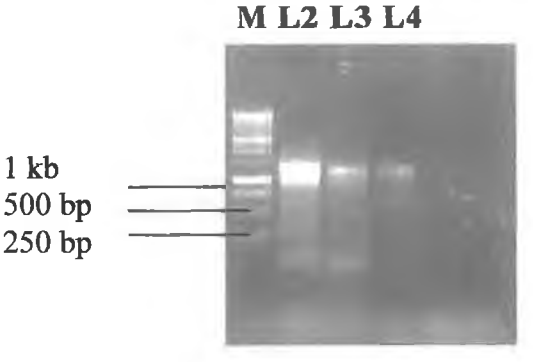


Figure 14: SmCC full-length cDNA

The usual 1 kb ladder (M) is depicted in lane 1. The 1.3 kb SmCC cDNA is visible in L2 as amplified with SmCCF2 and SmCCR6. In L3 the 1.3 kb gene is also present, this time amplified with SmCCF and SmCCR6. The full-length SmCC cDNA can also be amplified with SmCCF4 and SmCCR6 (L4).

The strong 1.3 kb band obtained as a result of PCR with SmCCF2 and SmCCR6 was gel purified and re-amplified by PCR using primers SmCCF and SmCCR, which contain the restriction sites for cloning into pPIC9K (refer to section 3.2).

However, after many unsuccessful attempts at amplification of the SmCC cDNA containing the necessary restriction sites it decided to clone the purified 1.3 kb cDNA amplified using SmCCF2 and SmCCR6 into the pGEM-T vector.

3.1.3: Cloning of 1.3 kb SmCC cDNA into pGEM

Gel purified and concentrated 1.3 kb SmCC was cloned into the pGEM-T vector system (refer to 2.2.2).

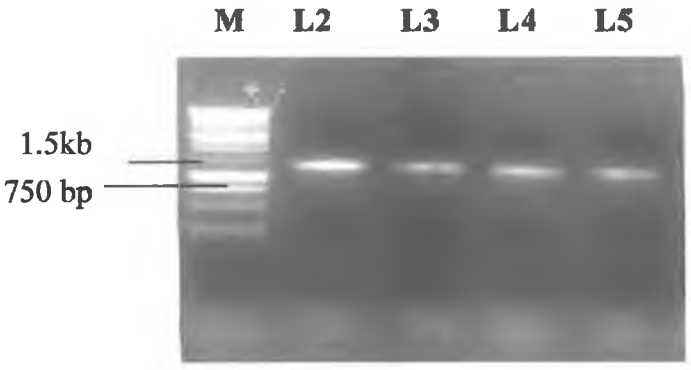


Figure 15: Gel purified 1.3 kb SmCC as seen on a 1% agarose gel

L1 contains the 1 kb ladder (M). In L2-L5 the gel purified 1.3 kb SmCC amplified with SmCCF2 and SmCCR6 is clearly visible. All four purified DNA samples were pooled and concentrated by ethanol precipitation prior to ligation to the pGEM vector.

The concentration of 1.3 kb DNA used in the ligation procedure was 395 ng/μL.

Table 30: Ligation of 1.3 kb SmCC cDNA to pGEM

Component	A	B	C	Positive	Background
2X Buffer	5 μL	5 μL-	5 μL	5 μL	5 μL
Vector	3 μL	2 μL	1 μL	1 μL	1 μL
1.3 kb gene	1 μL	3 μL	3 μL	——	———
Control insert	———	———	——	2 μL	———
Ligase	1 μL	1 μL	1 μL	1 μL	1 μL
H_20	--------	1 μL	--------	1 μL	3 μL

Overnight ligation and transformation was carried out (refer to Materials and Methods, Section 2.2.2).

Table 31: Results of transformation process

Plates	A	B	C
White Colonies	A1: 1 A2: 1	B1: 0 B2: 0	C1: 10 C2: 9

White colonies were selected off plate C only and were cultured overnight in LB-Amp broth. Plasmids were subsequently purified by alkaline lysis (refer to 2.2.2). The success of the purification procedure was determined by visualisation of the purified plasmids on a 1% agarose gel and was deemed to be successful.

Purified plasmid DNA was used as a template in a PCR reaction employing SmCC specific primers and restriction digest analysis with *PstI* proved that the SmCC cDNA was contained in pGEM.

The plasmid DNA served as a template in the PCR reaction using SmCCF and SmCCR primers to amplify the SmCC cDNA with *SnaBI* and *AvrII* restriction sites.

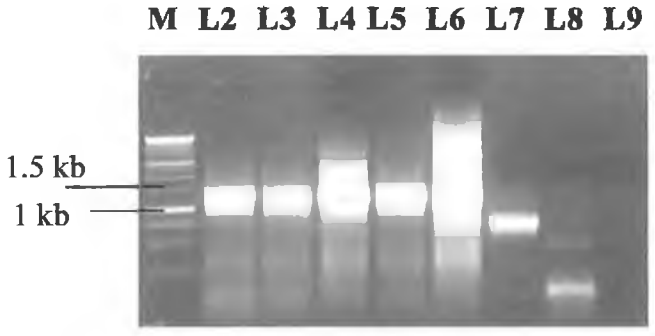


Figure 16: PCR amplification of SmCC containing restriction sites for cloning

The marker in lane 1 is the 1 kb DNA ladder. In L2-L6 purified plasmid served as the template for the re-amplification of SmCC with primers SmCCF and SmCCR. L7 contains the PCR positive control, which is cDNA template amplified with SmCCMF and SmCCR2 and the bright band seen at 728 bp proves the PCR to be a success. In L8 pGEM vector was used as template for PCR with SmCCF and SmCCR. No band should be visible but faint band at 450 bp and stronger band at 150 bp indicate that non-specific binding of primers to the pGEM vector occurred. The absence of DNA in L9, the negative control shows that the PCR reaction components were not contaminated.

3.1.4: Cloning, Sequencing and Alignment of the 450 bp PCR product

To determine what the 450 bp fragment was that was repeatedly amplified with the SmCC specific primers as described earlier, it was decided to sequence this fragment. The 450 bp product was amplified using SmCCF2 and SmCCR2 forward and reverse primers. A gel purified, concentrated 450 bp PCR product was cloned into the pGEM-Easy vector system. PCR analysis of the purified plasmids using the pGEM specific primers pGEM F and pUC/M13 R demonstrated that the 450 bp product was contained in the pGEM vector (Figure 17).

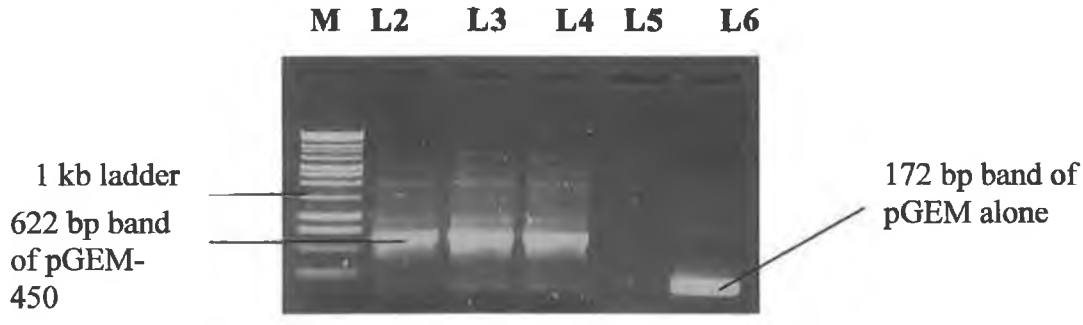


Figure 17: Bands showing the 450 bp band in the pGEM vector

The 1 kb ladder (M) is in lane 1. PCR amplification of the pGEM-450 construct with pGEM F and pUC/M13 R results in a 622 bp band (L2-L4) being produced. PCR of pGEM vector alone gives a band at 172 bp (L6). The negative control, which contains all reaction components except template DNA did not produce any band proving that no contaminants were present (L5). Non-specific bands are also visible at 1 kb and 2 kb in L2-L4.

Sequencing was carried out in the 3'→5' direction using the pUC/M13 reverse primer (Figure 18).

1 **pGEM forward primer**

GGGCGAATTGGGCCCGACGT**CGCATGCTCCCGGCCGCCAT**GGCGGCC

60

GCGGGAATTCGAT*ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGAC

CATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTAT

PUC/M13 reverse primer

AGTGTCACCTAAATAGCTTGGCGTAATCATG**GTCATAGCTGTTTCCTG**

TGTGAAAT

200

* denotes the point where the pGEM vector was linearised with EcoR V and where a T-residue was added to both 3' ends for ligation of the PCR product to the vector

Figure 18: Location of pGEM specific primers on pGEM-T Easy vector

pGEM-450 amino acid sequence aligned with 99.2% identity (129/130 amino acids) with the first 130 amino acids of Sm procathepsin C. However, the last 9

amino acids of pGEM-450 namely LGVRFDPVL had 100% identity with the last 9 amino acids of SmCC. This indicated that the middle of the SmCC cDNA was missing from the 450 bp product, giving rise to a shorter truncated SmCC.

```
                10        20        30        40        50
                 |         |         |         |         |
SmCC  ADTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHWTL
450   ADTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHWTL
Cns.  ADTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHWTL

    60        70        80        90       100       110
     |         |         |         |         |         |
SmCC IYNQGFEVAINHRKWLVIFAYKSNGEFNCHKSMPMWTHDTLIRQWKCFVAEKIGVHDKFH
450  IYNQGFEVTINHRKWLVIFAYKSNGEFNCHKSMPMWTHDTLIRQWKCFVAEKIGVHDKFH
Cns. IYNQGFEVaINHRKWLVIFAYKSNGEFNCHKSMPMWTHDTLIRQWKCFVAEKIGVHDKFH

         120       130
          |         |
SmCC      INKLFGSKSFGRTLYHIN
pGEM450   INKLFGSKSLGVRFDPVL--
Cns.      INKLFGSKSlGrrldh!n
```

Figure 19: Alignment of SmCC with pGEM-450

Note: The first amino acid (A) in figure 18 is actually the 20th amino acid of preprocathepsin C because a 19 amino acid signal sequence precedes procathepsin C. The consensus sequence is also shown (cns).

3.2: Cloning of the SmCC cDNA into pPIC9K vector and sequence analysis

3.2.1: Cloning and verification of the presence of SmCC in pPIC9K

The SmCC cDNA and pPIC9K were successfully linearised with *SnaBI* and *AvrII*. An overnight ligation was then performed, using a 3:1 molar ratio of vector:insert. Transformation of *E. coli* with the vector/SmCC ensued the next day. Initially, transformation efficiency proved to be poor and PCR using purified plasmid DNA from an overnight bacterial culture of the selected transformant with the pPIC9K specific primers AOX3' and α-factor produced only the 195 bp band characteristic of pPIC9K that does not contain an inserted gene (Figure 20). The presence of the insert in the vector would be indicated by a band at 1.5 kb (195 bp of pPIC9K + 1.3 kb of SmCC). A *PstI* digest was also performed to verify the presence of SmCC in the vector. The expected fragment sizes produced when vector alone and vector containing the SmCC cDNA were cut with *PstI* are depicted in Table 32.

On one occasion, only 75 ng/μL of SmCC and 20 ng/μL of vector were available. In this instance the usual 3:1 vector:insert ratio was not maintained in the subsequent ligation. 2μL of SmCC and 10μL of vector were used. Transformation was carried out as usual. On analysis of the purified plasmid DNA by PCR using AOX3' and α-factor primers, the 195 bp band of pPIC9K appeared on the gel together with a faint 1.5 kb band for two out of three of the plasmids. This could mean that contamination of the purified plasmids with pPIC9K occurred, or that a mixed colony was used which had plasmids that contained SmCC but also those that did not (Figure 22). A double digest was performed on the samples using *SnaBI* and *AvrII* to liberate the 1.3 kb SmCC cDNA from pPIC9K. However, the 1.3 kb band was not seen, only the

characteristic 9.3 kb pPIC9K was visible on the gel of this digest which showed that SmCC was not contained in the vector.

Figure 20: Screening of eight transformants by PCR using AOX3' and α-factor primers

Lane 1 contains the 1 kb DNA ladder (M). In L2-L9 the eight transformants all produce the 195 bp band expected for pPIC9K without an insert. As a control pPIC9K vector was also amplified with the primers and the expected 195 bp band can be seen clearly (L10). A negative control was also included in the PCR that contained all components of the reaction except template DNA. No band was expected but the amplification of a 195 bp band in L11 is suggestive of pPIC9K contamination of the PCR.

Typically, prior to ligation, digested vector and plasmid DNA was cleaned up using a Wizard kit (Promega). However the transformation process was seen to be more efficient when DNA that had not been cleaned-up was used in the ligation process. After overnight ligation was performed ligation products were run on a gel (Figure 23). Ligation was deemed unsuccessful for those samples in which a distinct 1.3 kb band was visible on gel analysis as this indicated that the SmCC cDNA had not been ligated to the vector. Only samples that did not show the distinct 1.3 kb SmCC band on the gel were selected for the transformation process. Transformation efficiencies were high in this instance with hundreds of colonies present on plates (plates B1 and B2). Nevertheless,

screening with AOX3' and α-factor primers amplified only the 195 bp band of pPIC9K.

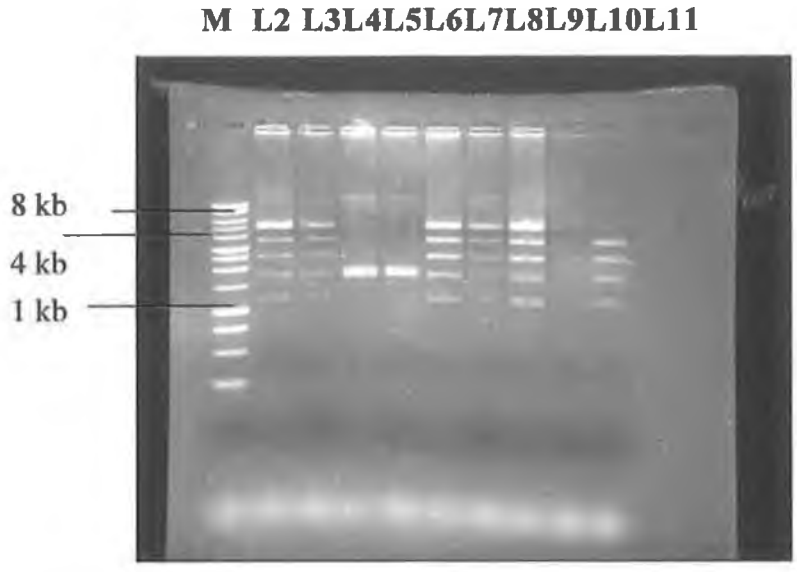


Figure 21: *PstI* digest of eight pPIC9K clones expected to contain SmCC

The usual 1 kb ladder is visible in lane 1 (M). Lane 2-9 contains the eight purified plasmids isolated from the selected transformants. The expected bands at 1.241 kb, 1.827 kb, 2.55 kb and 3.6 kb can be seen for all samples except 3 and 4 (L4 and L5). Also a very strong band at 5 kb is present for all samples except 3 and 4. If the plasmids prove positive for the presence of SmCC it is expected that a band of 4.9 kb would be produced but the 3.6 kb band should be absent. In this case both the 4.9 kb and 3.6 kb bands are present, which provides an inconclusive result. A control was included (L11) which was pPIC9K vector alone cut with *PstI* and bands produced are of the expected sizes (refer to Table 32).

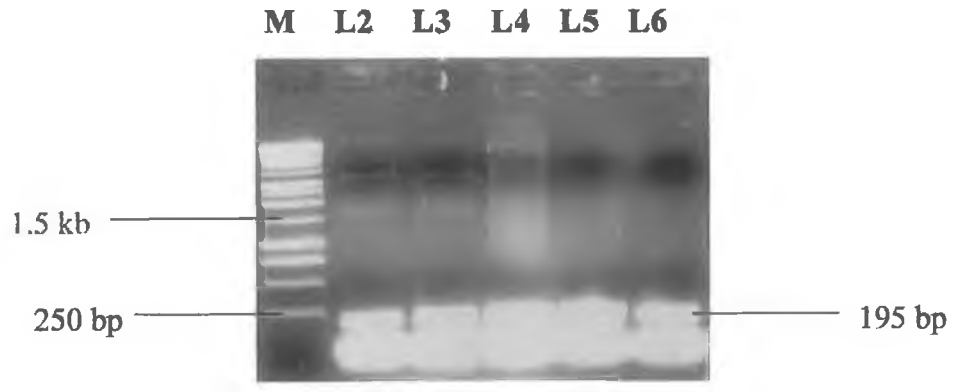


Figure 22: PCR amplification of three pPIC9K clones expected to contain SmCC using AOX3' and α-factor primers

The 1 kb DNA ladder is depicted in L1 (M). In L2-L4 the 195 bp band of pPIC9K is clear. However, a band is also visible ~1.5 kb which would indicate that pPIC9K does in fact contain the 1.3 kb SmCC cDNA. The control in L5 is pPIC9K vector amplified with AOX3' and α-factor primers. The negative control in L6 has a distinct 195 bp band, which is indicative of pPIC9K contamination of the PCR as no template DNA was added in this reaction.

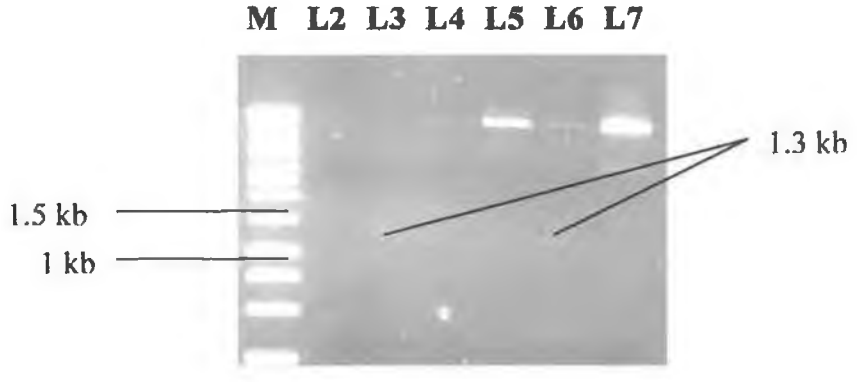


Figure 23: Gel of overnight ligation products

L1 contains the 1 kb DNA ladder (M). In all lanes 9 μL of ligation mix was used. In L2-L6 pPIC9K 9.3 kb DNA is clearly visible but for samples 2 and 5 (L3 and L6) the 1.3 kb SmCC cDNA can also be seen

which illustrates that ligation was not successful for these samples. L7 contains pPIC9K as a control.

A series of digests were performed to establish whether the insert was present in the vector (Table 32).

Table 32: Restriction digests performed to verify that pPIC9K contained the SmCC cDNA

Restriction enzyme	**Fragments produced (pPIC9K)**	**Fragments produced (pPIC9K and SmCC)**	**Result**
KpnI	612 bp 8,683 bp	612 bp 9,983 bp	Digestion incomplete
PstI	1,241 bp 1,827 bp 2,555 bp 3,677 bp	1,241 bp 1,827 bp 2,555 bp 4,977 bp	Inconclusive
StuI	1,425 bp, 7,870 bp	1,425 bp 9,170 bp	Digestion incomplete
XbaI	2,034 bp 7,261 bp	1,935 bp 2,583 bp 6,712 bp 6,729 bp	As above
XhoI	1,195 bp 3,583 bp 4,517 bp	1,195 bp 3,583 bp 4,517 bp	As above

Double digests were also performed using *SnaBI* and *AvrII* to liberate the SmCC cDNA from pPIC9K if present. Digestion proved to be incomplete after numerous attempts.

Colonies were screened from the highly efficient transformation (plate B2). Plasmids were isolated from an overnight culture of the selected transformant by the methods detailed in 2.2.2. Colonies used for screening were replica plated to enable further analysis of the same colonies. A new screening method was implemented using SmCC specific primers in conjunction with pPIC9K specific ones.

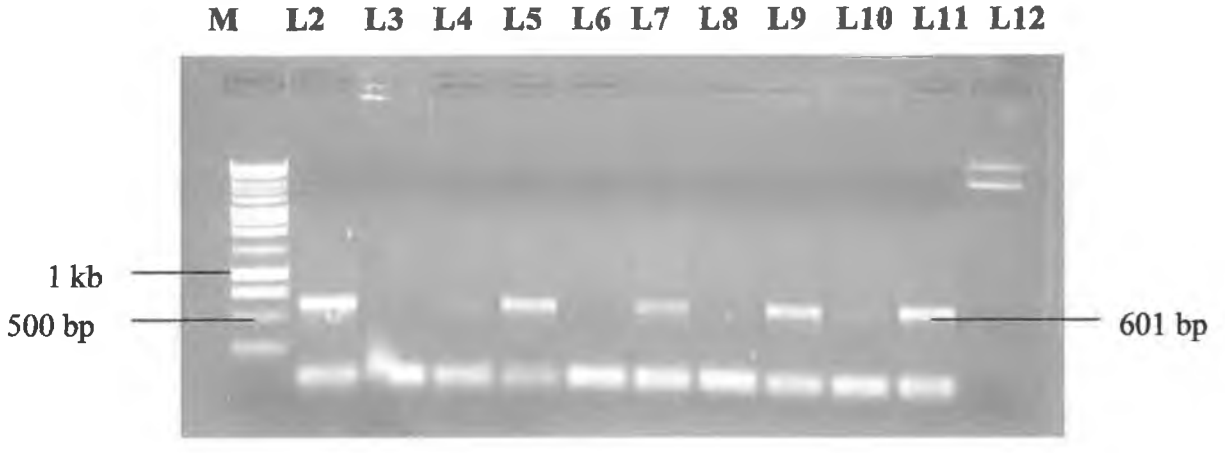


Figure 24a: PCR using SmCCF2 and SmCCMR primers for the detection of SmCC in pPIC9K in ten selected transformants from plate B2

L1 contains the 1 kb DNA ladder (M). 10 μL of each PCR product and 2 μL of loading dye was loaded into each well. A strong 601 bp band is visible in L2, L5, L7, L9 and L11 for clones 1, 4, 6, 8 and 10. This band is faint in L4 and L10. L12 contains 5μL of pPIC9K and 2 μL of loading dye

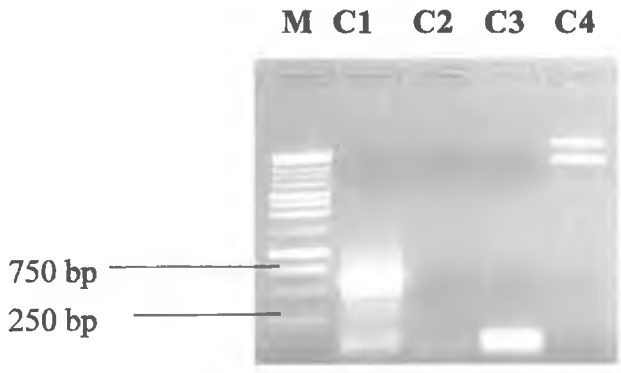


Figure 24 b: Controls for PCR (24a)

C1, which was purified SmCC template amplified with SmCCF2 and SmCCMR produced the expected 601 bp band. No band was visible for C2 as expected because only one primer was used in this PCR with SmCC purified template. A low band present for C3 could be as a result of primer-dimer formation. The template for C3 PCR was pPIC9K and SmCCF2 and SmCCMR primers were used. No band is expected and none is seen. The last lane C4 contains 5μL of pPIC9K and 2 μL of loading dye.

NotI was used to linearise plasmid DNA from 6 clones (clones 1, 4, 6, 8, 9 and 10), which appeared positive when screened with SmCCMF2 and SmCCMR (Figure 24a). All clones were originally from transformant plate B2. The linearised plasmids were then compared by size on a gel to both digested and undigested pPIC9K vector as an increase in size of the vector would indicate that it contained SmCC.

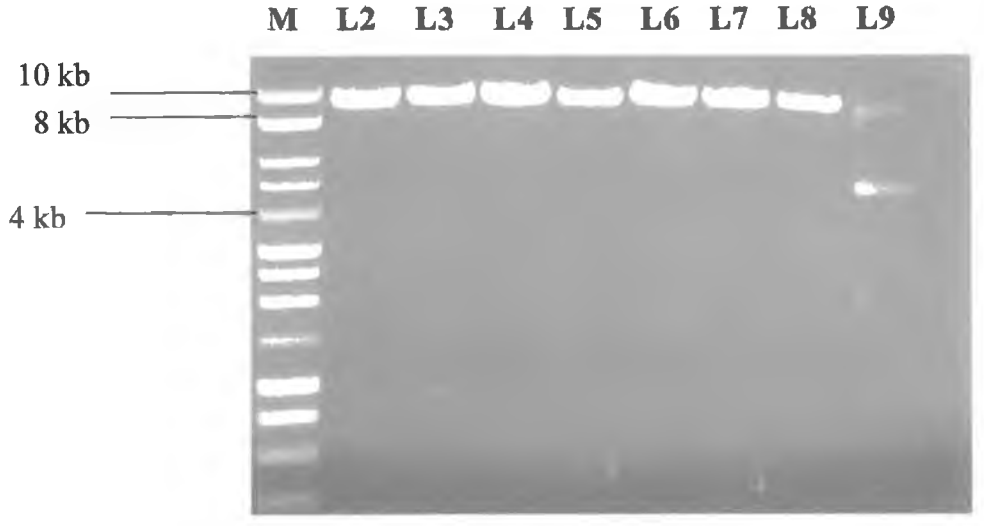


Figure 25: *NotI* digest to linearise plasmid DNA

L1 (M) contains the 1 kb ladder. Each lane contains 20 µL of plasmid DNA and 5 µL of loading dye. The linearised plasmid DNA in L2-L8 appears to run slightly higher than digested pPIC9K. However, this could be because the DNA in L2-L8 is more concentrated than that in L9.

Another screening strategy employed involved the use of the AOX3' primer with SmCCMF and the α-factor primer with SmCCMR to screen for plasmids which contained SmCC (Table 33). The same isolated plasmids were screened as in the PCR in Figure 24a.

Table 33: Sizes expected for screening with SmCC and pPIC9K primers

Primer sets	Product size
α-factor SmCCMR	664 bp
AOX3' SmCCMF	814 bp

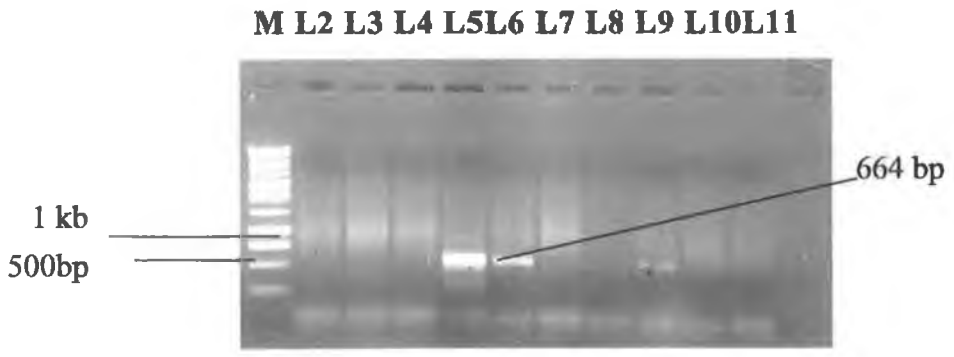


Figure 26: Screening of transformants for SmCC using α-factor and SmCCMR

The 1 kb ladder is depicted in lane 1(M). The 664 bp band indicative of a positive result for SmCC is seen in L5, L6 and L9 for clones 4, 5 and 8 respectively. Only clones 4 and 8 were positive for SmCC when screening with SmCCF2 and SmCCMR primers.

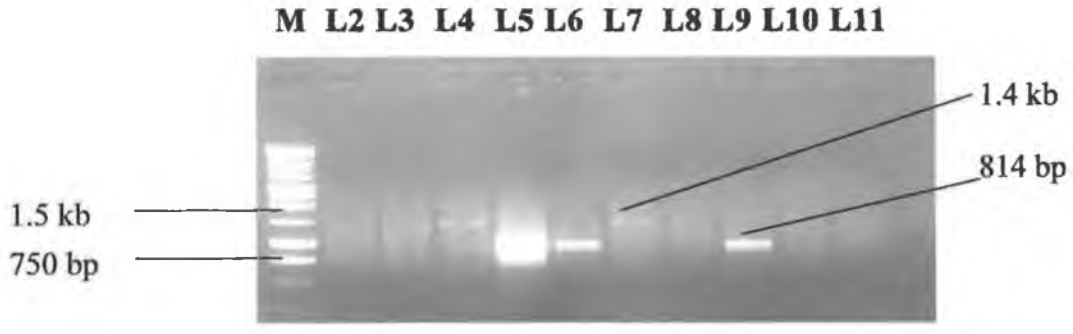


Figure 27: Screening of plasmids for SmCC using AOX3' and SmCCMF

The 1 kb ladder is clearly visible in lane 1(M). The band at 814 bp, which is indicative of a positive result for SmCC is seen in L5, L6 and L9 for clones 4, 5 and 8. For those clones which do not appear to be positive for SmCC (clones 1, 2, 3, 6, 7, 9 and 10) a distinct band at 1.4 kb is visible.

Purified plasmid DNA was sent for sequencing. The resulting sequence data proved that the selected clones did not contain the SmCC cDNA (from transformation plate B2).

For further analysis transformation plate B1 was used. A new SmCC specific primer, SmCCMF2 was designed, which when used along with SmCCMR would amplify a 200 bp SmCC band (Figure 28). PCR screening with α-factor and SmCCMR, and AOX3' and SmCCMF revealed that 4 out of 8 of the selected colonies, namely transformants 2, 4, 6 and 8 appeared positive for SmCC.

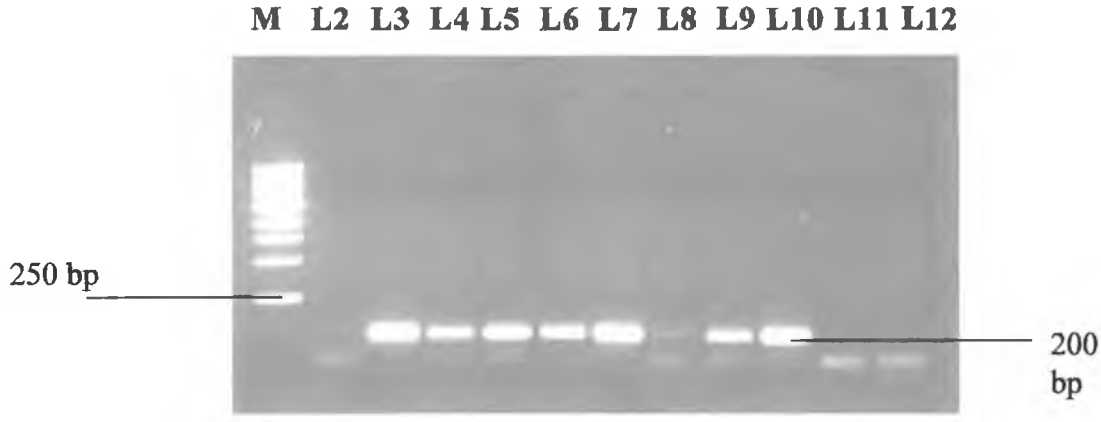


Figure 28: Screening of clones using primers SmCCMF2 and SmCCMR

The usual 1 kb ladder is depicted in lane 1 (M). The 200 bp band expected if clones were positive for SmCC can be seen clearly in 6 of the 8 clones (L3-L7 and L9) and is faintly visible in the other two (L2 and L8). SmCC DNA was also amplified with the primers and the resultant 200 bp band is visible in L10. PCR using the primers and pPIC9K template did not produce a 200 bp product as expected (L11) and the absence of a band in L12, the negative control, shows that no contaminants were present in the PCR.

Plasmids that appeared positive for SmCC on PCR screening with SmCCMF2 and SmCCMR (clone 2, 4 and 6) were used to transform E. coli cells. The transformation proved to be very efficient with ~200 colonies on each plate. One plate was selected to work with, the plate containing clone 2. However, although the previous PCR screenings revealed this clone to be positive for SmCC, subsequent screening failed to replicate this result. More colonies were selected off plate 2 and were screened. Further screening with SmCCMF2 and SmCCMR produced a positive result for all clones with a strong 200 bp PCR product visible on the gel. Screening with AOX3' and SmCCMF produced a 1.4 kb band for all clones, which in previous PCRs was associated with a negative result (Figure 27). Plates 4 and 6 were also screened by PCR for SmCC but results proved inconclusive.

Restriction digests were performed on all potentially positive clones using *SnaBI* and *AvrII* to liberate the 1.3 kb SmCC cDNA from the pPIC9K vector. However, for all clones analysed no 1.3 kb was seen when digested products were run on a gel which indicated that plasmids did not contain SmCC.

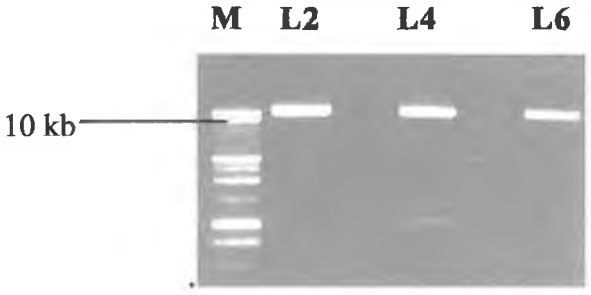


Figure 29: Double digest of plasmids using *SnaBI* and *AvrII*

Lane 1 contains the 1 kb ladder (M). For all lanes 20 μL of digested plasmid was loaded. L2 contains the digested plasmid expected to contain SmCC. L4 contains digested pPIC9K known to contain *F.*

hepatica cathepsin L1 (FhCL1). Digested pPIC9K without an insert is depicted in lane 6. The band in L2 is of a similar size to that in L6 so is pPIC9K DNA. The absence of a 1.3 kb band proves that SmCC is not present in the vector. In L4 the high band of pPIC9K is present but also a 1 kb band is visible which illustrates that pPIC9K contained FhCL1, which was released from the vector on digestion.

Purified SmCC cDNA with *SnaBI* and *AvrII* restriction sites was digested as before prior cloning into pPIC9K. The DNA was cleaned-up after digestion was complete. The concentration of DNA was 105 ng/μL and 115 ng/μL for vector and SmCC respectively. Two overnight ligations were performed using a 3:1 (ligation A) and 2:1 molar ratio of vector:insert (ligation B). The results of the subsequent transformation are outlined in Table 34.

Table 34: Results of transformation

Transformation	No. of colonies
A	A1: 20 A2: 23
B	B1: 25 B2: 42
E. coli control	No growth
pPIC9K +ve control	TNTC

Plate A1 only was selected for further analysis for simplicity.

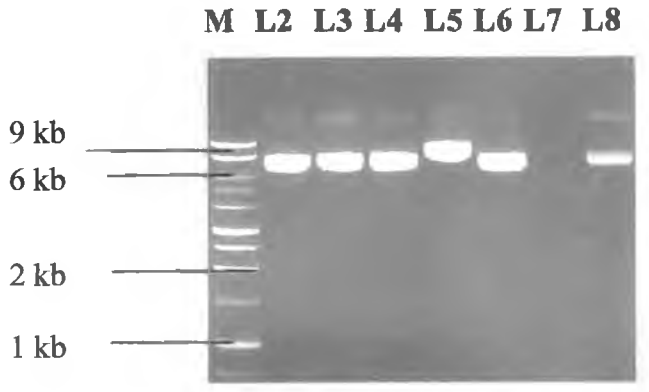


Figure 30: Gel of purified plasmid DNA (plate A1)

5 μL of plasmid and 3 μL of loaded dye were applied to each lane. All plasmids with the exception of plasmid # 4 (L5) appear to be the same size as the pPIC9K vector without insert (L8). Plasmid 4 appears to be larger than pPIC9K, which indicates that the SmCC cDNA is possibly contained in this plasmid.

PCR using SmCCF2 and SmCCMR (Figure 31) and α-factor and SmCCR primers (Figure 32) on all plamids (1-5) showed that SmCC was present in the vector.

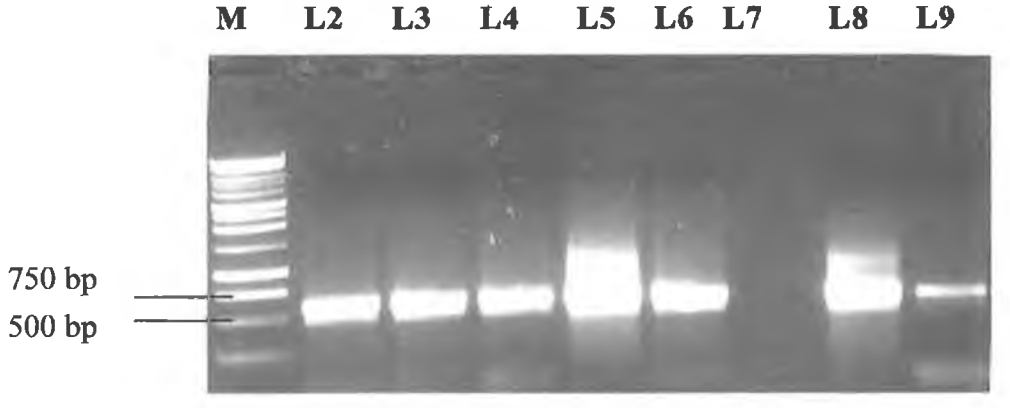


Figure 31: PCR amplification of SmCC specific band using SmCCF2 and SmCCMR

The SmCC 601 bp band can be clearly seen for all plasmids (L2-L6), which indicates that pPIC9K contains the SmCC cDNA. The positive control in L8 is a pure SmCC template amplified with SmCCF2 and SmCCMR. When pPIC9K was amplified with these primers (L9) a band at 601 bp can also be seen which indicates that contamination occurred in the PCR. However, this band does not appear as strong as the other 601 bp bands visible on the gel.

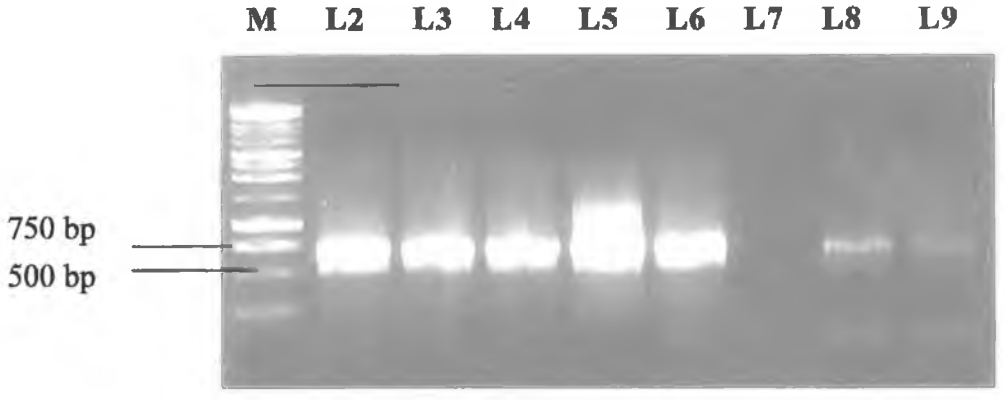


Figure 32: PCR amplification of SmCC specific band using α-factor and SmCCMR primers

The DNA of the 1 kb ladder appears to have degraded somewhat although the bands are still distinct. The band of 664 bp that is expected if SmCC is contained in the vector is visible for all plasmids (L2-L6). Pure SmCC template amplified with α-factor and SmCCMR produced a faint band at 664 bp, which should not be the case (L8) and when pPIC9K served as a template this faint band can also be seen (L9).

A double digest with *SnaBI* and *AvrII* was performed to prove conclusively whether plasmid # 4 contained SmCC (Figure 33). Plasmid 4 was chosen because of all the plasmids it looked most likely to contain SmCC on gel analysis (Figure 30).

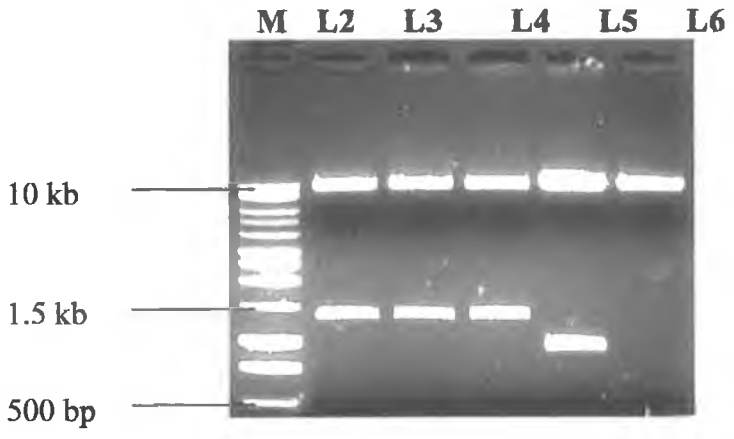


Figure 33: Double digest with *SnaBI* and *AvrII*

The 1 kb ladder is depicted in lane 1(M). 25 μL of three samples of plasmid # 4 (4A, 4B and 4C) were loaded into each lane (L2-L4). In each of these lanes the 1.3 kb band is clearly visible having been released from the 9.3 kb pPIC9K plasmid. In L5 pPIC9K known to contain FhCL1 was digested and the 1 kb FhCL1 band was successfully Liberated from the plasmid thus proving the success of the digestion process. The pPIC9K vector was also digested as a control (L6).

PCR was carried out using plasmid # 4 (4A, 4B and 4C) and AOX3' and α-factor primers as an additional verification of the presence of SmCC in the vector (Figure 34).

Using the plasmid # 4, E. coli cells were transformed and plasmids were isolated and confirmed to contain SmCC. This was done to provide a plentiful supply of the plasmid DNA for sequencing.

Figure 34a: Diagrammatic representation of SmCC/pPIC9K construct.

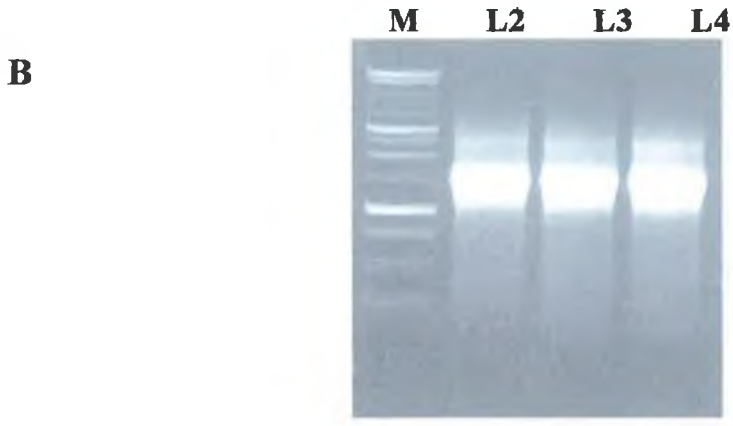


Figure 34 b: PCR with AOX3' and α-factor primers to show that pPIC9K contained the SmCC cDNA

The 1.5 kb bands in L2-L4 (4A, 4B and 4C) show conclusively that SmCC is present in pPIC9K. Controls (not shown) were also included in the PCR one of which was a pPIC9K template amplified with AOX3' and α-factor primers producing the characteristic 195 bp band for pPIC9K. The other contained all PCR components except template DNA showed that contaminants were not present in the PCR.

3.2.2: Sequencing of SmCC in pPIC9K

Sequencing was carried out on purified plasmid DNA # 4 to verify that it contained SmCC and also to confirm that changes had not occurred in the cDNA sequence, particularly in the active site. Sequencing was carried out in the 3'→ 5' direction using the following primers depicted in colour in Figure 35:

PRO REGION

Sna BI **MR4** **MR4**

GYVADTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPD
IA

IDEFGNRGHWTLIYNQGFEVTINHRKWLVIFAYKSNGEFNCHKSMPM
W

MR

THDTLIRQWKCFVAEKIGVHDKFHINKLFGSKSFGRTLYHINPSFVDKI
N

AHQKSWRAEIYPELSKYTIDELRNRAGGVKSMVTRPSVLNRKTPSKEL

MR **MATURE REGION**

ISLTGNLPLEFDWTSPPDGSRSPVTPIRNQGICGSCYAFASAAALEARIRLV
SNFS

MR5

EQPILSPQAVVDCSPYSEGCNGGFPFLIAGKYGEDFGFVSENCDPYTGEDT
GK

MR5

CTVSKNCTRYYTADYSYIGGYYGATNEKLMQLELISNGPFPVGFEVYEDF
QF

AOX3'

YKEGIYHHTTVQNDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVK
NS

AvrII

WGVEWGEQGYFRILRGTDECGVESLGVRFD PVLHHHHHH STOP PRAAN

Figure 35: Location of the primers used to sequence the SmCC cDNA

- **Pink:** AOX 3' primer, underlined pink region denotes overlap in the sequence obtained using AOX3' with that using SmCCMR5

- **Green:** SmCCMR5 primer, underlined green region denotes overlap in the sequence obtained using MR5 with SmCCMR primer

- **Red:** SmCCMR primer, underlined red region denotes overlap in the sequence using this primer with sequence obtained using MR4 primer

- **Blue:** SmCCMR4 primer which sequenced some pPIC9K amino acids (not shown) in addition to SmCC amino acids

The yellow highlighted regions indicate the primer binding sites on the cDNA.

DTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFG
NRGHWTLIYNQGFEVTINHRKWLVIFAYKSNGEFNCHKSMPMWTHDTLIR
QWKCFVAEKIGVHDKFHINKLFGSKSFGRTLYHINPSFVDKINAHQKSWR
AEIYPELSKYTIDELRNRAGGVKSMVTRPSVLNRKTPSKELISLTGNLPLEF
DWTSPPDGSRSPVTPIRNQGICGSCYAFASAAALEARIRLVSNFSEQPILSPQ
AVVDCSPYSEGCNGGFPFLIAGKYGEDFGFVSENCDPYTGEDTGKCTVSK
NCTRYYTADYSYIGGYYGATNEKLMQLELISNGPFPVGFEVYEDFQFYKE
GIYHHTTVQNDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVKNSWG
VEWGEQGYFRILRGTDECGVESLGVRFDPVL

Figure 36: The full amino acid sequence of the SmCC cDNA

The 198 amino acid propeptide is shown in blue and the 237 mature region is shown in black. The active site residues Cys 30, His 181 and Asn 203 are shown in red.

3.2.3: Alignment of SmCC cDNA with the Genbank sequence and other cathepsin Cs

An alignment was performed aligning the deduced amino acid sequence of SmCC from the Genbank (Butler et al., 1995) with the SmCC sequence detailed in Figure 36.

```
              10         20         30         40         50         60
               |          |          |          |          |          |
SmCC ADTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHWTLI
GB   ADTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHWTLI
Cns. ADTPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHWTLI

              70         80         90        100        110        120
               |          |          |          |          |          |
SmCC YNQGFEVTINHRKWLVIFAYKSNGEFNCHKSMPMWTHDTLIRQWKCFVAEKIGVHDKFHI
GB   YNQGFEVTINHRKWLVIFAYKSNGEFNCHKSMPMWTHDTLIDSGSV-CSGKIGVHDKFHI
Cns. YNQGFEVTINHRKWLVIFAYKSNGEFNCHKSMPMWTHDTLIrqgkc caeKIGVHDKFHI

              130        140        150        160        170        180
               |          |          |          |          |          |
SmCC NKLFGSKSFGRTLYHINPSFVDKINAHQKSWRAEIYPELSKYTIDELRNRAGGVKSMVTR
GB   NKLFGSKSFGRTLYHINPSFVGKINAHQKSWRGEIYPELSKYTIDELRNRAGGVKSMVTR
Cns  NKLFGSKSFGRTLYHINPSFVdKINAHQKSWRaEIYPELSKYTIDELRNRAGGVKSMVTR

              190        200        210        220        230        240
               |          |          |          |          |          |
SmCC PSVLNRKTPSKELISLTGNLPLEFDWTSPPDGSRSPVTPIRNQGICGSCYAFASAAALEA
GB   PSVLNRKTPSKELISLTGNLPLEFDWTSPPDGSRSPVTPIRNQGICGSCYASPSAAALEA
Cns. PSVLNRKTPSKELISLTGNLPLEFDWTSPPDGSRSPVTPIRNQGICGSCYAfaSAAALEA

              250        260        270        280        290        300
               |          |          |          |          |          |
SmCC RIRLVSNFSEQPILSPQAVVDCSPYSEGCNGGFPFLIAGKYGEDFGFVSENCDPYTGEDT
GB   RIRLVSNFSEQPILSPQTVVDCSPYSEGCNGGFPFLIAGKYGEDFGLPQKIVIPYTGEDT
Cns  RIRLVSNFSEQPILSPQaVVDCSPYSEGCNGGFPFLIAGKYGEDFGlpqencdPYTGEDT

              310        320        330        340        350        360
               |          |          |          |          |          |
SmCC GKCTVSKNCTRYYTADYSYIGGYYGATNEKLMQLELISNGPFPVGFEVYEDFQFYKEGIY
GB   GKCTVSKNCTRYYTTDYSYIGGYYGATNEKLMQLELISNGPFPVGFEVYEDFQFYKEGIY
Cns. GKCTVSKNCTRYYTaDYSYIGGYYGATNEKLMQLELISNGPFPVGFEVYEDFQFYKEGIY

              370        380        390        400        410        420
               |          |          |          |          |          |
SmCC HHTTVQNDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVKNSWGVEWGEQGYFRILRGT
GB   HHTTVQTDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVKNSWGVEWGEQGYFRILRGT
Cns. HHTTVQnDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVKNSWGVEWGEQGYFRILRGT

              430
               |
SmCC DECGVESLGVRFDPVL
GB   DECGVESLGVRFDPVL
Cns. DECGVESLGVRFDPVL
```

Figure 37: Amino acid alignment of SmCC with the Genbank SmCC (GB), (accession no: Z32531). The consensus sequence is also shown (Cns)

The results of the alignments performed on SmCC are discussed in detail in Section 4.3. Although SmCC cDNA was seen to be 94.5% identical to that in the Genbank some differences were apparent (Figure 37). SmCC cDNA was aligned with rat, mouse, human and S. japonicum cathepsin Cs and significant changes in amino acid sequence among the species were noted (Figure 38 and 39).

Rat CC	RNWACFVGK
Mouse CC	RNWACFVGK
Human CC	RNWACFTGK
SmCC	RQWKCFVAE
SmCC (Genbank)	DSG__SVCSG
SjCC	CQWHCFTAT

Figure 38: Amino acid differences between SmCC and Genbank SmCC proregions and the conservation of these changes among different species

Rat CC	VVEENCF
Mouse CC	VVEESCF
Human CC	LVEEACF
SmCC	FVSENCD
Genbank SmCC	LPQKIVI
SjCC	FVEEKCN

Figure 39: Amino acid differences between SmCC and Genbank SmCC mature enzymes and the conservation of these changes among different species

Amino acid identities of SmCC with the S. japonicum cathepsin C (SjCC) and mammalian cathepsin Cs are shown in Table 35.

Table 35: Amino acid identities of aligned sequences

Alignment (pro and mature regions)	% amino acid identity
SmCC and Genbank SmCC	94.5% (412/436)
SmCC and SjCC	60.78% (265/436)
SmCC and Human	47.99% (215/448)
SmCC and Rat	48.2% (216/448)
SmCC and Mouse	48.44% (217/448)

As the amino acid changes that occurred did not affect the key sites in particular the active site on SmCC it was possible to proceed with the functional expression of this fragment in *Pichia pastoris*.

3.3: FUNCTIONAL EXPRESSION OF SMCC IN *PICHIA PASTORIS*

3.3.1: TRANSFORMATION OF *PICHIA PASTORIS* WITH SMCC

The restriction enzyme *BglII* was used to linearise the pPIC9K/SmCC expression cassette prior to transformation of *P. pastoris*. *BglII* stimulates gene replacement at the AOX1 loci in *P. pastoris*. This restriction enzyme was selected because it did not cut SmCC whereas the other two enzymes that could be used, *SalI* and *SacI* cut within the SmCC cDNA.

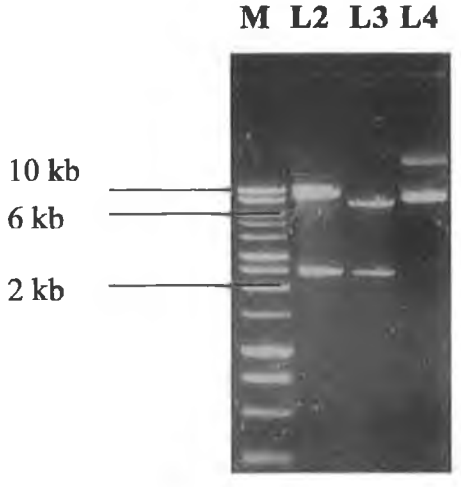


Figure 40: *BglII* restriction digest of pPIC9K/SmCC expression cassette

The 1 kb marker is depicted in lane 1 (M). The two distinct bands at 8.2 kb and 2.4 kb in L2 show that pPIC9K/SmCC digestion with *BglII* was successfully completed. As a control pPIC9K alone was digested with BglII and the expected bands at 6.8 kb and 2.4 kb are clearly visible in L3. Undigested pPIC9K was included in L4 for comparative purposes.

For the first transformation carried out, 1.8 µg of digested pPIC9K/SmCC was used to transform *P pastoris* GS115 cells. After 5 days incubation at 30^0C the confluent growth of spheroplasts on RDBH and no growth on RDB was observed. Tiny colonies appeared on the transformed *P. pastoris* plates. A cell

count revealed that 3 x 10^5 and 1 x 10^5 cells/ml were present in the top agar of each duplicate plate. As a cell count of 5 x 10^5 cells/ml is recommended for plating onto G418, each cell suspension (200 µL) was plated neat onto each G418 plate. After 6 days incubation no growth was observed on any of the plates. However, *P. pastoris* grew well on YPD, indicating that yeast cells were viable.

For the second transformation, 5 µg of pPIC9K/SmCC was used. As a control, 2.4 µg of *BglII* digested pPIC9K was used to transform *P. pastoris*.

Table 36: Results of *P. pastoris* transformation # 2

Plate	Days incubation	Cell count
A(SmCC/pPIC9K)	7	5.45 x 10^7 cells/ml
B (pPIC9K)	10	3 x 10^6 cells/ml

When plating on G418, yeast cell suspensions were diluted to 5 x 10^5 cells/ml. Neat cell suspension was also plated (plate C).

Table 37: Results of G418 plating

Plate	Days incubation	Colonies (duplicate plates)
A1	10	2
A2		11
B1	7	150
B2		147
C1	4	Confluent
C2		growth

The above results are for 0.25 mg/ml G418 plates only as no growth was observed at higher G418 concentrations. After single colony purification on YPD plates, re-plating onto 0.25 mg/ml G418 ensued. Single colonies from these plates were blotted onto MM/MD plates for phenotypic analysis. Colonies grew well on MD but not on MM and so were deemed to be of the Mut^S phenotype. Colonies on the MD plates were single colony purified on YPD prior to commencing the induction protocol.

PCR Analysis of *P. pastoris* Integrants

PCR was carried out using 5'AOX1 and 3'AOX1 primers to determine whether SmCC had been integrated into the *P. pastoris* genome. PCR of pPIC9K alone using 5'AOX1 and 3'AOX1 results in a 492 bp PCR product. Therefore, a 1.8 kb PCR product is indicative of the presence of SmCC in the *P. pastoris* genome (492 bp + 1,300 bp).

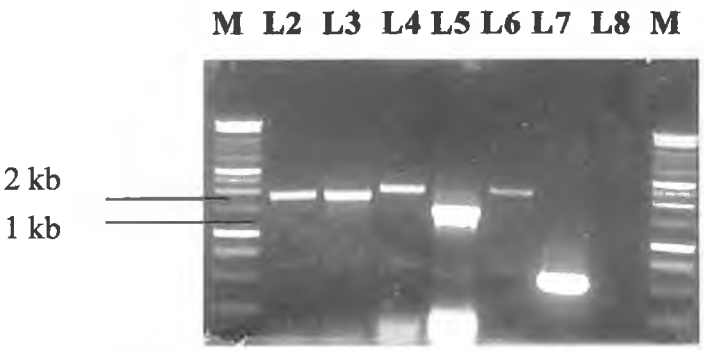


Figure 41: PCR screening of *P. pastoris* clones using 5'AOX1 and 3'AOX1 primers

The 1 kb ladder is depicted in lanes 1 and 9. In lane 2 a band is visible at 1.8 kb which illustrates that SmCC has been successfully integrated into the *P. pastoris* genome for transformants A2. The same is true for transformants from plate C2, with the appearance of a 1.8 kb band in L3.

In L4, PCR with *P. pastoris* strain GS115 results in a product at 2.1 kb as expected. *F. hepatica* CL1 (FhCL1) in *P. pastoris* was included as a control in the PCR (L5). For clone B2 (pPIC9K), only the characteristic 2.1 kb band of P. pastoris is seen as expected. (L6). In L7 the 492 bp product expected when pPIC9K is amplified with 5'AOX1 and 3'AOX1 primers is clearly visible. The absence of a band in L8, the negative control shows that no contaminants were present in the PCR.

3.3.2: METHANOL INDUCTION OF SMCC IN *P. PASTORIS* AND ANALYSIS OF EXPRESSED PROTEIN

A series of inductions were carried out, each one varying parameters of the protocol so as to optimise the expression of an active, functional SmCC enzyme.

3.3.2.1: INDUCTION # 1

Culture Conditions

100 ml of BMGY were inoculated with a single colony from the following plates:

- Plate A1 (SmCC and pPIC9K)
- Plate A2 (SmCC and pPIC9K)
- Plate B2 (pPIC9K background control)
- Plate C2 (SmCC)
- A14 clones (FhCL1 in P.pastoris)

An OD of 2.7 was reached by all cultures after 41 hours incubation. Induction began after resuspension of each culture in 20 ml of BMMY, pH 6.0. Methanol was added to culture flasks to a final volume of 1% daily for six days. Samples were taken each day, from day 0-day 6 (D0-D6).

SDS-PAGE Analysis

Culture supernatants were analysed by SDS-PAGE.

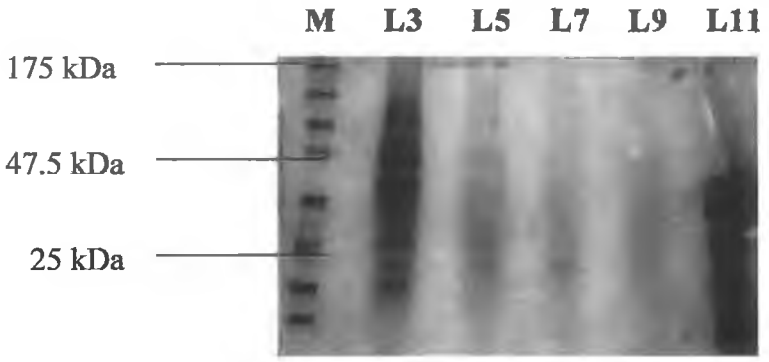


Figure 42: SDS-PAGE of D3 supernatants from induction # 1

10 µL protein and 5 µL 4 X reducing buffer were applied to each well. The molecular weight marker (M) is depicted in L1. Many bands can be seen in L3 for A1 but are not significant as they are due to cell debris. L5 contains A2 supernatant and a faint band at ~ 22 kDa is visible. C2 supernatant also shows this 22 kDa band in L7.

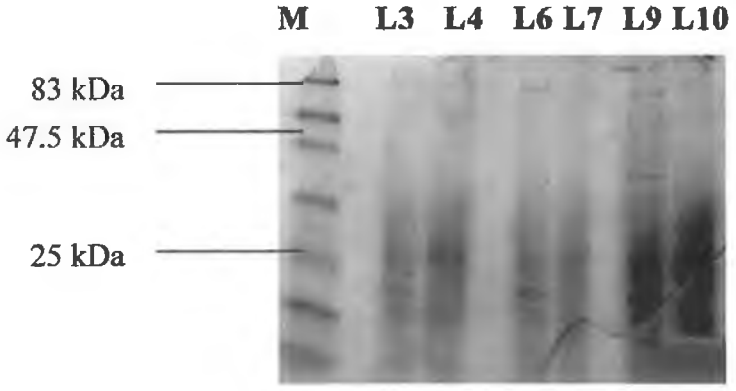


Figure 43: SDS-PAGE of D4-D6 supernatants for B2 and C2 clones

Lane 1 contains the molecular weight marker (M). C2 D4 supernatant in L3 has a band at 22kDa, which is not seen for B2 D4 (L4). This 22kDa band is also visible for C2 D5 in L6 along with an additional band at 83

kDa, both of which are absent for B2 D5. The 22 kDa and 83 kDa bands are visible again on analysis of C2 D6 supernatant along with a band at ~ 40 kDa which was not previously seen (L9). All three bands are absent from B2 D6 (L10).

SDS-PAGE was carried out using D6 supernatants and Ni-NTA beads. Bands were seen on the gel at 22 kDa for A1, A2 and C2 but not for B2, as expected. Pellets were disrupted for D1-D6 samples and were analysed by SDS-PAGE. As no D0 sample had been taken for this induction in error, it was impossible to ascertain which resolved bands were indicative of induced protein.

Western Blot

A western blot probing the D3 supernatants (A2, B2 and C2) with Anti-His_6 antibody did not show any binding. Binding was observed using FhCL1 and Sm cathepsin B (SmCB) samples as expected. A western blot was also carried out on D1-D6 disrupted pellets but the Anti-His_6 antibody did not bind to any of them (data not shown).

Enzyme Activity Assays

Enzyme activity assays were carried out for A2, B2 and C2 D1-D6 supernatants to detect if rSmCC was present using the cathepsin C specific substrate H-Gly-Arg-NHMec. All assays were read at slit width 10, 2.5. All assays were performed overnight (~ 18 hrs) at 37^0C unless stated otherwise.

An NHMec standard curve was prepared for each enzyme assay that was carried out in this project (data not shown). The standard curves prepared are shown in the Appendix.

Table 38: Enzyme activity assay using supernatants from induction # 1

Day	A2	B2	C2
D1	0.91	0.58	0.77
D2	0.98	1.12	0.77
D3	1.01	1.30	0.47
D4	1.01	1.01	0.96
D5	1.06	0.95	1.21
D6	1.21	1.17	1.55

The positive control contained all the usual reaction components except the enzyme was *F. hepatica* somatic extract. The average fluorescence of this positive control was 192.9 fluorescence units.

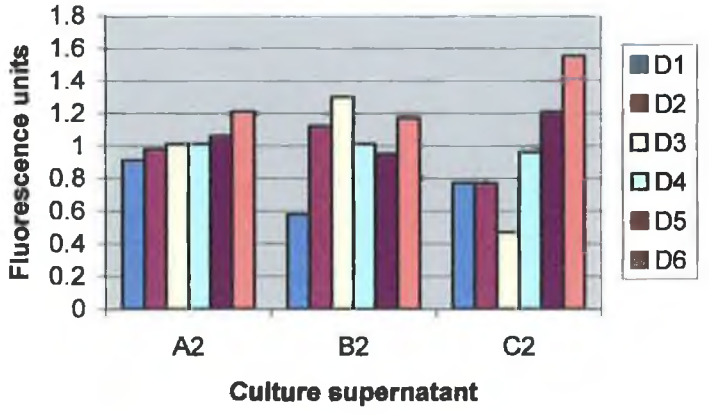


Figure 44: Enzyme activity assay for rSmCC from induction # 1

Enzyme activity of D1-D6 supernatants using the cathepsin C specific substrate H-Gly-Arg-NHMec is depicted in the graph above. For both A2 and C2 supernatants, which are expected to exhibit rSmCC activity, activity increases reaching a maximum at D6. However, cleavage of the SmCC specific substrate also occured for B2, and this should not be the case as SmCC cDNA was not present in B2 transformants.

Enzyme activity assays were also carried out on D6 pellets from culture supernatants A2, B2 and C2 (Table 39).

Table 39: Enzyme activity assay using D6 pellets from induction # 1

Pellet	Average fluorescence
D6 A2	7.09
D6 B2	3.99
D6 C2	8.66
Background control	0.957
Positive control	180.43

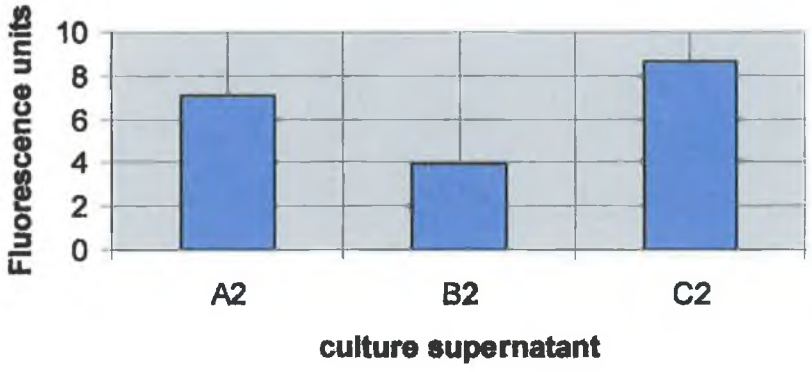


Figure 45: Enzyme assay using D6 pellets from induction # 1

For B2, which should not contain SmCC, non-specific cleavage of the substrate occurred. D1, D3 and D5 pellets were also assayed. A D0 control from a later induction was used but on average five times more substrate was cleaved for D0 control sample than for the pellet samples, which reinforces the notion that yeast proteases capable of hydrolysing the SmCC substrate were produced.

3.3.2.2: INDUCTION # 2 (SCALE-UP)

Culture Conditions

Using the YPD replica plate of C2 transformants, 2 colonies were selected and were used to inoculate 250 ml of BMGY. Resupsended pellets were combined, giving a final volume of 100 ml BMMY, pH 6.0 in the culture flask. Methanol was added daily to a final concentration of 1% for six days as before.

SDS-PAGE Analysis

SDS-PAGE analysis on D0-D6 supernatants did not reveal any significant bands. The numerous bands present on the gels were attributed to yeast cell debris. SDS-PAGE was also carried out after addition of Ni-NTA beads to the supernatants as before. A weak band was seen at 22 kDa for the D6 supernatant sample but not for any of the others. The D0-D6 pellets from this induction were resupsended in dH_20 and were applied to the gel as usual (Figure 46).

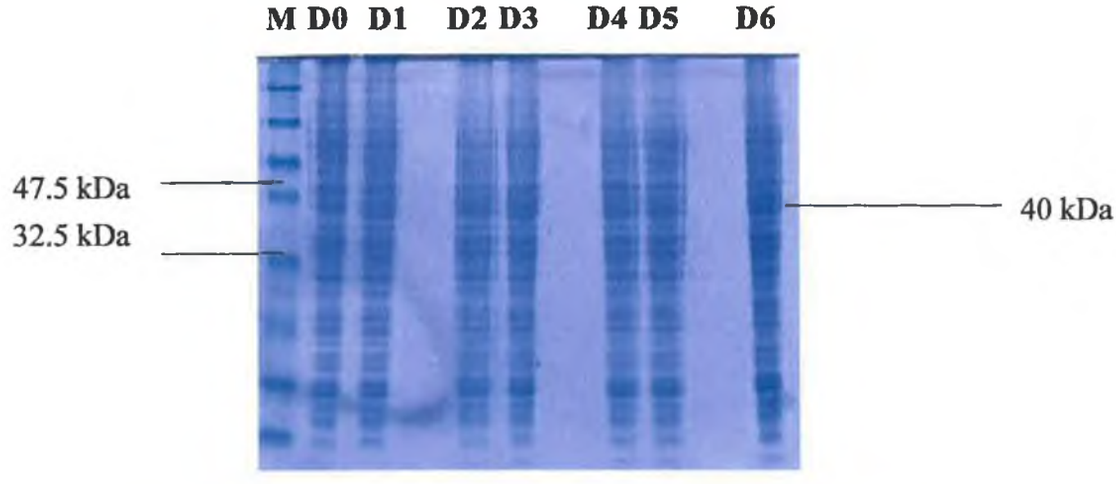


Figure 46: SDS-PAGE analysis of D0-D6 pellets from induction # 2

The molecular weight marker is depicted in lane 1 (M). It is clear to see that most of the bands present in D1-D6 pellets are also found in D0, which indicates that they do not represent induced protein. However, a

band is visible in D1-D6 at ~ 40 kDa that is not present in the D0 pellet sample, which indicates that this band could be expressed SmCC.

Western Blot

A western blot was carried out using the D0-D6 pellets and Anti-His_6 antibodies. No binding was observed except for SmCB purified protein included as a control.

Affinity Chromatography on a Ni-NTA column

The D6 sample from induction # 2 was diluted in a 1:5 sample:lysis buffer ratio before it was applied to the column. SDS-PAGE of column fractions did not reveal the presence of any protein except for the dialysed sample, which showed a faint band at 62 kDa. A BCA assay was carried to determine the protein concentration of all column fractions (Table 40). As is clear from the BCA assay results, the protein concentration of the sample before it was applied to the column (0.89582 mg/ml) was practically the same as that of the run through sample (0.884099 mg/ml), which illustrates that very little, if any protein was bound to the Ni-NTA beads in the column.

Table 40: BCA assay of Ni-NTA affinity column fractions

Average							
BSA mg/ml	Lysis	Wash	Elution	Flush	PBS		
0	0.099	0.114	0.127	0.123	0.0875		
0.2	0.539	0.241	0.2065	0.195	0.2175		
0.4	0.734	0.3415	0.279	0.252	0.357		
0.6	0.4225	0.81	0.357	0.3325	0.5865		
0.8	0.5105	0.495	0.4235	0.3945	0.76		
1	0.619	0.598	0.512	0.4695	0.895		
1.2	0.6425	0.6775	0.57	0.5005	0.864		
1.4	1.0325	0.6765	0.5955	0.5565	1.049		
1.6	1.1355	0.686	0.6725	0.5965	1.1335		
1.8	1.3115	0.8425	0.706	0.6525	1.2095		
2	2.3485	1.006	0.821	0.6985	1.332		
	Pre	Run Thru	Wash	Eluate	Flush	Post Dialysis	Day 6
	0.797	0.762	0.21	0.139	0.134	0.101	2.027
	0.751	0.768	0.213	0.137	0.136	0.107	2.027
Average	0.774	0.765	0.2115	0.138	0.135	0.104	2.027
Conc. Mg/ml	0.89582	0.884099	-0.06503	-0.00605	-0.04404	-0.08459	3.037668
Lysis	y = 0.7679x + 0.0861						
Wash	y = 0.3552x + 0.2346			Flush	y = 0.2861x + 0.1476		
Elution	y = 0.3305x + 0.14			PBS	y = 0.6159x + 0.1561		

Enzyme Activity Assays

A regular enzyme assay and a leupeptin inhibition assay were performed on D0, D4, D5 and D6 culture supernatants (Table 41). Leupeptin was used at a concentration of 10 μM and this was observed to be insufficient to cause inhibition of SmCC. Higher leupeptin concentrations were employed in later assays, it was seen that a 0.5 mM final concentration of leupeptin was necessary to result in at least 55% inhibition of SmCC.

Table 41a: Results of enzyme activity assays of induction # 2 supernatants

Day	Average Fluorescene
D0	0.71
D4	0.743
D5	0.892
D6	0.786

Table 41b: Controls for induction # 2 assays

Control	Average Fluorescence
Positive (Fh somatic extract)	205.7
Background	0.643

As is evident from the results above, no specific SmCC activity was present as the fluorescence detected for D0 culture supernatant was only slightly lower to that for D6 supernatant.

Assays were carried out in sodium citrate (SC), sodium acetate (SA), sodium phosphate (SP) and Tris-HCL (T) buffers at different pHs to determine whether rSmCC had a different pH optimum to of pH 5.5 reported in the literature (Table 42). Day 6 culture supernatant was used for all pH profile assays. The average fluorescence read for each pH was plotted against pH (Figure 47).

Table 42a: Controls for pH profile assay

Controls	Average Fluorescence
Positive control (Fh somatic extract)	216
Background	0.603

Table 42b: pH profile assays for SmCC, induction # 2

Buffer	Average fluorescence units
SC pH 3.0	0.97
SC pH 4.0	0.59
SC pH 5.0	1.0
SC pH 6.0	1.49
SC pH 7.0	1.74
SA pH 4.0	0.66
SA pH 4.5	0.71
SA pH 5.0	0.72
SA pH 5.5	1.22
SA pH 6.0	1.02
SP pH 5.0	1.78
SP pH 5.5	3.46
SP pH 6.0	6.28
SP pH 6.5	9.6
SP pH 7.0	13.04
SP pH 7.5	15.43
T pH 6.0	0.96
T pH 7.5	4.73
T pH 8.0	2.94
T pH 8.5	6.11

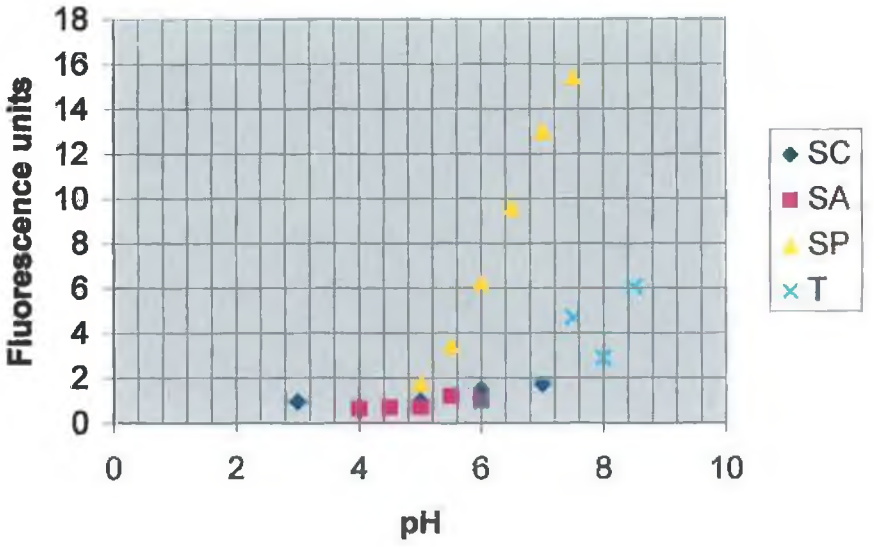


Figure 47: Graph of pH profile assays using the SmCC substrate H-Gly-Arg-NHMec and D6 culture supernatants from induction # 2

Disrupted D6 pellets from C2, induction # 2 were assayed for SmCC activity at pH 5.5 but activity was seen to be less than that of the B2 control. This would suggest that the enzyme activity seen was not due to rSmCC.

3.3.2.3: INDUCTION # 3

Culture Conditions

For this induction A2, B2 and C2 clones were used to inoculate 100 ml of BMGY each. After 41 hrs incubation an OD_{600} of 2.6 was reached for all cultures. Induction was initiated by resupsension of each culture in 20 ml of BMMY pH 6.0.

SDS-PAGE Analysis

On SDS-PAGE analysis of D0-D7 supernatant samples, the only bands seen for C2 were bands of approximately 200 kDa. No significant bands were seen for any of the others. Analysis of disrupted D0-D7 pellets by SDS-PAGE did not reveal the presence of any induced bands. Bands observed at D1-D7 were also present in D0 pellets.

Western Blot and ELISA

A western blot and an ELISA (Table 43) were performed on D0-D7 supernatants and disrupted pellets. Positive controls worked well for both procedures but no binding of Anti-His_6 antibody to the SmCC hexahistidine tag was observed for any of the D0-D7 supernatants or pellets.

Table 43: ELISA results for D0-D7 supernatants (average A450 nm of duplicate wells)

	1	2	3	4	5
A	FhCL1 0.190	SmCB 0.655	SmCB (dialysed) 0.0765	Blank 0.030	
B	D0 C2 0.046	Blank 0.047	D3 C2 0.048	D6 C2 0.049	D7 C2 0.051
C	D0 A2 0.042	D1 A2 0.0505	D3 A2 0.074	D6 A2 0.058	D7 A2 0.058
D	D0 B2 0.043	D1 B2 0.044	D3 B2 0.048	D6 B2 0.045	D7 B2 0.045

Enzyme Assays

Enzyme assays were performed on D0, D1, D3, D6 and D7 supernatants from cultures A2, B2 and C2. Assays were carried out in a sodium acetate pH 5.5 buffer and a bar chart was plotted depicting the average fluorescence obtained using each supernatant (Table 44a below and Figure 48).

Table 44a: Enzyme Assay of culture supernatants, induction # 3

Supernatant	Average fluorescence
A2 D0	0.6
A2 D1	0.52
A2 D3	0.48
A2 D6	0.59
A2 D7	0.48
B2 D0	0.46
B2 D1	0.60
B2 D3	0.55
B2 D6	0.66
B2 D7	0.66
C2 D0	0.66
C2 D1	0.57
C2 D3	0.50
C2 D6	0.53
C2 D7	0.66

Table 44b: Controls used in the enzyme assay

Control	Average fluorescence
Background	1.011
Positive	163.03

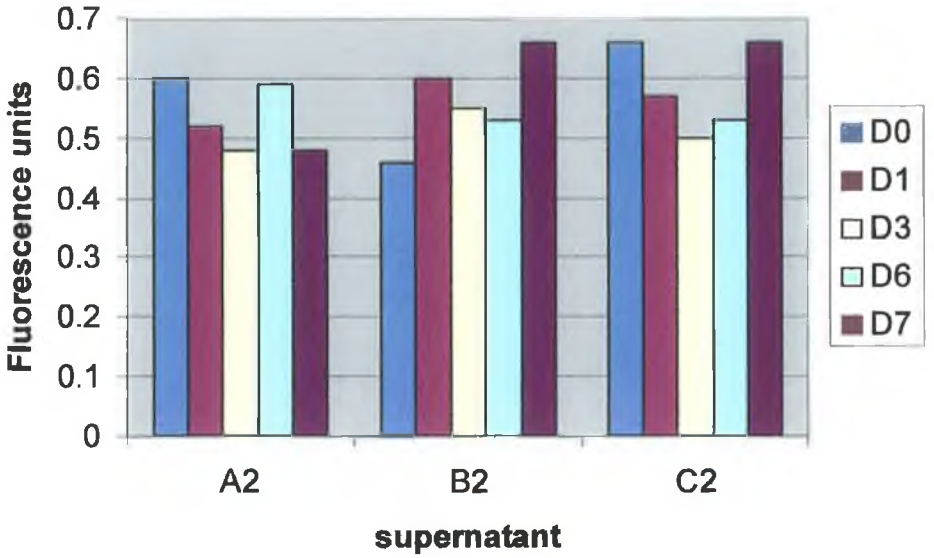


Figure 48: Bar chart depicting the results of the enzyme assay for D0-D7 supernatants, induction 3

Enzyme assays were also performed on disrupted cell pellets (Table 45). Assays were performed in sodium acetate pH 5.5 and a bar chart was prepared showing the average fluorescence (NHMec released) for pellets from each day of the induction (Figure 49).

Table 45a: Controls used in the enzyme assay

Control	Average fluorescence
Background	0.60
Positive	177.05

Table 45a: Enzyme Assay of pellets, induction # 3

Pellet	Average fluorescence
D0	5.60
D1 A2	1.63
D3 A2	3.07
D6 A2	11.92
D7 A2	7.91
D1 B2	4.47
D3 B2	4.81
D6 B2	10.89
D7 B2	5.96
D1 C2	7.19
D3 C2	7.95
D6 C2	9.61
D7 C2	4.25

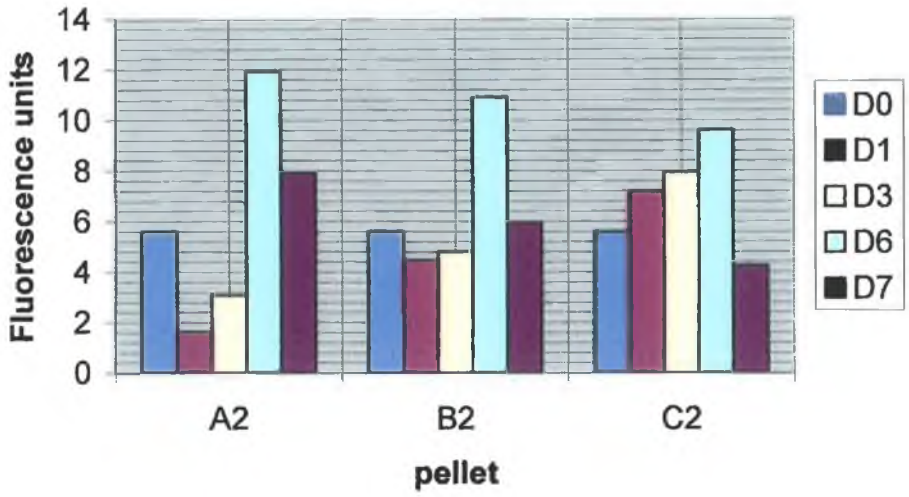


Figure 49: Bar chart of enzyme assay of D0-D7 pellets, induction # 3

3.3.2.4: INDUCTION # 4

Culture Conditions

Colonies from B2 and C2 replica plates were used to inoculate 500 mls of BMGY, which was then split into 2 x 250 ml cultures. An OD_{600} of 2.2 was reached by both cultures after 23 hours incubation. Each duplicate 250 ml culture pellet was pooled after resuspension in 50 ml BMMY to give a final culture volume of 100 ml. Methanol was added twice daily for 7 days to a final concentration of 1.5% in the culture medium.

SDS-PAGE Analysis

Culture supernatants were analysed by SDS-PAGE. Pellets were disrupted by the usual method but recovery proved insufficient to permit further analysis to be undertaken.

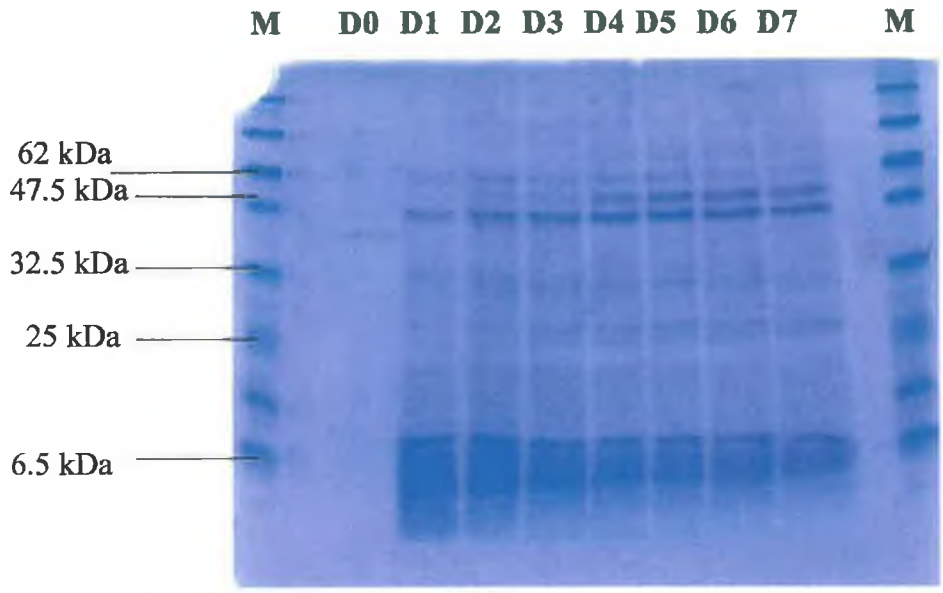


Figure 50: SDS-PAGE of D0-D7 supernatants from C2, induction # 4

The usual molecular weight markers (M) are depicted in the first and last lanes. The D0 supernatant displays a band at ~ 40 kDa. D1 supernatant shows a band at 47kDa, which is also present for D2-D7 samples. A band at ~ 58 kDa is visible for D2-D7 that is only faintly visible at D1. At D4 a band appears of 55 kDa, which was absent from previous days but present in later ones. As the predicted molecular size of the pro-mature SmCC is 49.5 kDa, these bands at > 50 kDa could be indicative of differentially glycosylated forms of pro-mature SmCC. The appearance of a faint 25 kDa band in D2-D7 samples could indicate that the mature enzyme is being produced. The fact that enzyme activity was detected for these samples supports this notion. For D1-D7 supernatants the bands < 6.5 kDa that appear on the gel could be breakdown products that arose from proteolysis of the pro-mature SmCC.

Supernatants bound to Ni-NTA beads were resolved by SDS-PAGE and the same bands were seen for this gel as for regular SDS-PAGE.

Western Blot and ELISA

Two western blots were carried out on D0-D6 supernatants using the usual anti-His_6 antibody and anti-schistosome mouse sera. The secondary antibody in each case was goat anti-mouse IgG. In both cases binding occurred to SmCB but not SmCC protein. An ELISA was performed using both sets of antibodies and although binding occurred for SmCB and FhCL1 controls, none was observed for SmCC.

A western blot was performed using anti-*S. japonicum* cathepsin C rabbit sera. Purified baculovirus expressed rSjCC was included as a control.

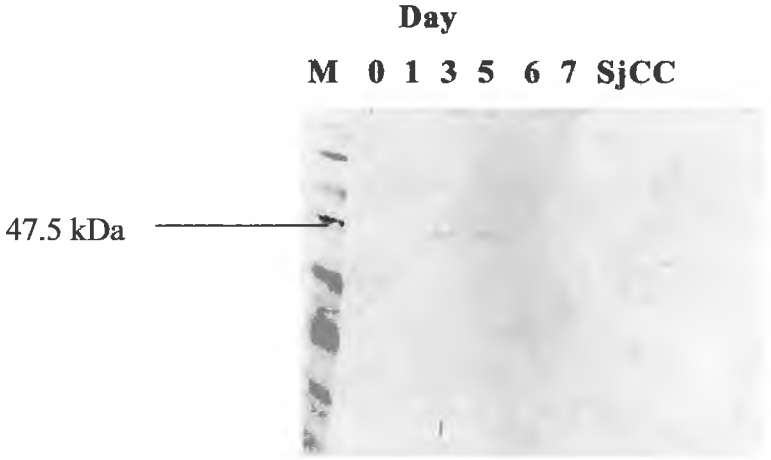


Figure 51a: Western Blot using anti-SjCC rabbit sera

No binding is seen for D0 or D1 supernatants. With D3 and D5 supernatants antibodies bound to a band at 47 kDa. Binding was not observed for any other supernatants or for the SjCC postive control. Figure 51b shows the supernatants used in this blot resolved by SDS-PAGE. For D3 and D5 supernatants the band at 47 kDa to which binding occurred in the western blot can be clearly seen. This band is also visible at D1, D6 and D7 despite the fact that binding did not occur for these. SjCC resolves as a band at ~ 150 kDa.

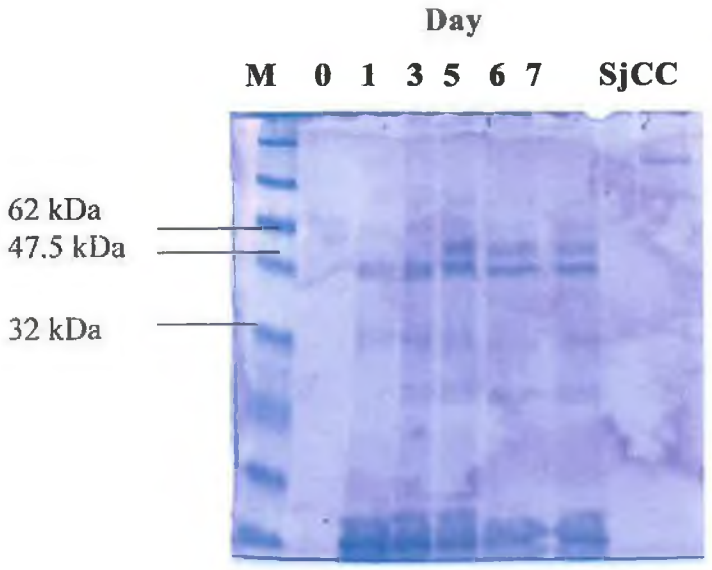


Figure 51 b: Samples used for western blot as resolved by SDS-PAGE

An ELISA was carried out using anti-SjCC rabbit sera. As depicted in Table 46, binding of the antibody to SmCC was minimal. Neat and diluted SjCC and S. mansoni somatic extract were included as controls but only weak binding was observed for these controls.

Table 46: ELISA results (average A450 nm of duplicate wells)

	1	2	3	4
A	SjCC (dilute) 0.068	SjCC (neat) 0.049	Sm somatic extract 0.084	Blank 0.030
B	D0 C2 0.048	D1 C2 0.0445	D2 C2 0.044	D3 C2 0.046
C	D4 C2 0.047	D5 C2 0.0435	D6 C2 0.045	D7 C2 0.044

Enzyme Assays

Enzyme assays were performed on B2 and C2 D0-D5 culture supernatants (Table 47a and b). The buffer used in the assays was sodium acetate pH 5.5. A standard curve was prepared as usual (refer to Appendix). A bar chart was prepared depicting average fluorescence (NHMec liberated) for B2 and C2 for each day of the induction (Figure 52).

Table 47a: Controls used in the enzyme assay

Control	Average fluorescence
Background	0.675
Positive	169.85

Table 47b: Enzyme Assay of induction # 4 culture supernatants

Supernatant	Average fluorescence
D0 B2	0.93
D1 B2	0.49
D2 B2	0.48
D3 B2	0.91
D4 B2	0.53
D5 B2	0.57
D0 C2	0.53
D1 C2	0.77
D2 C2	1.69
D3 C2	4.38
D4 C2	10.77
D5 C2	30.21

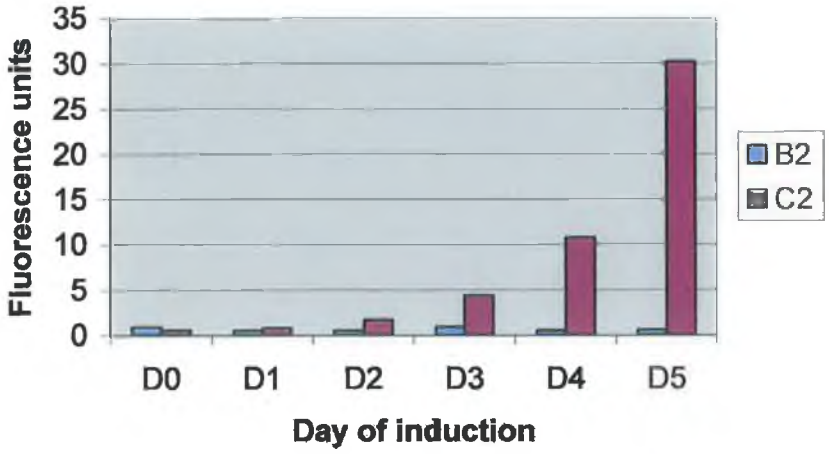


Figure 52: Bar chart of enzyme activity assay for B2 and C2 (D0-D7), induction # 4

An enzyme assay at pH 5.5 was set up to determine the difference in enzyme activity over a 4 hr. and 20 hr. incubation period using D5 supernatants (Table 48). The difference in fluorescence of duplicate assays over the seven-day period of both the 4 hr and 20 hr assays is detailed in Figure 53.

Table 48: Enzyme assay of C2 culture supernatants over a 4hr. and 20 hr. induction period

Day	4 hrs	20 hrs
D0	0.73	0.26
D1	0.81	0.41
D2	1.81	1.45
D3	2.38	3.80
D4	3.91	9.91
D5	9.01	31.43
D6	8.87	30.47
D7	9.24	28.46

The usual controls were included which proved that the assay was working properly (data not shown).

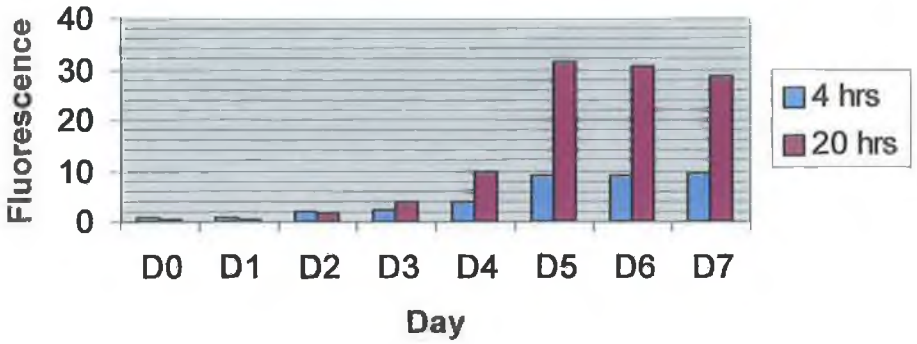


Figure 53: Bar chart showing the effect of a 4 hr. and 20 hr. incubation period on enzyme activity

An inhibition assay was performed using the D5 supernatant from induction # 4 and leupeptin to a final concentration of 0.5 mM (Table 49). The bar chart in Figure 54 clearly shows the inhibitory effect leupeptin had on rSmCC activity. Three controls were included in this assay as follows:

- Control 1 contained 10 µL of D5 supernatant and no leupeptin.
- Control 2 contained 10 µL of F. hepatica somatic extract and 0.5 mM leupeptin.
- Control 3 contained 10 µL of F. hepatica somatic extract and no leupeptin.

Table 49: Leupeptin inhibition assay using D5 supernatant induction # 4

Assay	Average fluorescence
Background control	0.55
Leupeptin + rSmCC	2.73
Control 1 (rSmCC)	29.13
Control 2 (Fh + leupeptin)	92.57
Control 3 (Fh alone)	159.9

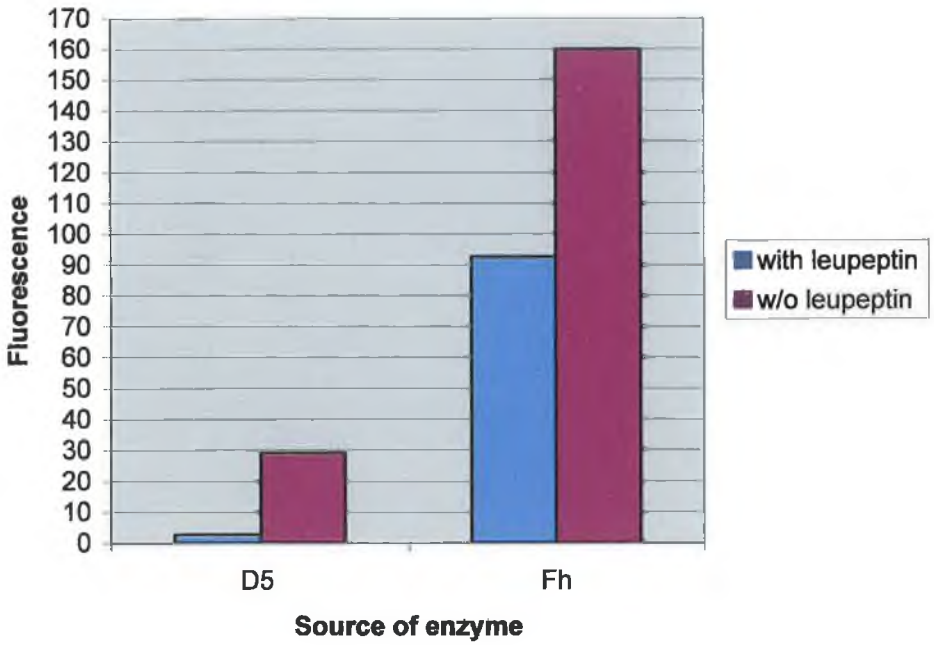


Figure 54: Bar chart of leupeptin inhibition assay using D5 supernatant, induction # 4

Leupeptin inhibited the rSmCC activity by 90.6%. The cathepsin C activity in the *F. hepatica* somatic extract exhibited 42% inhibition in the presence of leupeptin.

A similar assay was set up using iodoacetamide to a final concentration of 1 mM in the assay instead of leupeptin, to study its effect on rSmCC activity. Using the same D5 supernatant, only 9.32% inhibition of rSmCC activity was observed. Using *F. hepatica* somatic extract 94.6% inhibition of cathepsin C was observed with iodoacetamide.

An assay was set up to investigate if the cathepsin C specific substrate H-Gly-Arg-NHMec could be non-specifically cleaved by recombinant FhCL1 (Table 50). The bar chart in Figure 55 shows the cleavage of the substrate by rFhCL1 was low relative to that observed for rSmCC.

Table 50: Assay to determine whether rFhCL1 can cleave H-Gly-Arg-NHMec

	rSmCC	rFhCL1
D0	0.76	0.99
D1	0.97	1.65
D3	4.35	3.612
D6	26.77	4.58
Background	0.57	0.57
Positive control	156.3	329.4

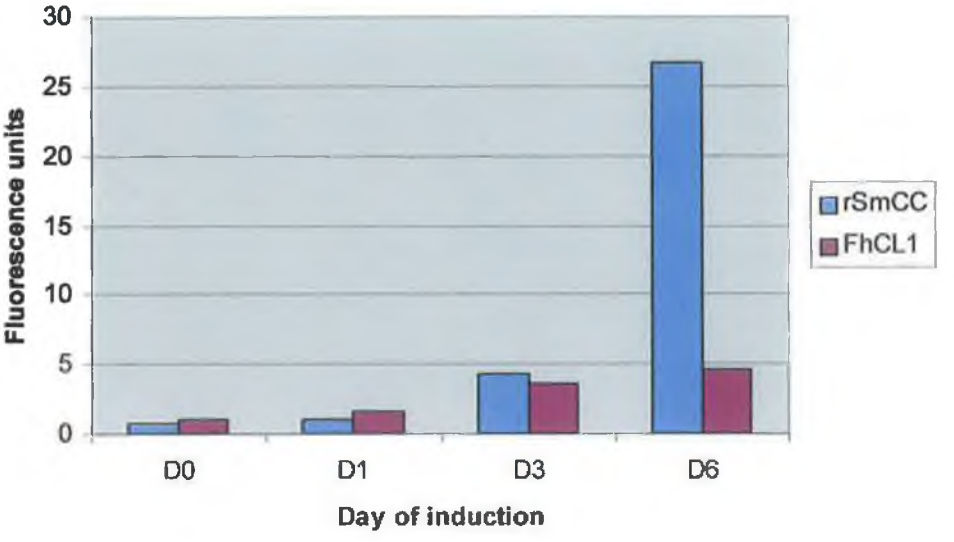


Figure 55: Bar chart depicting the cleavage of cathepsin C specific substrate by rSmCC and rFhCL1

3.3.2.5: INDUCTION # 5

Culture Conditions

To investigate the effect of pH on expression of SmCC, three separate inductions were carried out at pH 5.0, 6.0 and 7.0. The culture conditions were as before (refer to 3.3.2.4), using C2 transformants, the only difference being the pH of the induction media. After initial growth in BMGY pH 6.0, each culture was brought to an OD_{600} of 20 prior to initiation of induction. The final pH of the BMMY in each flask was determined at the end of the 6-day induction period (Table51).

Table 51: pH of BMMY culture media after 6 days induction

pH 5.0	**6.5**
pH 6.0	**7.5**
pH 7.0	**8.5**

SDS-PAGE Analysis

Figure 56 shows the supernatants from the 6-day induction period as resolved by SDS-PAGE.

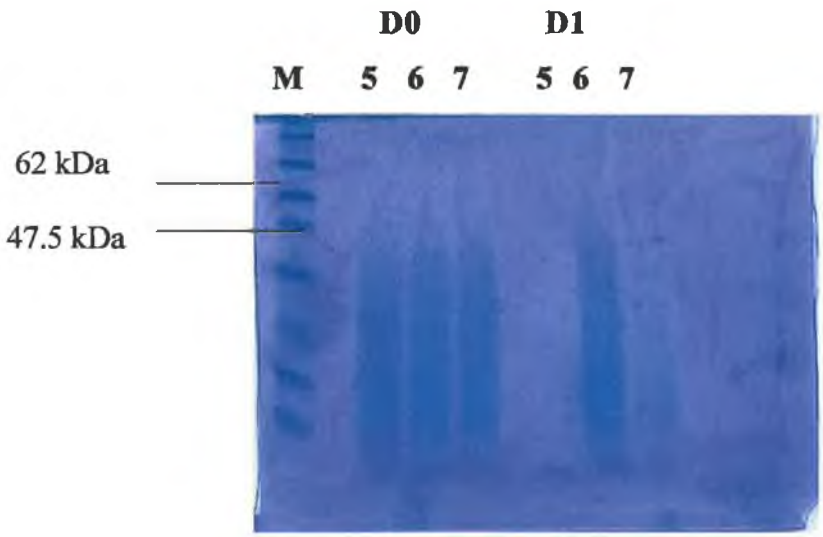


Figure 56a: pH optimum studies on the expression of SmCC (D0 and D1)

The usual molecular weight marker is shown in lane 1(M). No bands are visible for D0 supernatants at any pH. A single band appears at ~ 42 kDa for the D1 pH 7.0 supernatant but not for any of the other D1 samples.

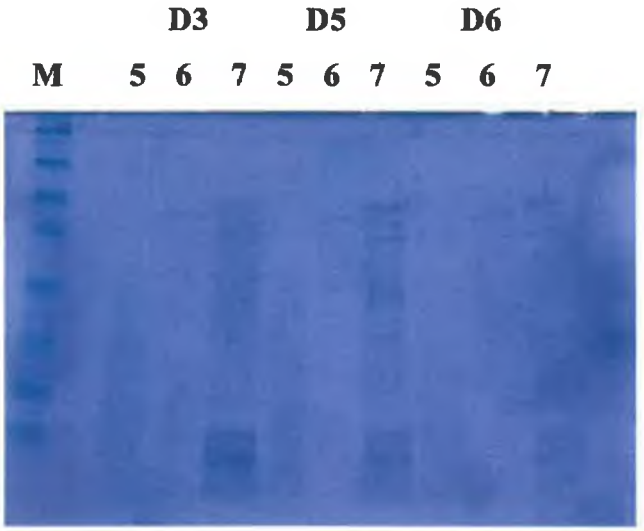


Figure 56b: pH optimum studies on the expression of SmCC (D3, D5, D6)

No band is apparent for D3 pH 5.0 supernatants. A single band at ~ 55 kDa can be see for D3 pH 6.0. Three distinct bands are visible for D3, pH 7.0, one at 60 kDa, a faint one at 55 kDa and one at 47 kDa. Degradation products can be clearly seen at the bottom of the gel. The same bands are seen for D5 and D6 supernatants.

Western Blot and ELISA

A western blot was performed using D0 pH 5.0, 6.0 and 7.0 and D4, pH 5.0, 6.0 and 7.0 supernatants and anti-*S. japonicum* cathepsin C antibodies (Figure 57). A western blot was also carried using the anti-His_6 antibody and although the SmCB positive control worked well, no binding of the antibodies to rSmCC his-tag was observed.

In order to determine whether C2 clones actually expressed rSmCC, a colony blot was performed (refer to 2.2.6). Probing of the colony blot with anti-His_6 antibodies failed to reveal any significant binding. Binding of anti-His_6

antibody to the His-tag of rFhCL1, the positive control in this experiment occurred. This indicated that rSmCC was not expressed in the C2 clones.

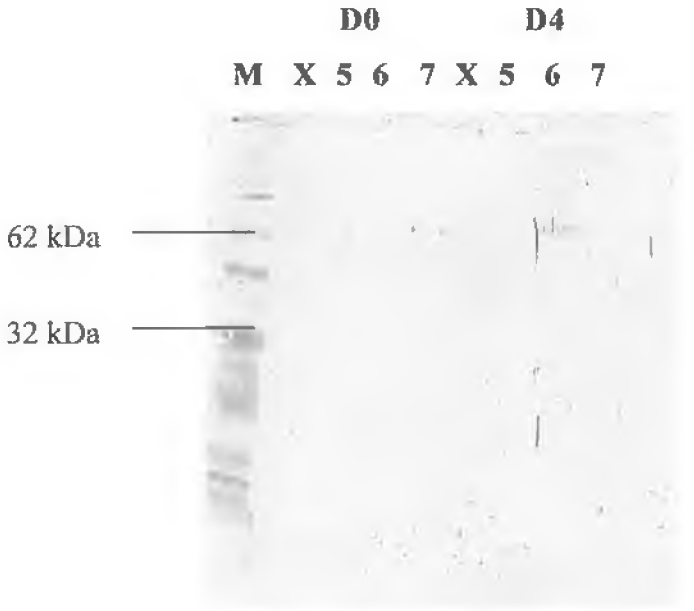


Figure 57: Western Blot of D0 and D4 pH 5.0, 6.0 and 7.0 supernatants using anti-SjCC antibodies

Binding occurred to a 62 kDa and 28 kDa band using D4 pH 6.0 supernatant. Binding to the 28 kDa band was also observed for D4 pH 7.0 supernatants.

Enzyme Assays

Enzyme assays were carried out to determine whether culturing at different pHs had an effect on the activity of rSmCC. All assays were performed at pH 5.5 (Table 52). A bar chart was prepared showing the average fluorescence obtained using D0-D6 supernatants from inductions performed at pH 5.0, 6.0 and 7.0.

Table 52a: Enzyme assays using D0-D6 supernatants at pH 5.0-7.0 (average fluorescence of duplicate assays)

	pH 5.0	pH 6.0	pH 7.0
D0	0.63	0.58	0.88
D1	3.58	0.46	0.76
D3	4.91	2.35	1.34
D4	4.75	3.01	1.42
D5	4.89	2.27	1.45
D6	4.46	1.78	1.43

Table 52b: Controls for the assay

Control	Average fluorescence
Background	0.69
Positive	154.37

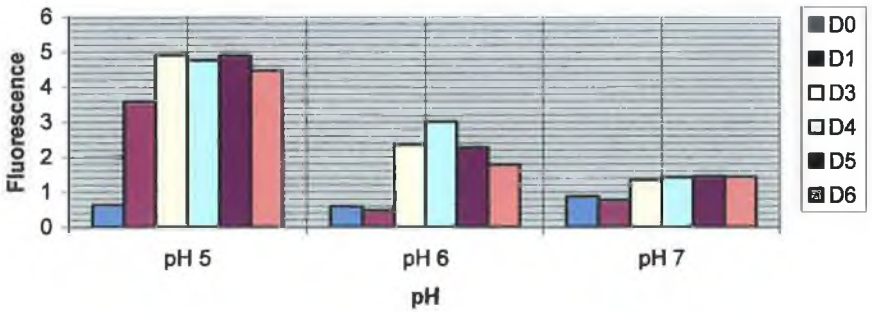


Figure 58: Bar chart depicting the enzyme activity of rSmCC of supernatants induced at pH 5.0, 6.0 and 7.0

3.3.2.6: INDUCTION #6

Culture Conditions

Culture conditions were exactly as for 3.3.2.4. Incubation for 18 hrs was sufficient for an OD_{600} 3.0 to be attained for both cultures. Resupsended pellets in BMMY pH 6.0 were combined and upon initiation of induction the culture volume was 100 ml. Methanol was added to a final concentration of 1.5% twice daily for 5 days.

SDS-PAGE

D0-D5 supernatant samples were resolved by SDS-PAGE

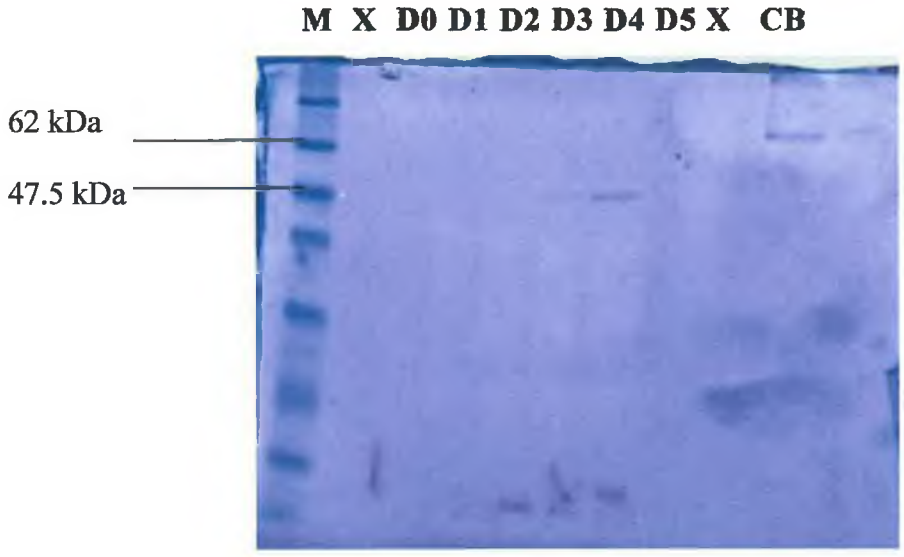


Figure 59: SDS-PAGE of D0-D5 supernatants from induction # 6

The usual molecular weight marker is depicted in lane 1 (M). Single bands at ~ 47 kDa are visible for D3 and D4 supernatants only. A band at ~ 70 kDa can be seen for SmCB in the final lane.

Western Blot

A western blot was carried out on D0-D5 samples using anti-His_6 antibodies and anti-SjCC sera. No binding was observed for either western blot, although the positive control SmCB reacted with anti-His_6 and anti-SjCC. This was a surprising result with anti-SjCC as it was not expected that SmCB would cross-react with anti-SjCC antibodies.

Enzyme Assays

An overnight enzyme assay was performed at pH 5.5 on D0-D5 culture supernatants to ascertain whether active rSmCC was produced (Table 53). A bar chart was prepared plotting the average fluorescence of duplicate assays performed over 4 and 20 hrs using D0-D5 culture supernatants (Figure 60).

Table 53: Enzyme assays using D0-D5 culture supernatants, induction # 6

Supernatant	Average fluorescence
D0	0.32
D1	0.15
D2	5.01
D3	6.65
D4	9.28
D5	9.09
Background	1.88
Fh somatic extract positive control	142.55

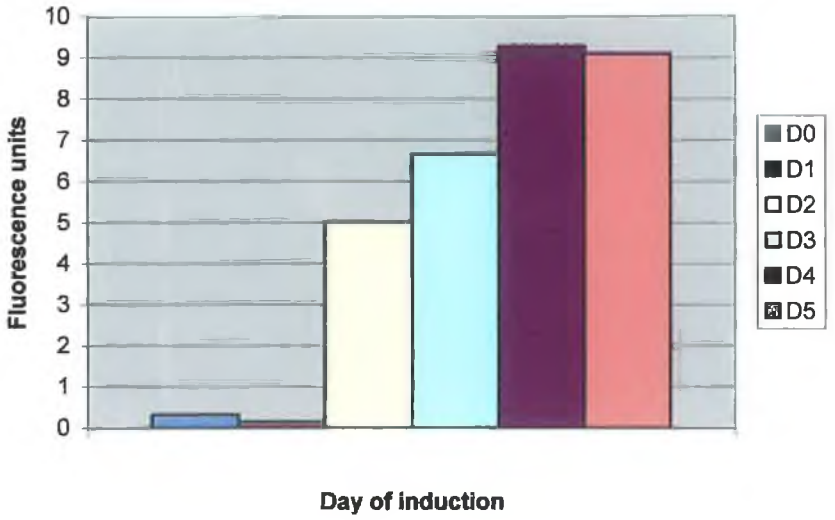


Figure 60: Bar chart of enzyme activity assay using D0-D5 supernatants from induction # 6

A 4 hr. and 20 hr. assay was performed to assess the effect of incubation time on the activity of the recombinant enzyme. The bar chart in Figure 61 depicts the average fluorescence obtained (NHMec released) after the assay was carried out for 4 hrs and 20 hrs.

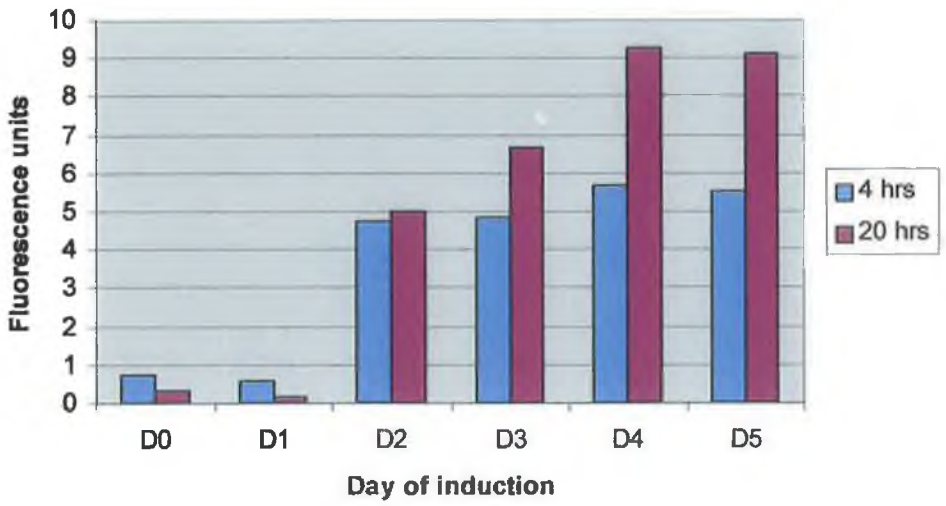


Figure 61: Bar chart showing rSmCC activity (fluorescence) after 4 hrs and 20 hrs incubation

The increase in rSmCC activity over time is non-linear. Maximum activity was observed at D4 for both the 4 hr. and 20 hr. assay.

3.4: BIOCHEMICAL CHARACTERISATION OF *F. HEPATICA* CATHEPSIN C

Cathepsin C from *F. hepatica* somatic extract (FhCC) was characterised to ascertain whether FhCC shared the following biochemical characteristics displayed by other cathepsin Cs, in particular SmCC:

- Enhancement of enzyme activity in the presence of the reducing agent DTT.
- Optimum enzyme activity at pH 5.5.
- Inhibition by the general cysteine protease inhibitors leupeptin, iodoacetamide and NEM and slow inibition by E-64.
- Enhancement by halide ions.

3.4.1: ENHANCEMENT OF ENZYME ACTIVITY BY DTT

A series of enzyme assays were performed in the presence and absence of the reducing agent DTT in order to determine its effect on FhCC activity.

- Addition of 5 µL of 2 mM DTT to the assay resulted in a five-fold increase in fluorescence (NHMec liberated) from 16.72 nm in the absence of DTT to 84.8 nm when DTT was added. The enzyme activity for the DTT assay was calculated to be 0.67 U (refer to Section 2.2.7). The enzyme activity could not be determined for the assay without DTT owing to the fact that the fluorescence value did not fall within the area of the NHMec standard curve (see Appendix for standard curves) but the enhancing effect of DTT on FhCC activity was evident.
- An assay was carried out using 15 µL DTT at pH 5.5 and 8.5 and all assays were read on the fluorimeter (slit width: 10, 10) (see Figure 62).

The addition of DTT resulted in 67% enhancement of FhCC activity at pH 5.5 and 50% enhancement at pH 8.5.

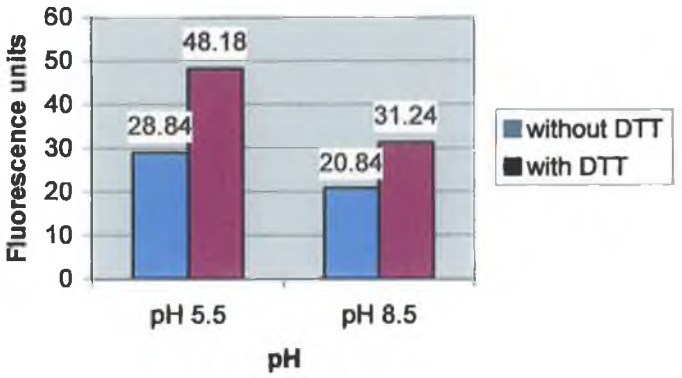


Figure 62: Bar chart showing the enhancing effect of DTT on FhCC activity

- In assays carried out for rSmCC activity, 100 µL of 20 mM DTT was used in a 1 ml assay total assay volume. At slit width 10, 10 on the fluorimeter FhCC activity proved to be off-scale giving a result of 1000. At 10, 2.5 an average reading of 180 was obtained. This DTT concentration was employed in all assays for rSmCC.

3.4.2: pH PROFILE ASSAYS OF FhCC

In order to determine the pH optimum of FhCC, a series of pH profile assays were performed. All assays were prepared as usual except 150 µL of DTT was added to all assays (Table 55).

Table 54: Buffers used in the pH profile assays

Buffer	pH
Sodium acetate (SA)	4.0, 4.5, 5.0, 5.5
Sodium phosphate (SP)	5.5, 6.0, 6.5, 7, 7.5, 8.0
Tris-HCl (T)	7.5, 8.0, 8.5, 9.0

Table 55: pH profile assay results (average fluorescence at 10, 10)

Buffer	Result	Control
SA pH 4.0	6.85	14.36
SA pH 4.5	21.51	14.0
SA pH 5.0	83.18	17.73
SA pH 5.5	138.03	16.87
SP pH 5.5	34.47	15.44
SP pH 6.0	43.96	16.14
SP pH 6.5	47.02	19.28
SP pH 7.0	60.25	22.86
SP pH 7.5	75.98	25.51
SP pH 8.0	91.68	26.57
T pH 7.5	106.45	14.8
T pH 8.0	137.45	13.37
T pH 8.5	205.45	17.06
T pH 9.0	11.84	18.75

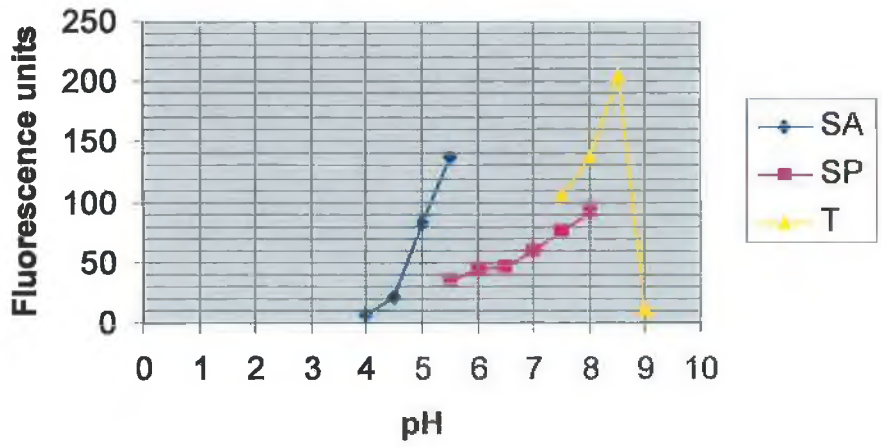


Figure 63: Graph depicting the results of the pH profile assay of FhCC

A pH profile assay was carried out following incubation of the enzyme with the cathepsin-L inhibitors Z-Phe-Ala-NH_2 and Z-Phe-Phe-OH to investigate whether the peak of activity observed at pH 8.5 was due to the presence of another cysteine protease, such as cathepsin L.

Table 56: Results of pH profile assay using Z-Phe-Ala-NH_2 inhibitor

Buffer	Average fluorescence
SA, pH 4.0	10.73
SA, pH 4.5	65.53
SA, pH 5.0	150.83
SA, pH 5.5	193.40
SP, pH 5.5	109.46
SP, pH 6.0	130.27
SP, pH 6.5	116.12
SP, pH 7.0	152.70
SP, pH 7.5	195.97
SP, pH 8.0	281
T, pH 7.5	418.63
T, pH 8.0	662.10
T, pH 8.5	782.73
T, pH 9.0	212.20

As is evident from Figure 64, the two peaks of activity at pH 5.5 and 8.5 that were previously observed (Figure 63) are present again.

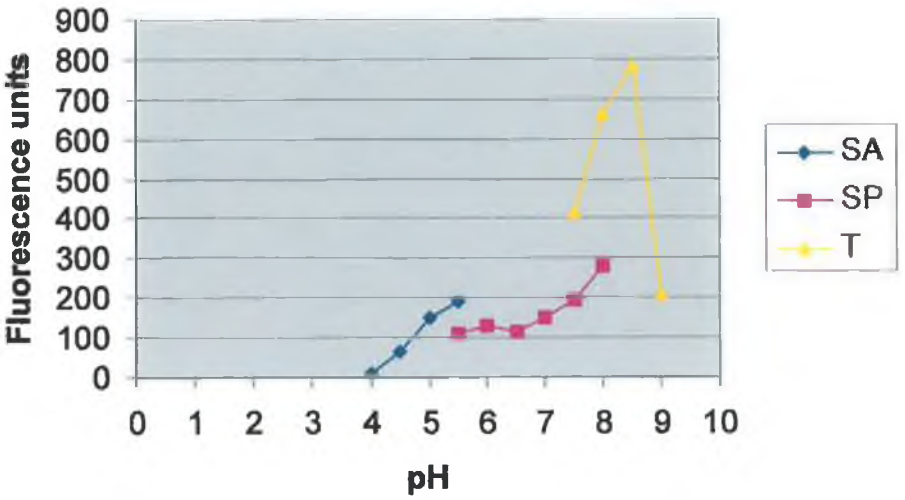


Figure 64: Graph of enzyme assay containing Z-Phe-Ala-NH_2

The assay using Z-Phe-Phe-OH was carried out at pH 5.5 and 8.5 only as they represented the pH at which highest FhCC activity was observed.

Table 57: Results of pH profile assay using Z-Phe-Phe-OH inhibitor

Buffer	Average fluorescence
SA, pH 5.5 (+ inhibitor)	599.50
SA, pH 5.5 (no inhibitor)	690.23
SP, pH 5.5 (+ inhibitor)	198.47
SP, pH 5.5 (no inhibitor)	198.27
T, pH 8.5 (+ inhibitor)	575.63
T, pH 8.5 (no inhibitor)	659.67

As Figure 65 illustrates, only a slight decrease in enzyme activity is observed in those assays that contain the inhibitor (refer to Discussion 4.5). Therefore, the enzyme with peak activity at pH 8.5 is not cathepsin L.

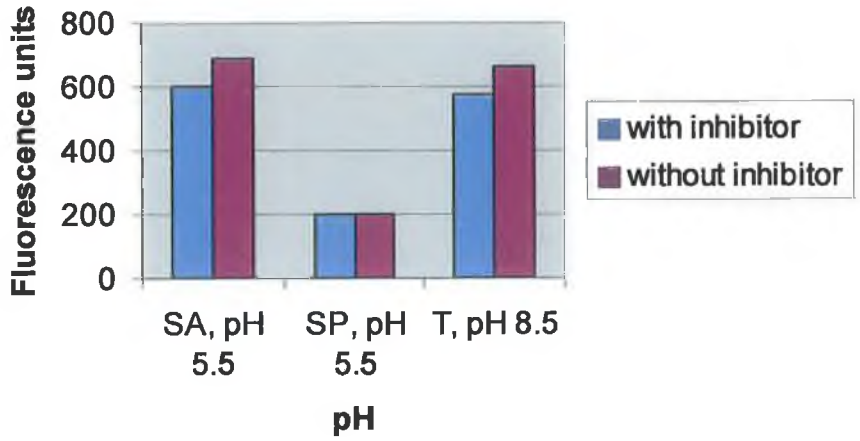


Figure 65: Bar chart of the assay carried with and without Z-Phe-Phe-OH

3.4.3: ENZYME ASSAYS USING CYSTEINE PROTEASE INHIBITORS

Assays were performed with the general cysteine protease inhibitors Iodoacetamide (Table 58), Leupeptin (Table 59) and NEM (Table 60) to examine the effect of these inhibitors on FhCC. Controls were included for comparative purposes, which were standard enzyme assays without inhibitor.

Table 58: Iodoacetamide inhibition assay results

Iodoacetamide (mM)	Fluorescence units	% inhibition
0	79.61	
0.01	76.58	3.8
0.1	68.15	14.4
1	47.56	40.3

Table 59: Leupeptin inhibition assay results

Leupeptin (mM)	Fluorescence units	% inhibition
0	605.75	
0.01	515.15	15
0.1	335.50	44.6

Table 60: NEM inhibition assay results

NEM (mM)	Fluorescence units	% inhibition
0	295.8	
1	198.33	33

Slow inhibition by the cysteine protease inhibitor E-64 is another feature of all cathepsin Cs. Other cathepsins such as cathepsin L are inhibited almost instantaneously by E-64. To investigate this enzyme assays were performed for cathepsin C and L in the presence and in the absence of 5 μM E-64. For the E-64 assays activity was expressed as a % of the total activity of the enzyme determined in the absence of the inhibitor (Table 61, Figure 66).

Table 61: E-64 inhibition assay of FhCC and FhCL

Time	FhCC% activity	FhCL% activity
0	89.58	8.87
5	90.33	8.06
10	81.35	7.84
20	74	6.89
40	62.47	6.84
80	62.13	6.11

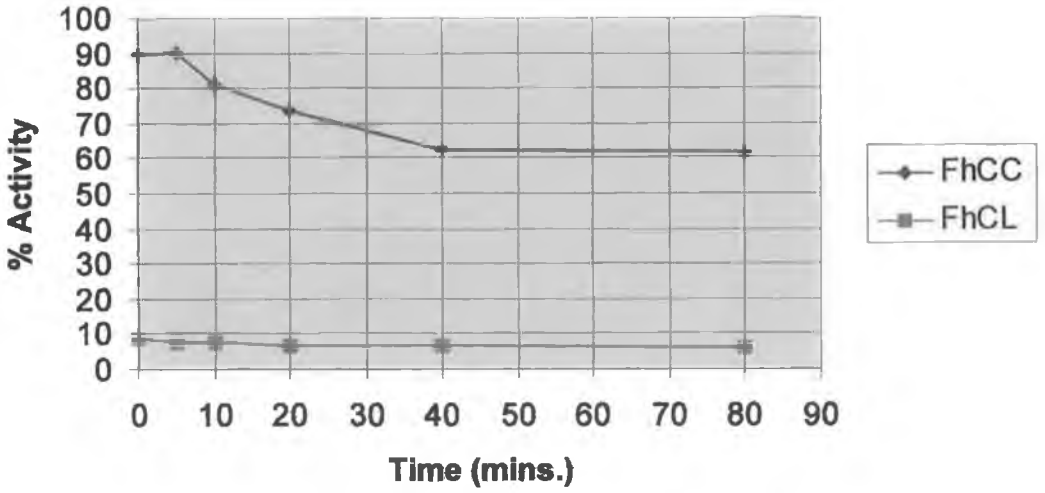


Figure 66: Graph showing E-64 inhibition assay of FhCC and FhCL

The slow inhibition of FhCC by E-64 is clearly depicted in the bar chart in Figure 66 whereas it can be seen that inhibition of cathepsin L by E-64 occurs almost instantaneously.

3.4.4: Activation of FhCC by halides

Cathepsin C activity is enhanced in the presence of halide ions. To investigate the effect of halides on the activity of FhCC, a series of assays were performed using halides at the concentrations outlined in Tables 62-65 below. Activity is expressed as a percentage of the activity of FhCC in the absence of halide ions.

Table 62: Results of activation of FhCC with NaCl

NaCl conc. (mM)	Fluorescence units	% activation
0	38.94	
0.1	39.18	0.6
1	51.45	32.12

Table 63: Results of activation of FhCC with NaBr

NaBr conc. (mM)	Fluorescence units	% activation
0	23.9	
0.1	26.97	12.8
1	27.08	13.32
10	28.55	19.44

Table 64: Results of activation of FhCC with NaI

NaI conc. (mM)	Fluorescence units	% activation
0	21.79	
0.1	25.03	14.87
1	29.72	36.36
10	31.43	44.21

Table 65: Results of activation of FhCC with NaF

NaF conc. (mM)	Fluorescence units	% activation
0	22.51	
0.1	22.20	-1.37
1	22.24	-1.18
10	21.59	-4.04

3.4.5: BCA Protein Assays

BCA protein assays were carried out on the *F. hepatica* somatic extract in order to determine the protein concentration of the extract so that the specific activity of FhCC can be calculated. A standard curve was prepared using known concentrations of BSA and the protein concentration of the somatic extract was determined from the graph using the A_{570} value obtained for the extract in the assay (Table 66, Figure 67).

Table 66: BCA Protein Assay

Name	**A_{570}nm of duplicates**
H_20 control	0.128
0.2 mg/ml BSA standard	0.482
0.4 mg/ml BSA standard	0.761
0.6 mg/ml BSA standard	1.049
0.8 mg/ml BSA standard	1.294
1.0 mg/ml BSA standard	1.915
1.5 mg/ml BSA standard	2.151
2.0 mg/ml BSA standard	2.667
Neat somatic extract	Off-scale
1/5 dilution of somatic extract	1.778
1/10 dilution of somatic extract	0.854
1/100 dilution of somatic extract	0.180

From the BSA standard curve (Figure 67):

- The 1/5 dilution of somatic extract had an A_{570} of 1.778 nm. From the standard curve the protein concentration in this sample was 0.945 mg/ml. This represented a 1/5 dilution of the extract so the total protein concentration was 4.725 mg/ml.

- The 1/10 diltuion of somatic extract had an A_{570} of 0.854 nm. From the standard curve the protein concentration was determined to be 0.56 mg/ml. Taking the dilution factor into account, the overall protein concentration was 5.6 mg/ml.

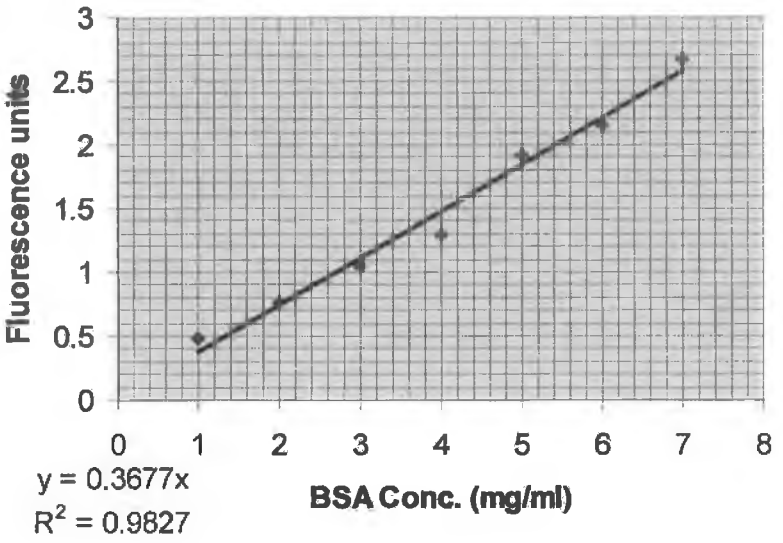


Figure 67: BSA Standard Curve

The protein concentration was used to determine the specific activity of FhCC in the somatic extract. Using the result from the DTT assay performed in Section 3.4.1, the activity of FhCC was 0.67 U. The protein concentration is 4.725 mg/ml determined from the 1/5 dilution of somatic extract. The specific activity was determined to be 0.142 U/mg of protein.

When somatic extract was prepared again from homogenised flukes it was necessary to carry out a BCA assay in order to determine the total protein concentration. The protein concentration of a somatic liver fluke extract was determined to be 16.2 mg/ml of protein.

4: DISCUSSION

4: DISCUSSION

Proteases undoubtedly play an indispensable role in the biology of parasitic organisms, playing a pivotal role in tissue penetration, digestion of host tissue for nutrition and evasion of immune responses (Tort *et al.*, 1999; Sajid and McKerrow, 2002). A number of proteinases have been implicated in Hb digestion by schistosomes including the exopeptidase cathepsin C, which is the focus of this project (refer to 1.2.6). It has been suggested that any immune response directed against the Hb digesting enzymes of schistosomes could block the digestive process and deliver a detrimental effect on the parasite (Dalton *et al.*, 1995; Brindley *et al.*, 1997).

Because of the widespread applications of gene cloning technologies, including the reliability of PCR it has been possible to clone and produce by recombinant means many of the important proteases of parasites, many of which have potential as vaccine candidates (Tort *et al.*, 1999). These technologies were exploited to the full in this project, in order to isolate, clone and express recombinant SmCC.

Schistosome cathepsin C shares many characteristics with those of mammals in that it is a lysosomal cysteine protease belonging to the papain superfamily, active at acid pH and containing an unusually long propeptide (Hola-Jamriska *et al.*, 1998; Tort *et al.*, 1999; Turk *et al.*, 2001). The full-length cDNA of *S. mansoni* cathepsin C was cloned by Butler *et al.*, 1995 and can be retrieved from the public database (accession no: Z32531). Proenzyme cathepsin C from *S. japonicum* was previously expressed in a baculovirus expression system utilising *Trichoplusia ni* (High-Five) strain host insect cells. Approximately 1 mg/L of recombinant enzyme was purified by affinity chromatography per litre of cell culture supernatant (Hola-Jamriska *et al.*, 2000). The rat cathepsin C precursor was also expressed in a baculovirus system and the yield of mature enzyme produced

by this system was 50 mg/L (Lauritzen *et al.*, 1998). *S. mansoni* cathepsin C has not been expressed in any system to date.

4.1: Isolation and PCR amplification of *S. mansoni* cathepsin C (SmCC) cDNA

Prior to cloning of the SmCC cDNA into the pPIC9K vector for expression in *Pichia pastoris*, the cDNA had to be isolated from the cDNA library. Initially primers were designed based on the Genbank cDNA sequence to amplify the pro and mature regions of SmCC, which contained the *SnaBI* and *AvrII* restriction sites for cloning into the pPIC9K vector. In initial PCRs using the SmCCF and SmCCR primers, a strong 450 bp band was seen on gel analysis of the amplified products instead of the expected 1.3 kb SmCC cDNA. This could be explained by the fact that the primers contained additional bases not specific to SmCC cDNA, such as the aforementioned restriction sites and the hexahistidine tag at the 3' end. These extra sequences could not anneal specifically to the cDNA template in the PCR and this could explain why PCR with SmCCF and SmCCR did not amplify the 1.3 kb cDNA.

"Faithful" primers were designed based solely on the Genbank sequence i.e., no recognition sites or hexahistidine tail was present. However, the 450 bp band was seen again using the new SmCCF2 and SmCCR2 primers. In order to increase the specificity of the PCR reaction various parameters of the reaction were manipulated such as the annealing temperature and $MgCl_2$ concentration but the 1.3 kb cDNA was not isolated. On one occasion a faint 1.3 kb band was visible on the gel after amplification with SmCCF2 and SmCCR2 but a very bright 450 bp band was also present (Figure 9). When this faint 1.3 kb band was gel purified and used as template in the re-amplification of the fragment with the same primers no PCR product was obtained which would indicate that the concentration and quality of the purified DNA was not conducive to the successful re-amplification of SmCC. It is inevitable that some DNA is lost during the purification process and if

the DNA concentration was low prior to purification this could explain why the expected 1.3 kb cDNA was not amplified. Another explanation is that inhibitory reagents of PCR may have been introduced during the purification process thus inhibiting the action of *Taq* polymerase.

It was thought that perhaps the cDNA library was degraded or defective or that the end of the cDNA was folded in such a manner that it was impossible for the reverse primer to anneal. Alternatively, perhaps the full-length SmCC cDNA was absent from the library. To investigate whether this was the case a set of primers, SmCCMF and SmCCMR were designed, based on the sequence from the middle of the SmCC cDNA, which when used with SmCCR2 and SmCCF2 respectively, would amplify cDNA fragments of 728 bp and 601 bp (refer to Figure 7). Gel analysis of the PCR products from amplification with these primers showed that these fragments were produced, each comprising "half" of the SmCC cDNA, thereby verifying the presence of the full-length cDNA in the library (Figure 10). A new reverse primer, SmCCR3, was designed that was moved down several bases past the SmCC stop site towards the 3' end of the cDNA (Figure 7, Table 2 and 3). PCR using this primer along with SmCCF2 still resulted in the amplification of the small, 450 bp product.

Another strategy employed involved the use of a primer specific for the plasmid pB42AD together with an SmCC specific primer as this was the vector used in the construction of the cDNA library. As the exact orientation of the library in the vector was unknown SmCCF2 was used with both pFor and pRev primers. The "middle" primer SmCCMF was also used with both vector specific primers (Figure 11). Gel analysis of PCR products revealed the presence of bands at 900 bp, 350 bp and 150 bp on amplification with pRev and SmCCF2. The 900 bp band was discovered to be a vector specific band. The 1.3 kb SmCC cDNA was not amplified.

New primers were designed, called SmCCF4 and SmCCR4. The forward primer was designed to anneal to a region of the signal peptide of SmCC

cDNA, whereas the reverse primer was designed to anneal to the 3' end twenty bases past the stop site. PCR amplification with these primers produced a 480 bp product only.

The obstacle to the successful amplification of the full length SmCC cDNA appeared to be the reverse primer. A convoluted secondary structure, or missing bases at the 3'end could have been an obstacle to the reverse primer annealing. Alternatively, it was possible that the reverse primer annealed properly to its complementary sequence on the template cDNA but was unable to complete the synthesis of the full-length cDNA, thus giving rise to a truncated smaller product.

To investigate this possibility it was necessary to design another reverse primer, SmCCR5. This primer was designed to anneal at bases 1351-1380 on the cDNA sequence, which was 5' to the stop site. The stop site (base 1382) was not amplified. PCR amplification with either SmCCF2 or SmCCF4 with this reverse primer resulted in the full-length SmCC cDNA being produced. This indicated that a problem had indeed existed with the 3' end of the SmCC cDNA past the stop site. This was demonstrated by PCR. It was seen that when the reverse primer was designed to anneal seven bases past the stop site of SmCC, as is the case for SmCCR and SmCCR2, the full-length cDNA could not be amplified. The same was true when SmCCF4 was used which annealed eighteen bases after the stop site (3' end). However, by designing the primer to anneal before the stop site as a result of moving the primer closer to the 5' end of the cDNA, the full-length product was amplified. Because it was vital that the stop site was present in the amplified cDNA, it was necessary to design another primer, SmCCR6, which annealed just four bases past the stop site. Amplification with this primer proved successful and also illustrated that whatever problem occurred at the 3'end of the SmCC cDNA occurred after base 1386. The 1.3 kb PCR product amplified with SmCCF2 and SmCCR6 was gel purified and concentrated.

To increase the concentration of DNA present to facilitate the successful re-amplification of SmCCF with SmCCR primers, the purified SmCC cDNA was cloned into the pGEM vector. After successful transformation of *E. coli* cells with the vector, the presence of the insert in isolated, purified pGEM was confirmed by size comparison between the vector containing the insert and pGEM vector alone, by performing a *PstI* digest on the clones, and by PCR analysis using both pGEM and SmCC specific primers. Isolated plasmid DNA confirmed to contain SmCC was used as template for the amplification of the cDNA using SmCCF and SmCCR primers containing *SnaBI* and *AvrII* restriction sites for cloning. The resultant 1.3 kb product was purified for use in the next cloning step into the pPIC9K vector.

4.2: The 450 bp fragment

Owing to the fact that the 450 bp PCR product appeared so frequently with so many different primers during the amplification process, much curiosity existed about it. It was decided to sequence this fragment after it was successfully cloned into pGEM to determine whether it was an artifact of the cDNA library, a smaller, truncated version of SmCC, or perhaps another previously undiscovered or related gene with sequence similarity to SmCC. Subsequent alignment of the amino acid sequence of the 450 bp fragment with that of SmCC revealed a 99.2% identity of this sequence with amino acids 20-148 of SmCC. Also the last nine amino acids of the 450 bp fragment were 100% identical to the last nine bases of SmCC. This proved conclusively that the 450 bp fragment was a smaller, truncated version of the SmCC DNA. The region of SmCC from amino acids 128-447 was missing from the 450 bp fragment. A BLAST was performed using all reverse primers against SmCC cDNA to investigate whether these primers bound to any other region of the SmCC cDNA but no alternative binding sites were revealed for any of the reverse primers (Atschul *et al.*, 1990). It is possible that non-specific binding of the reverse primer to another site on the SmCC cDNA did occur but this is unlikely because of the fact that the 450 bp fragment was seen to contain 9 amino acids from the end of SmCC, which is

indicative of the reverse primer successfully binding to the end of the cDNA. Another theory is that the primer could be defective and may not be able to amplify the full-length cDNA. The probability that a primer will not be fully extended is dependent on distance, secondary structure, the extent of DNA degradation in the template, enzyme limitation and the time allowed for polymerase extension (Ehrlich *et al.*, 1991). Any of these factors could explain why the truncated SmCC 450 bp fragment was produced. In conclusion, the 450 bp product was probably produced together with the full-length cDNA but because of the increased sensitivity of PCR to the amplification of smaller fragments, the 450 bp fragment appeared as a strong band on gel analysis, whereas the 1.3 kb band was faint (Figure 9).

4.3: Cloning of the SmCC cDNA into pPIC9K vector and sequence analysis

The next stage in this project involved cloning the SmCC fragment into the pPIC9K vector. Initial transformation efficiencies proved low, and analysis of transformants by PCR using the pPIC9K specific AOX3' and α-factor primers revealed that the 1.3 kb insert was not contained in the vector. Only the 195 bp band characteristic of pPIC9K was visible on gel analysis of the PCR products instead of the 1.5 kb band expected if pPIC9K contained the insert. A confirmatory *PstI* digest did not reveal the presence of the 1.3 kb cDNA in the vector (Figure 21). It was expected that if SmCC was contained in pPIC9K, a larger fragment of 4.9 kb would replace the 3.6 kb fragment expected when the enzyme cuts pPIC9K. For 6 of the 8 clones analysed in this manner, the aforementioned 3.6 kb fragment was produced on digestion in addition to the 4.9 kb fragment. This could be indicative of the presence of mixed plasmids i.e., some plasmids contained the SmCC cDNA while others did not. Further evidence for the mixed plasmid theory came from PCR analysis of plasmids isolated from a different transformation (Figure 22). Gel analysis of a PCR using AOX3' and α-factor primers revealed the presence of a 1.5 kb band for 2 of the 3 clones analysed, which suggests that the plasmid contained the insert. However, the 195 bp band of pPIC9K alone

was also visible, which should not be the case. These conflicting results could be explained by the mixed population theory or alternatively could indicate that the PCR was contaminated with pPIC9K. The later is likely because the negative control for this PCR, which contained no template DNA revealed pPIC9K contamination. A double digest with *SnaBI* and *AvrII* was performed on these clones to liberate the 1.3 kb cDNA from the vector but on gel analysis only the 9.3 kb pPIC9K band was seen.

Prior to initiating any further transformations the efficiency of the overnight ligation step was ascertained. Only those ligation reactions that did not show a 1.3 kb SmCC band on gel analysis were used for transformation because the absence of this insert suggested that it had been successfully ligated to the vector (Figure 23). It was decided to screen the subsequent transformants by restriction digestion because PCR preferentially amplifies smaller pieces of DNA, and if a mixed colony was used, it was thought that the smaller 195 bp DNA would be always preferentially amplified at the expense of the larger SmCC cDNA. However, restriction enzyme digests performed using *KpnI*, *StuI*, *XbaI* and *XhoI* were incomplete even after 3 hrs, at 37^0C (Table 32). Plasmid DNA wasn't cut either, which indicates that the problem lay with the enzymes themselves. An explanation for this could be that traces of phenol, chloroform, detergents or other inhibitory agents remained after isolation of the plasmid DNA by alkaline lysis. Methylation of plasmid DNA could also account for the inactivity of the enzymes. Additionally, compared to linear DNA, intact bacterial plasmids often require a higher unit quantity of restriction enzyme in order to cut the DNA owing to the fact that it is in a supercoiled, covalently closed circular form. Perhaps the activity of the enzymes employed had declined over time so that the activity stated by the manufacturer did not represent the true activity of the enzyme and consequently was not sufficient to complete the digestion (New England Biolabs and Promega manuals).

Transformants from plate B2 were selected for PCR analysis using the SmCC specific SmCCF2 and SmCCMR primers (Figure 24a and b). PCR analysis

of eleven isolated plasmids revealed that five contained SmCC cDNA. However, gel analysis also revealed the presence of bands < 250 bp, which could represent the 195 bp bands of pPIC9K even for those plasmids shown to contain SmCC cDNA. This finding provides further support for the mixed population theory. These small bands appeared brighter for those plasmids that did not have SmCC cDNA.

Size comparison of *NotI* linearised plasmids on a gel with pPIC9K alone proved inconclusive (Figure 25). It was difficult to ascertain precisely from the gel whether the plasmids contained the insert because although on inspection it appeared that some plasmids were larger than pPIC9K alone this could be attributed to the fact that may have contained more concentrated DNA than pPIC9K and thus produced stronger, brighter bands.

Another detection strategy employed involved carrying out PCR using an SmCC specific primer in conjunction with a pPIC9K specific one. This would determine whether the SmCC cDNA detected in the previous PCR was actually contained in the vector and not the result of insert carry-over from the ligation procedure. Three out of ten clones, namely clones 4, 5 and 8 produced the anticipated 664 bp product when screened with α-factor and SmCCMR primers (Figure 26). Only clones 4 and 8 proved positive for SmCC when screened with SmCCF2 and SmCCMR primers. The same clones produced the expected 814 bp product when they were screened with AOX3' and SmCCMF primers (Figure 27). For those clones which did not reveal the 814 bp band on gel analysis a band at 1.4 kb was visible, which arose as a result of the SmCCMF binding non-specifically to a site on the vector. Plate B1 was screened with SmCCMF2 and SmCCMR. The presence of a 200 bp band was indicative of a positive result for SmCC and half of the clones screened were seen to be positive (Figure 28). The SmCC positive plasmids from this PCR were used to transform *E. coli*. Transformation proved very efficient in this instance and clones transformed with plasmid # 2 were selected for further analysis. On PCR analysis some isolated plasmids proved positive for SmCC while others did not which reinforces the

mixed population theory. Restriction digestion with *SnaBI* and *AvrII* showed that the plasmids originating from transformation plates B1 and B2 did not contain the 1.3 kb SmCC cDNA.

All the evidence supported the notion that a mixed plasmid population was present.

Because of the difficulties experienced in obtaining a pure culture, containing plasmids with SmCC, it was decided to transform again with SmCC cDNA. Transformation efficiencies proved high in this instance (Table 34). Purified plasmids from the transformation were analysed on a gel. One of the plasmids on size comparison, referred to as plasmid # 4 , proved to be larger than the other plasmids and the pPIC9K control (Figure 30). Further analysis of this plasmid by PCR using the SmCCF2, SmCCMR (Figure 31) and α-factor, SmCCMR (Figure 32) primers confirmed the SmCC cDNA to be present in the vector. Restriction digestion with *SnaBI* and *AvrII* liberated the 1.3 kb SmCC cDNA from the 9.3 kb pPIC9K (Figure 33). The appearance of a distinct 1.5 kb band upon gel analysis of the products of a PCR carried out with AOX3' and α-factor primers proved conclusively that SmCC was contained in pPIC9K (Figure 34b). No pPIC9K 195 bp band was visible as was expected if the vector contained the SmCC fragment.

Sequence Analysis

Sequencing was carried out on plasmid # 4 using the sequencing primers detailed in 3.2.2 (Figure 35). The deduced amino acid sequence of this SmCC consisted of a propeptide of 198 residues and a mature enzyme sequence of 237 residues. The signal peptide of 19 amino acids was not cloned. When the SmCC amino acid sequence of the propeptide and mature regions were aligned with SmCC from the Genbank 412 amino acids out of 436 were seen to be 94.5% identical.

Using BLAST, SmCC (pro and mature) was found to 47.99% identical to human dipeptidyl peptidase I (DPPI), 48.44% identical to mouse DPPI and 48.2% identical to rat DPPI (Table 35). The deduced SmCC sequence exhibited 21.34% identity with SmCB and 21.87% identity with SmCL at the amino acid level. As anticipated active site residues Cys 30, His 181 and Asn 203 were conserved among species, as were the residues in proximity to the active site.

When SmCC was aligned with three mammalian cathepsin Cs from human, mouse and rat, Genbank SmCC and SjCC (see Appendix) the following points were noted:

- The nine amino acid residues RQWKCFVAE, (residues 102-110 in Figure 37), which are not found in the Genbank SmCC are found in other species. For instance, arginine (R) is conserved in rat, mouse and human cathepsin C, glutamine (Q) is conserved in SjCC, while tryptophan (W), cysteine (C) and phenylalanine (F) are all found in rat, mouse, human and SjCCs. Valine (V) is conserved in rat and mouse CC only. Alanine (A) is found in SjCC and the additional glutamic acid (E) appears to be unique to the sequenced SmCC, it is not conserved in any other species examined (Figure 38).

- The amino acid sequence FVSENCD, located at position 287-293 (Figure 37) in SmCC was different to that in the Genbank (Figure 39). E was the only residue that was conserved in all species but not the Genbank SmCC.

Since the publication of the SmCC sequence in 1995, advances have been made which improve the efficiency and accuracy of DNA sequencing. The amino acid differences between the SmCC used in this project and that in the Genbank are conserved among many species so it is likely that the sequence of the SmCC used in this project represents the correct sequence of SmCC.

Only 60.78% identity was observed on alignment of the pro and mature regions of SmCC with that of SjCC. On alignment of the Genbank SmCC with SjCC, 56% identity was seen. It was expected that SmCC would exhibit a higher identity with SjCC and this implies that SmCC and SjCC may be discrete enzymes rather than species homologues (Hola-Jamriska *et al.*, 1998).

The tyrosine residue adjacent to the active site Cys 30, which is unique to all cathepsin C members of the papain superfamily, was conserved in SmCC. It is believed that this acidic residue may play an important role in determining substrate specificity by being involved in the binding of the amino-terminus of the peptide substrate and an isoleucine at the P2 side chain binding cleft (McGuire *et al.*, 1997; Hola-Jamriska *et al.*, 1998).

At position 71 in the mature enzyme, an uncharged polar amino acid, asparagine (N) is found in SmCC and its Genbank counterpart. Mouse, rat and SjCCs all possess an acidic residue, aspartic acid (D) at this position. Additionally, an isoleucine (I) residue at position 229 in the mature enzyme, found in all mammalian cathepsin Cs and SjCC was replaced by a leucine (L) in SmCC. This isoleucine residue is believed to interact with the side group of the P_2 residue of the substrate. These key differences between SmCC and SjCC and mammalian cathepsin Cs could be indicative of differences in substrate specificity and function (Hola-Jamriska *et al.*, 1998).

In mammalian cathepsin Cs it is thought that a conserved sequence of the prosegment, TA/DEIQQQ/KIL, situated in proximity to the cleavage site of the pro and mature regions, regulates the processing of the mature enzyme from the inactive zymogen. This sequence is not conserved in cathepsin C from schistosomes (Hola-Jamriska *et al.*, 1998). Also, a conserved sequence in mammalian cathepsin Cs, GLS/RD adjacent to the site where the enzyme is cleaved during conversion to the two-chain form was not conserved in schistosomes (Dolenc *et al.*, 1995; McGuire *et al.*, 1997). A unique insert was seen in schistosomes at this site; NNH in SjCC and TDH in the Genbank

SmCC. For the SmCC used in this project, the sequence at this location was NDH.

N-Glycosylation

The four potential N-glycosylation sites, two in the proregion and two in the mature enzyme region of the Genbank sequence are conserved in the deduced SmCC (Butler *et al.*, 1995). Three potential N-glycosylation sites are present in the three reported mammalian cathepsin Cs, one in the mature enzyme and the other two in the prosegment. The glycosylation site in the mature region is conserved in SmCC but is absent from SjCC. Because the recognition sites for targeting molecules to the lysosome are present on the carbohydrate moieties it is believed that SmCC and SjCC may have different cellular locations with Sj being secreted (Hola-Jamriska *et al.*, 1998).

Disulfide Bonds

As for the Genbank SmCC, Cys 27 is capable of forming a disulfide bond with Cys 70 and Cys 63 can form a bond with Cys 104. Cys 109 is unpaired as is the active site Cys.

Prosegment

Owing to the fact that cysteine proteases are initially synthesised as inactive zymogens, the proregion plays a number of vital roles. It can serves as a signal sequence, an endogenous inhibitor and also an intramolecular chaperone (Sajid and McKerrow, 2002). The unusually long prosegment characteristic of cathepsin C members of the papain superfamily was evident in SmCC. Alignment of the prosegments of the schistosome and mammalian cathepsin Cs revealed a high level of conservation; of 209 residues 49 were absolutely conserved and 68 were >50% conserved. The long prosegment was attributed to an extension at the NH_2 terminus. This extension contained

two blocks of residues highly conserved among the cathepsin Cs. These two blocks were conserved within the SmCC used in this project. At the COOH terminus of the proregion two conserved blocks are present which share structural similarity with cathepsins B and L. The prosegment of SmCC contained ERFNIN-like motifs, which are characteristic of cathepsin L. This was also observed for SjCC (Hola-Jamriska *et al.*, 1998).

Phylogenetic Analysis

Phylogenetic analysis of all the genes of members of the papain superfamily in the public database revealed that the cathepsin C genes cluster with the cathepsin Bs (Hola-Jamriska *et al.*, 1998). Controversy exists with regards to the ancestry of cathepsin C. Due to the fact that cathepsin C possesses a simple exon/intron structure in comparison to the complex genes of cathepsin B, H and L, it has been suggested that cathepsin C may not share a common ancestry with cathepsin B but instead perhaps arose through convergent evolution (Rao *et al.*, 1997). However, murine DPPI was demonstrated to have a complex genomic structure. It contains seven exons and spans more than 20 kb. Cathepsin B and H also span more than 20 kb, comprising 10 and 12 exons respectively. Cathepsins C, H and L all share a common insertion position that isolates the exon containing the active site residue. In cathepsins B, C, H and L the active site histidine and asparagines residues are found within one exon although the exon on which they reside and the positions of the intron-exon insertion sites differ from gene to gene. These similarities indicate that the four proteases arose from a common ancestral gene, but that their structures have evolved to include intron losses and insertions (Pham *et al.*, 1997). Phylogenetic analysis of SmCC and SjCC and three mammalian species placed cathepsin C firmly on a clade with the cathepsin B group. Given the common ancestry of cathepsin C and B it is unusual that SmCC and SjCC contain the ERFNIN motifs mentioned earlier in their prosegments as they are characteristic of cathepsin L proteases (Hola-Jamriska *et al.*, 1998).

Restriction Sites

Prior to transformation of *P. pastoris* with SmCC, one of the restriction enzymes *SalI*, *SacI* or *BglII* must be used to linearise pPIC9K containing the insert. The enzyme of choice should cut pPIC9K but not SmCC. Restriction site analysis was performed on the Genbank SmCC and none of the enzymes were shown to cut the cDNA. However, restriction site analysis of the sequenced SmCC revealed that base changes had occurred within the sequence, which resulted in the production of both a *SalI* and a *SacI* site. This meant that only *BglII* could be used to linearise the expression vector prior to transformation of pPIC9K.

TIGR Gene Indices

The TIGR gene indices represent a public database of Expressed Sequence Tags (EST). ESTs have provided a first glimpse of the collection of transcribed genes for a variety of organisms. Through careful analysis of this sequence data additional functional, structural and evolutionary information can be provided. Gene indices are constructed for selected organisms by first clustering, then assembling ESTs and annotated gene sequences from the Genbank, thereby producing a set of unique high fidelity virtual transcripts or tentative consensus sequences (Quackenbush *et al.*, 2000). TC 4520 from the TIGR database revealed 97.48% identity to the Genbank SmCC and 96.58% identity to the SmCC used in this project at amino acid level (http://www.tigr.org).

Secondary Structure Prediction

A predicted secondary structure scan was carried out on the deduced SmCC protein sequence (refer to Appendix). It was predicted that 42.5% of the amino acids of SmCC could assume a random coil conformation, while 19.31% of amino acids are likely adopt alpha helical structures. A further

26.9% of the amino acids were expected to form extended strands and 11.26% of the amino acids were predicted to form beta turns.

4.4: FUNCTIONAL EXPRESSION OF SmCC IN *PICHIA PASTORIS*

Transformation of *P. pastoris* was carried out by spheroplasting instead of the electroporation method because it is considered the most effective technique for the generation of multi-copy inserts. Transformation proved successful when 5μg of SmCC/pPIC9K DNA was used. Transformants grew only on 0.25 mg/ml G418, which means that it is likely that only one copy of SmCC was integrated into the *P. pastoris* genome as the level of G418 resistance is proportional to the number of plasmids integrated. PCR using AOX5' and AOX3' primers showed that the SmCC cDNA was successfully integrated into the *P. pastoris* genome for clones A2 and C2 as the anticipated band at 1.8 kb was present for these clones (Figure 41). Clone B2, which represents *P. pastoris* transformed with pPIC9K reveals the 2.1 kb band characteristic of the GS115 *P. pastoris* strain as expected.

Induction # 1

An initial 16-18 hr growth period was recommended to achieve an OD_{600} 2.0-6.0 prior to induction. However, it was found that this could only be reached after 41 hrs incubation. It was later discovered that the spectrophotometer used was inaccurate which explained why such a long growth period was required to attain the recommended cell density. This error was not discovered until induction # 5. Therefore, for all inductions prior to this an accurate determination of cell density could not be made.

On SDS-PAGE analysis of D3-D6 supernatants from this induction a distinct 22 kDa band was visible, which was not seen for the pPIC9K control samples (B2). This 22 kDa could be representative of the mature enzyme as the predicted molecular size of the mature enzyme is 26.5 kDa. This band also appeared on SDS-PAGE of D6 supernatants after incubation on Ni-NTA

beads. This indicated that an intact hexahistidine tag was present on the recombinant protein. If this 22 kDa protein was a yeast protein it would not possess this tag. However, no hexahistidine tag was detected by western blot analysis.

Enzyme activity assays were performed which revealed that the SmCC specific substrate H-Gly-Arg-NHMec was cleaved using A2 and B2 supernatants. However, using B2 supernatants, the substrate was also cleaved. This should not be the case as this is the control containing *P. pastoris* transformed with pPIC9K without SmCC cDNA. This activity was attributed to non-specific cleavage by yeast proteases. Enzyme assays carried out using culture pellets revealed low SmCC specific activity. Due to the fact that no D0 sample was available for this induction it was impossible to determine whether this activity was due to non-specific substrate cleavage by yeast proteases or whether it was due to SmCC.

Induction # 2

Although this was a scale-up of induction # 1, the 22kDa band from the previous induction was not seen. Western blot analysis did not detect the hexahistidine tag. Affinity chromatography was carried out on the D6 supernatant. Determination of the protein concentration of the column fraction revealed that the protein concentration of the supernatant before application to the column (0.896 mg/ml) was virtually the same as that of the run through sample (0.884 mg/ml), which indicated that very little, if any protein bound to the column. This means that either the hexahistidine tag is not present, i.e., the protein detected was not SmCC or alternatively, it may mean that the tag was degraded or folded in such a manner that it was not exposed and therefore could not bind to the Ni-NTA beads or the anti-His_6 antibodies.

Enzyme activity for the culture supernatants was the same as that detected for the D0 sample, which indicates non-specific hydrolysis of the substrate

(Table 41a). A pH profile assay was employed to investigate activity at pH 3.0-8.5. Highest activity was observed at pH 7.5 (Table 42c, Figure 46). However, the background control used for this assay was prepared in sodium acetate pH 5.5. When this control was prepared in sodium phosphate at pH 7.5 unusually high levels of substrate hydrolysis were observed relative to that of the control prepared in sodium acetate. As a result the high activity observed at pH 7.5 cannot be attributed to the change in pH but was a result of intrinsic properties of the buffer, which lead to non-specific cleavage of the substrate.

Pellets were disrupted and were analysed by SDS-PAGE to detect any intracellular SmCC that may not have been secreted (Figure 45). A band was observed at 40 kDa in D1-D6 pellets that was not present at D0. Therefore this band represented an induced protein, which could be SmCC. However, the enzyme activity observed using the pellets was the result of non-specific cleavage of H-Gly-Arg-NHMec and not due to rSmCC activity.

Induction # 3

Induction # 3 was performed as for # 1 for seven days. SDS-PAGE analysis of supernatants revealed the presence of bands at 200 kDa and analysis of disrupted pellets did not reveal the presence of any induced bands. Western blot and ELISA analysis of supernatants and pellets did not reveal any binding (Table 43). Enzyme activity assays were carried out using D0-D7 supernatants and no increase in activity was observed between the D0-D7 assays (Table 44a, Figure 47). B2 control supernatants exhibited similar activity to A2 and C2, which proves that the low activity observed was not due to rSmCC activity. A similar assay was performed using disrupted pellets (Table 45a, Figure 48). Using A2, B2 and C2, activity was seen to decline after D0, but increased after D3, reaching a peak at D6. Activity declined after D6. However, the observed activity could not be accredited to rSmCC because similar levels of substrate hydrolysis were seen for B2 controls as for A2 and C2.

Induction # 4

Initial growth was performed in a 500 ml volume for this scale-up induction. Although an accurate determination of cell density could not be made for reasons mentioned previously, the cell density was, by inspection, very high. The effect of feeding methanol twice daily to a final concentration of 1.5% was investigated. Culture supernatants were initially analysed by SDS-PAGE (Figure 49). A non-specific 40 kDa band was seen at D0. A strong band at 47 kDa was seen at D1-D7. A band at 58 kDa was visible at D1-D7 but became fainter after D3. Incidentally, at D4 a band appeared at 55 kDa that was absent at D0-D3 but present at D5-D7. As the predicted molecular size of the SmCC proenzyme is 49.5 kDa, it is likely that these bands >50 kDa could represent differentially glycosylated forms of the enzyme. These bands resolved by SDS-PAGE even after the supernatants had been incubated with Ni-NTA beads, which indicated an intact hexahistidine tag. Purified baculovirus expressed SjCC resolved as three forms in SDS-PAGE gels, of 55, 39 and 38 kDa. It was thought that the 39 and 38 kDa polypeptides were derived from the larger 55 kDa polypeptide by differential processing of the NH_2 –terminus during expression of the proenzmye or during purification (Hola-Jamriska *et al.*, 2000). The mature enzyme has a predicted molecular mass of 25 kDa. Faint bands were visible at ~ 25 kDa on the gel, which could account for some of the SmCC specific activity observed. Degradation products were visible for all supernatants at < 6.5 kDa on the gel. These degradation products could indicate that processing of the proenzyme occurred or could demonstrate that yeast proteases were active in the media.

In mammalian cathepsin Cs part of the N-terminal portion of the propeptide is retained in the mature enzyme (Dolenc *et al.*, 1995; Turk *et al.*, 2001; Horn *et al.*, 2002). Referred to as the residual propart, this unique peptide contributes to the tetrahedral structure and creates an extension of the active site cleft, providing features that endow cathepsin C with dipeptidyl aminopeptidase activity (Turk *et al.*, 2001). The fact that insect-cell expressed SjCC was functionally active while retaining part of its proregion

provides proof that this residual propart is also retained in schistosomes (Hola-Jamriska *et al.*, 2000). In that case, the mature enzyme would be expected to have a larger molecular mass than the predicted mass of 25 kDa. Therefore, the bands seen at >47 kDa on the gel could be representative of the mature enzyme. For example, in human cathepsin C, the monomeric proenzyme of 55 kDa was discovered to be only slightly larger than the 53 kDa monomeric form of the mature enzyme (Dolenc *et al.*, 1995).

Two western blots were performed using anti-His_6 and anti-Schistosome mouse sera as the primary antibodies. Despite the promising result obtained using the Ni-NTA beads, the hexahistidine tag was not detected for any of the supernatants analysed. No binding was observed with the anti-schistosome mouse sera. The positive controls worked well for both cases. A western blot and an ELISA were performed using anti-SjCC rabbit sera (Figure 51 a, Table 46). Using D3 and D5 supernatants, antibodies bound to a band at 47 kDa. This shows that Sj antibodies cross-react with SmCC. The ELISA showed that binding to the purified SjCC controls was minimal and no binding was observed with D0-D7 supernatants. This could indicate that the 47 kDa band seen for the western blot could possibly be due to non-specific binding but the fact that a band of similar size appears on SDS-PAGE of D1-D7 supernatants increases the likelihood of this band being rSmCC.

Enzyme assays were performed on the D0-D5 supernatants (Table 47b, Figure 52). B2 supernatants served as a control in this assay. Low levels of non-specific substrate cleavage were observed for B2 supernatants. Using D0-D5 C2 supernatants, activity was detected, increasing steadily from D0-D5. The increase in enzyme activity after D4 coincides with the appearance of the aforementioned 55 kDa band at D4 on SDS-PAGE. However, activity was still low, D4 supernatants possessed 0.6 U of enzyme activity whereas D5 supernatants had 1.5 U of activity, which is low considering that the assay was performed overnight. Time was not taken into account when calculating enzyme activity in this instance. The positive control, FhCC had

9 U of activity for the overnight assay. The change in activity over a 4 hr and 20 hr period was investigated. For the 4 hr assay activity peaked at D5 (0.5 U), then declined slightly at D6, and then peaked again at D7. For the 20 hr assay activity was seen to peak at D5 (1.7 U) and declined for subsequent days. As 1 U of enzyme activity is a function of time (min), although an increase in the amount of NHMec liberated over time was observed, this does not represent an increase in activity of the enzyme because the assay was performed over a longer time-scale.

A leupeptin inhibition assay was carried out using the D5 supernatant (Table 49, Figure 54). A 90.4% decline in enzyme activity was observed in assays performed in the presence of leupeptin. For the positive control, leupeptin inhibited FhCC in somatic extract by only 42%. A similar assay performed using iodoacetamide revealed it to be only a weak inhibitor of rSmCC, inhibiting the enzyme by only 9.32%. Iodoacetamide inhibited FhCC by 94.6%. Since *F. hepatica* somatic extract was used as the source of FhCC instead of purified enzyme and because iodoacetamide is a general cysteine protease inhibitor, it is likely that other cysteine proteases, capable of cleaving the cathepsin C substrate were also inhibited in the extract. This finding questions the reliability of using *F. hepatica* somatic extract as a positive control.

Another assay was undertaken, to discover the extent to which rFhCL1 cleaves the cathepsin C substrate H-Gly-Arg-NHMec (Table 50). As the bar chart in Figure 55 illustrates, although some non-specific cleavage occurs, the activity of FhCL1 towards H-Gly-Arg-NHMec is low relative to that of rSmCC.

These findings for induction # 4 indicate that although rSmCC appeared to be produced, production was by no means optimal. By increasing the cell density of the starting culture and by increasing the concentration of methanol in the induction flasks, the production of rSmCC was enhanced. As the cells are all of the slow-growing Mut^S phenotype, such a high cell

density may be a necessity in order for a recombinant protein such as SmCC to be functionally expressed.

Induction # 5

Induction was performed at pH 6.0, 7.0 and 8.0 to investigate the effect of pH on rSmCC expression. Cultures were grown to an OD of 20.0 prior to induction. The pH in the induction flasks was monitored over the 6-day induction period and it was observed that pH increased by 1.5 pH units over the course of the induction (Table 51). Resolution of D0-D6 supernatants by SDS-PAGE revealed the presence of a single band for pH 6.0 supernatants from D3-D6 at 55 kDa. It is possible that this single band represents the mature enzyme with the residual prosegment attached For D3-D6 pH 7.0 supernatants, three distinct bands are resolved at 60 kDa, 55 kDa and 47 kDa, which are similar to those bands observed on analysis of induction # 4 supernatants. Low molecular mass degradation products can be seen at pH 7.0 but not pH 6.0. No bands were visible for any of the supernatants prior to D3.

Western blot analysis using the anti-His_6 antibodies did not detect the presence of the his tag. This provides further evidence that the his tag is damaged or that the secondary structure of the protein obscures the his tag. This will add to the complexity of subsequent purification steps. A western blot employing the anti-SjCC sera revealed that binding occurred to a 62 kDa and 28 kDa polypeptide for D4 pH 6.0 and D4 pH 7.0 supernatants. As these bands are not visible following resolution by SDS-PAGE for either sample, it is likely that they are due to non-specific antibody binding.

Enzyme activity assays were carried out on all supernatants (Table 52, Figure 58). Activity was low in all cases compared to that observed for supernatants from previous inductions, peaking at D3 for pH 5.0 supernatants, D4 for pH 6.0 and D5 for pH 7.0. Surprisingly, pH 5.0

supernatants exhibited the highest activity of the three. This is unusual because no bands were visible on SDS gel analysis of these supernatants.

Induction # 6

Cells were grown to an OD_{600} of 3.0 prior to induction. This induction served as a comparison to induction # 4, in which cells were initially grown to a very high density. All other conditions were as for induction #4. After 5 days induction, supernatants were resolved by SDS-PAGE. A single band of 47 kDa was observed at D3 and D4 (Figure 59). This is an encouraging result, because it signifies that the pro-mature enzyme is being produced and no degradation of the product occurs. Perhaps this band represents the mature enzyme with the residual propart. This can only be proven by N-terminal sequencing of the protein.

Western blot analysis revealed no binding with anti-His_6 and anti-SjCC sera. Surprisingly, the positive control for the blot, SmCB was shown to cross-react with SjCC.

Activity in the supernatants was low in all instances (Table 53, Figure 60). Activity peaked at D4 and when results from an overnight assay were compared with an assay carried out over a 4hr period very little difference in activity between the two was seen (Figure 61).

4.5: Biochemical Characterisation of *F. hepatica* Cathepsin C

Schistosome dipeptidyl peptidase I (DPPI) or cathepsin C was characterised by Hola-Jamriska *et al.*, 1999. In this study, the optimum pH for cathepsin C activity was discovered to be pH 5.5. Activity was enhanced by reducing conditions and by addition of the halide ions Cl^-, Br^- and I^-. Schistosome cathepsin C was only slowly inhibited by E-64. Iodoacetamide, leupeptin and NEM, general cysteine protease inhibitors, which as thiol blockers

inhibit the highly reactive thiol group that characterises the cysteine protease, rapidly blocked the schistosome's DPPI activity. Therefore, schistosome DPPI exhibited similar characteristics to mammalian cathepsins (refer to Section 3.4).

Most of the enzymes characterised from *F. hepatica* have been endoproteinases. Previously, a dipeptidyl peptidase was characterised from *F. hepatica* but it did not cleave the DPPI substrate H-Gly-Arg-NHMec. This enzyme cleaved DPPII and DPPIV substrates but their respective inhibitors, puromycin and bactitracin did not have any effect on the activity of the enzyme (Carmona *et al.*, 1993).

Given the biochemical characteristics of schistosome DPPI, a series of enzyme assays were performed to determine if FhCC shared these characteristics and to investigate if FhCC possessed any novel or original attributes.

Initial assays were carried out to investigate whether FhCC, like its schistosome counterpart, was enhanced in the presence of DTT. DTT is a reducing agent that reduces the cysteines in the enzyme and aids in substrate binding by changing the enzyme conformation. The addition of even 5 µL of 2 mM DTT was seen to have a marked effect on enzyme activity. Addition of 15 µL of 2 mM DTT resulted in a 67% enhancement of FhCC activity at pH 5.5 and a 50% enhancement at pH 8.5 (Figure 62). Addition of 100 µL of 20 mM DTT was recommended for optimal activity (Hola-Jamriska *et al.*, 1999). This resulted in a 524% enhancement of activity. Therefore, 100 µL of 20 mM DTT was used in all assays to detect the presence of recombinant protein.

A series of pH profile assays were undertaken to investigate the pH optimum of FhCC. The graph in Figure 63 illustrates that two definite peaks of enzyme activity occurred, one at pH 5.5 and the other at pH 8.5. In this assay 1.5 times more NHMec was liberated from the fluorogenic substrate at

pH 8.5 than at pH 5.5. Activity at pH 5.5 was 67% of that at pH 8.5. However, other assays revealed that the activity at pH 5.5 was greater than at 8.5. When the buffer was changed from sodium acetate pH 5.5 to a sodium phosphate buffer of the same pH, a four-fold decline in FhCC activity resulted. This decline in activity did not occur on changing the buffer from sodium phosphate pH 7.5 to Tris-HCl, pH 7.5. The peak at pH 5.5 using the sodium acetate buffer was consistent with the pH optimum of schistosome cathepsin C using the same H-Gly-Arg-NHMec substrate. Optimum DPP III activity was previously reported at pH 8.5 but this does not explain the peak observed using H-Gly-Arg-NHMec as the substrate for DPP III is H-Arg-Arg-NHMec and DPP III does not cleave H-Gly-Arg-NHMec at pH 8.5 (Hola-Jamriska *et al.*, 1999). In another study carried out in our laboratory, the peak at pH 8.5 was suggested to represent a second pH preference for cathepsin C. Alternatively, another enzyme could be present in the somatic extract capable of cleaving the cathepsin C substrate.

Cathepsin L1 and L2 are the dominant proteases present in *F. hepatica* (Spithill *et al.*, 1999; Dalton and Mulcahy, 2001). To investigate if the peak at pH 8.5 was due to the activity of cathepsin L, the cathepsin L inhibitors Z-Phe-Ala-NH_2 and Z-Phe-Phe-OH were incubated with the enzyme for 10 min before addition of the substrate (Table 56 and 57). For the assay performed with Z-Phe-Ala-NH_2 highest activity was observed at pH 8.5 as before and the smaller peak at pH 5.5 was also present (Figure 64). Four times more NHMec was released at pH 8.5 than at pH 5.5. For the assay using Z-Phe-Phe-OH, assays were performed at pH 5.5 and pH 8.5 (Table 57). Only a slight decrease in enzyme activity was observed in assays containing the inhibitor. The FhCC activity for the assay carried out at pH 5.5 in the absence of inhibitor was 5 U. When inhibitor was present, activity was 4.1 U, which represents a modest decline in activity of only 8%. At pH 8.5 3.67 U of enzyme activity was observed in the presence of inhibitor, which represents a 30% decline in activity relative to the assay performed without inhibitor, which had 5.27 U of activity. When assays were performed in sodium phosphate buffer, pH 5.5, no decline in activity was observed on

addition of inhibitor. Therefore, it is unlikely that the peak of activity at pH 8.5 is due to the action of a cathepsin L-like enzyme but could perhaps be due to the action of a previously uncharacterised enzyme. Alternatively, this could signify another pH preference of FhCC not evident in its schistosome counterpart.

Inhibition studies were carried out to further characterise the activity responsible for the hydrolysis of H-Gly-Arg-NHMec. In the study of schistosome DPP I iodoacetamide (1 mM) was the most potent inhibitor of DPP I resulting in 99% inhibition of the enzyme. NEM (1 mM) caused 95% inhibition of DPP I activity and leupeptin (10 μM) inhibited the enzyme by 72% (Hola-Jamriska *et al.*, 1999). Iodoacetamide (1mM) inhibited FhCC by 40.3% (Table 58), 0.1 mM leupeptin inhibited the enzyme by 44.6 % (Table 59) and 1 mM NEM led to 33% inhibition of FhCC (Table 60). As previous studies carried out on schistosome cathepsin C showed leupeptin to be a less potent inhibitor than iodoacetamide and NEM, it is surprising that it was the most potent inhibitor of FhCC activity of the three. Inhibition of FhCC was low in all assays relative to the levels obtained for schistosome cathepsin C. Perhaps a higher concentration of inhibitor is needed to inhibit FhCC activity than is the case for schistosomes.

Mammalian cathepsin Cs are more slowly inhibited by E-64 relative to other cysteine proteases such as cathepsin B, H and L (Barrett *et al.*, 1982). A time course study was performed on FhCC using the inhibitor E-64 to investigate if this was the case for FhCC. An assay was carried out with and without E-64 for cathepsin C and cathepsin L1 using their respective substrates. The time for which FhCC and FhCL1in the somatic extract were incubated with E-64 before the addition of substrate was varied (Table 61). As depicted in Figure 66, for the FhCC assay with E-64 only slight inhibition was evident. After 10 min 81.3% of the activity remained, whereas after 80 min incubation FhCC retained 62.13% of its activity. By contrast, hydrolysis of Z-Phe-Arg-NHMec by cathepsin L was blocked by 91% almost immediately. In schistosomes, SjCC was only slightly inhibited by E-64

after 10 min. More than 70% of the activity remained after 40 min and only 50% of the activity was blocked after 80 min. Cathepsin L activity was immediately blocked in the presence of E-64 (Hola-Jamriska *et al.*, 1999). The results demonstrate the slow inhibition by E-64 observed for mammalian and schistosome cathepsin C is also a characteristic of FhCC.

Mammalian cathepsin C is enhanced in the presence of chloride ions (Butler *et al.*, 1969). It was later discovered that the location of the chloride ion in the S2 subsite of cathepsin C directed the interaction between the ion and the substrate and therefore enhanced cathepsin C activity. The activation of Cl^- is pH dependent, being stronger at pH values below 7.0 (Cigic and Pain, 1999). In studies of schistosome cathepsin C it was found that H-Gly-Arg-NHMec hydrolysis was increasingly enhanced in the presence of NaCl and NaBr in the range 0.1-10.0 mM. Activity was enhanced by NaI ions at a concentration of 1.0 mM but was slightly reduced at 10.0 mM. Fluoride ions at 0.1-10.0 mM did not enhance activity (Hola-Jamriska *et al.*, 1999).

Halide enhancement assays performed on somatic extract using cathepsin C specific substrate revealed that 1.0 mM NaCl enhanced activity by 32.12% (Table 62). NaBr was a weaker enhancer than NaCl; 1.0 mM and 10.0 mM NaBr led to a 13.32% and 19.44% enhancement in activity respectively (Table 63). 1.0 mM NaI activated FhCC by 36.36%, whereas 10.0 mM NaI resulted in 44.21% activation (Table 64). This finding does not concur with that obtained for SjCC in the presence of NaI, which showed a slight reduction in activity at a higher iodide concentration (Hola-Jamriska *et al.*, 1999). As anticipated, NaF did not enhance activity. In fact, it had an inhibitory effect (Table 65).

It is possible that like mammalian cathepsin C, FhCC has an absolute requirement for chloride ions. The fact that activity was observed even in the absence of exogenous halides could be due to the presence of endogenous halides in the somatic extract.

A BCA protein assay was carried out to determine the protein concentration of the extract in order to calculate the specific activity of FhCC. The specific activity was calculated to be 0.142 U/mg of protein. Another preparation of somatic extract was shown to have a specific activity of 0.0414 U/mg of protein, which represents a very low activity.

In conclusion, FhCC was biochemically characterised and was seen to exhibit similar characteristics to its schistosome counterpart. Some differences exist but overall the biochemical properties of cathepsin C from the two different helminths were similar.

4.6: Conclusion

The cDNA encoding SmCC was successfully amplified from a cDNA library using specific primers containing sites for *SnaBI* and *AvrII* digestion. This facilitated cloning of the cDNA into the *Pichia pastoris* expression vector, pPIC9K in frame with the initiation codon of the yeast α-factor secretion signal. *P. pastoris* was transformed with the SmCC cDNA and transformants were selected on the basis of G418 resistance.

Protein expression was initially induced with BMMY media supplemented with 1% methanol. A 22kDa band was visible on resolution of culture supernatants by SDS-PAGE. This polypeptide was not reactive against anti-His_6 antibodies and affinity chromatography on Ni-NTA using the D6 supernatant revealed the absence of the his tag as no protein bound to the column. An induced band of 40 kDa appeared on SDS-PAGE analysis of culture pellets but further investigation did not prove this band to be SmCC.

When a twice-daily feed of methanol to a final concentration of 1.5% was employed, SDS-PAGE analysis of culture supernatants revealed the presence of distinct bands at 58, 55 and 47 kDa that were absent from D0 samples. It is highly likely that these bands represented differentially glycosylated forms of the proenzyme since *P. pastoris* is capable of adding both O-and N-linked

carbohydrate moieties to secreted proteins (Cereghino and Cregg, 2000). Enzyme activity was observed and although low, this indicated that the mature enzyme was produced. Due to the fact that mammalian cathepsin Cs retain part of their proregion in the mature enzyme any of aforementioned bands could represent the mature enzyme with this residual propart attached (Dolenc *et al.*, 1995; Turk *et al.*, 2001; Horn *et al.*, 2002). Degradation products at <6.5 kDa were visible for all supernatant samples analysed which indicates that proteolysis occurred, possibly due to the action of yeast proteases. When this induction was repeated using BMMY media at pH 5.0, 6.0 and 7.0, no protein was produced at pH 5.0, a single protein of 55 kDa was expressed at pH 6.0 and at pH 7.0 the bands previously observed at 58, 55 and 47 kDa were produced.

Despite the fact that an intact his tag could not be detected for any of the culture supernatants that showed potential cathepsin C bands when resolved by SDS-PAGE, SjC antisera bound to a protein of 47 kDa, which implies that this is SmCC. Perhaps secondary folding at the COOH terminus obscured the his tag or perhaps it was degraded by yeast proteases. A colony blot was performed and revealed that the protein was not being expressed. N-terminal sequencing is the only way to conclusively prove that the proteins produced are SmCC but before this is carried out different parameters of the fermentation process need to be manipulated to optimise the production of the recombinant protein.

4.7: RECOMMENDATIONS

When future research on the expression of SmCC by *Pichia pastoris* is carried out, it may be useful to implement some of the following strategies, which could have an effect on the expression levels of the recombinant protein:

- Although *BglII* linearised plasmids are more likely to generate high copy number transformants than *SalI* or *SacI*, all transformants used in this project to initiate transformation were single copy. Generally the

use of multi-copy inserts has improved yield. For instance, the production of tetanus toxin fragment C was related to the plasmid copy number (Clare *et al.*, 1991b). Another transformation could be performed using a higher concentration of DNA, in an attempt to obtain multi-copy transformants. Alternatively, a vector could be constructed with multiple head to tail copies of the expression cassette in order to increase the copy-number and improve yield (Cereghino and Cregg, 2000).

- Expression levels in shake flasks are low relative to those obtained with fermenters (Higgins and Cregg, 1998). It is only in the controlled environment of a fermenter that it is possible to grow *P. pastoris* to high cell densities and as the concentration of product is roughly proportional to the concentration of cells in culture, this is an important factor. It has been demonstrated that by switching from a shake flask to a fermenter protein production in *P. pastoris* has increased by 140% (Cino, 1999). Aeration is a critical factor for the expression of a protein and Mut^S strains are more likely to become O_2 limited (Sreekrishna *et al.*, 1997). In fermenters dissolved oxygen level is strictly monitored so that oxygen limitation does not occur. Studies have shown that a continuous fermentation process can result in a 5.0-6.0 fold higher productivity than a fed-batch system so perhaps when a fermentation is being carried out in future, it might be worth considering the possibility of operating it in a continuous mode (Goodrick *et al.*, 2001).

- It has been suggested that for Mut^S strains, addition of glycerol together with methanol during the induction phase may increase the growth rate of the cells, which could increase the production rate of the recombinant protein (Files *et al.*, 2001). This could be investigated for rSmCC production.

- The level of expression at pH 5.0, 6.0 and 7.0 was investigated in this project. Although no protein was produced at pH 5.0, the effect of low pH on SmCC expression should be determined. The recommended pH range for expression in *P. pastoris* was between pH 2.8-6.5 (Thiry and Cingolani, 2001). Media manipulation and pH changes can significantly alter the profile of protein components and can result in proteins that were previously unnoticed becoming evident.

- All inductions carried out in this project were performed at 30^0C. Protein expression ceases at 32^0C. High heat loads can result when *P. pastoris* is actively growing or expressing high levels of protein. In shake flasks, temperatures can increase by 25^0C if left uncontrolled (Cino *et al.*, 1999). Therefore, temperature should be monitored strictly for all future inductions. In one study, lowering the induction temperature from 30^0C to 25^0C resulted in a four-fold increase in yield of galactose oxidase in *P. pastoris* (Whittaker, 2000). On the basis of these findings, the effect of lowering the temperature on rSmCC expression should be investigated.

- Proteolysis represents the most common reason for failure of an expression process (Romanos, 1995). Proteolytic degradation occurred in induction # 4 as many low bands at < 6.5 kDa were visible on SDS-PAGE analysis of supernatants. To minimise proteolysis, amino acid or peptide supplements or EDTA should be added to the growth media. The media could also be buffered to pH 3.0, which would be unfavourable for the production of the harmful proteases (Sreekrishna *et al.*, 1997). If proteolysis was discovered to be a major obstacle to the production of functional SmCC perhaps protease deficient *P. pastoris* strains could be utilised.

5: BIBLIOGRAPHY

5: Bibliography

- Ada, G.L. (1993) Chapter 7: Vaccines and the challenge of parasitic infections. In: Immunology and Molecular Biology of Parasitic Infections, 3rd edition, pp. 126-137. Blackwell Scientific Publications, Cambridge MA, USA.

- Adrio, J.L., Velasco, J., Soler, G., Rodriguez Saiz, M., Barredo, J.L. and Moreno, M. A. (2001) Extracellular Production of Biologically Active Deacetoxycephalosporin C Synthase from *Streptomyves clavuligerus* in *Pichia pastoris*. *Biotechnology and Bioengineering* **75**, (4): 487-491.

- Andersen, D.C. and Krummen, L. (2002) Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology* **13**, (2): 117-123.

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

- Balloul, J.M., Grzych, J.M., Pierce, R.J. and Capron, A. (1987a) A purified 28,000 Dalton protein from *Schistosoma mansoni* adult worms protects rats and mice against experimental schistosomiasis. *The Journal of Immunology* **138**, 3448-3453.

- Balloul, J.M., Sondermeyer, P., Dreyer, D., Capron, M., Grzych, J.M., Pierce, R.J., Carvallo, D., Lecocq, J.P. and Capron, A. (1987b) Molecular cloning of a protective antigen of schistosomes. *Nature* **326**, 149-154.

- Baneyx, F. (1999) Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology* **10**, 411-421.

- Barrett, A.J., Kembhaui, 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 cathepsin B, H and L. *The Biochemical Journal* **47**, 280-293.

- Barrett, A.J., Rawlings, N.D. and Woessner, J.F. (1998) Introduction to cysteine peptidases and their clans. In: Handbook of proteolytic enzymes, Barrett, A.J., Rawlings, N.D. and Woessner, J.F (Ed.), pp. 545-555. Academic Press, San Diego, USA.

- Bergquist, N.R. (2002) Schistosomiasis: from risk assessment to control. *Trends in Parasitology* **18**, 309-314.

- Berti, P.J. and Storer, A.C. (1995) Alignment/Phylogeny of the Papain Superfamily of Cysteine Proteases. *Journal of Molecular Biology* **246**, 273-28

- Boulanger, D., Reid, G.D., Sturrock, R.F., Wolowczuk, I., Balloul, J.M., Grezel, D., Pierce, R.J., Otieno, M.F., Guerret, S., Grimaud, J.A., Butterworth, A.E. and Capron, C. (1991) Immunization of mice and baboons with the recombinant Sm28 GST affects both worm viability and fecundity after experimental infection with *Schistosoma mansoni*. *Parasite Immunology* **13**, 473-490.

- Bradley, J.E. (2001) Chapter 16: Immunity to Protozoa and Worms. In: Immunology, Roitt, I., Brostoff, J. and Male, D. (Ed.), Sixth Edition, pp. 259-276. Mosby International, New York, USA.

- Brindley, J.P., Kalinna, B.H., Dalton, J.P., Day, S.R., Wong, J.Y., Smythe, M.L., and McManus D.P. (1997) Proteolytic degradation of host haemoglobin by schistosomes. *Molecular and Biochemical Parasitology* **89**, 1-9.

- Brindley, P.J., Kalinna, B.H., Wong, J.Y.M., Bogitsch, B.J., King, L.T., Smyth, D.J., Verity, C.K., Abbenante, G., Brinkworth, R.I., Fairlie, D.P., Smythe, M.L., Milburn, P.J., Bielefeldt-Ohmann, H., Zheng, Y. and McManus, D.P. (2001) Proteolysis of human hemoglobin by schistosome cathepsin D. *Molecular and Biochemical Parasitology* **112**, 103-112.

- Brophy, P.M. and Barrett, J. (1990) Glutathione Transferases in Helminths. *Parasitology* **100**, 345-349.

- Buckholz, R.G. and Gleeson, M.A.G. (1991) Yeast Systems for the Commercial Production of Heterologous Proteins. *Biotechnology* **9**, 1067-1072.

- Bush, A.O., Fernandez, J.C., Esch, G.W. and Seed, J.R. (2001) Chapter 1: Introduction. In: Parasitism–The diversity and ecology of animal parasites, 1st edition, pp. 1-11. Cambridge University Press, Cambridge, UK.

- Butler, R., Michel, A., Kunz, W. and Klinkert, M-Q. (1995) Sequence of *Schistosoma mansoni* cathepsin C and its structural comparison with papain and cathepsins B and L of the parasite. *Protein and Peptide Letters* **2**, 313-320.

- Caffrey, C.R., Mathieu, M.A., Gaffney, A.M., Salter, J.P., Sajid, M., Lucas, K.D., Franklin, C., Bogyo, M. and Mc Kerrow J.H. (2000) Identification of a cDNA encoding an active asparaginly endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*. *FEBS Letters* **466**, 244-248.

- Caffrey, C.A., Hansell, E., Lucas, K.D., Brinen, L.S., Hernandez, A.A., Cheng, J., Gwaltney, S.L., Roush, W.R., Stierhof, Y.D., Bogyo, M., Steverding, D. and Mc Kerrow, J.H. (2001) Active site mapping, biochemical properties and subcellular localization of rhodesain, the

major cysteine protease of *Trypanosoma brucei rhodesiense*. *Molecular and Biochemical Parasitology* **118**, 61-73.

- Caffrey, C.R., Salter, J.P., Lucas, K.D., Khiem, D., Hsieh, I., Lim, K.C., Ruppel, A., McKerrow, J.H. and Sajid, M. (2002) SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. *Molecular and Biochemical Parasitology* **121**,49-61.

- Capron, A. (1998) Schistosomiasis: Forty years' war on the worm. *Parasitology Today* **14** (10), 379-384.

- Capron, A. (2001) Vaccine strategies against schistosomiasis: from concepts to clinical trials. *International Archives of Allergy and Immunology* **124**, 9-15.

- Carmona, C., Mc Gonigle, S., Dowd, A.J., Smith, A.M., Coughlan, S., Mc Gowran, E. and Dalton, J.P. (1993) A dipeptidylpeptidase secreted by *Fasciola hepatica*. *Parasitology* **109**, 113-118.

- Cereghino, G.P.L. and Cregg, J.M. (1999) Applications of yeast in biotechnology: protein production and genetic analysis. *Current Opinion in Biotechnology* **10**, 422-427.

- Cereghino, J.L. and Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews* **24**, 45-66.

- Cheng, T.C. (1986) Chapter 1: Parasitism and Symbiosis, pp. 1-12. Chapter 10: Digenea: The Digenetic Trematodes, pp. 332-342. In: General Parasitology, Second Ed. Academic Press Inc., Orlando, Florida 32887, USA.

- Chitsulo, L., Engels, D., Montresor, A. and Savioli, L. (2000) The global status of schistosomiasis and its control. *Acta Tropica* **77**, 41-51.

- Chlichlia, K., Bahgat, M., Ruppel, A. and Schirrmacher, V. (2002) DNA vaccination with asparaginyl endopeptidase (Sm32) from the parasite *Schistosoma japonicum*: anti-fecundity effect induced in mice. *Vaccine* **20**, 439-447.

- Cigic, B. and Pain, R.H. (1999) Location of the binding site for chloride ion activation of cathepsin C. *European Journal of Biochemistry* **264,** 944-951.

- Cigic, B., Dahl, S, W. and Pain, R.H. (2000) The residual pro-part of cathepsin C fulfils the criteria required for an intramolecular chaperone in folding and stabilizing the human proenzyme. *Biochemistry* **39**, 12382-12390.

- Cino, J. (1999) High Yield Protein Production from *Pichia pastoris* Yeast: A Protocol for Benchtop Fermentation (http://www.nbsc.com).

- Clare, J.J., Romanos, M.A., Rayment, F.B., Rowedder, J.E., Smith, M.A., Payne, M.M., Sreekrishna, K. and Henwood, C.A. (1991a) Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. *Gene* **105**, 205-212.

- Clare, J.J., Rayment, F.B., Ballantine, S.P., Sreekrishna, K. and Romanos, M.A. (1991b) High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Biotechnology* **9**, 455-460.

- Coulson, P.S. (1997) The radiation attenuated vaccine against schistosomes in animal models: paradigm for a human vaccine? *Advances in Parasitology* **39**, 271-336.

- Cregg, J.M., Tschopp, J.F., Stillman, C., Siegel, R., G, M., Craig, W.S., Buckholz, R.G., Madden, K.R., Kellaris, P.A., Davis, G.R., Smiley, B.L., Cruze, J., Torregrossa, R., Velicelebi, G. and Thill, G.P. (1987) High-level expression and efficient assembley of hepatitis B surface antigen in the methylotrophic yeast, *Pichia pastoris*. *Biotechnology* **5**, 479-485.

- Da'dara, A.A., Skelly, P.J., Wang, M.M. and Harn, D.A. (2002) Immunization with plasmid DNA encoding the integral membrane protein, Sm23, elicits a protective immune response against schistosome infection in mice. *Vaccine* **20,** 359-369.

- D'Angou, M.C. and Daugulis, A.J. (2001) A Rational Approach to Improving Productivity in Recombinant *Pichia pastoris* Fermentation. *Biotechnology and Bioengineering* **72**, (1): 1-11.

- 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.

- Dalton, J.P., Smith, A.M., Clough, K.A. and Brindley, PJ. (1995) Digestion of haemoglobin by schistosomes: 35 years on. *Parasitology Today* **11** (8), 299-302.

- Dalton, J.P., Clough, K.A., Jones, M.K. and Brindley, P.J. (1996) Characterization of the Cathepsin-Like Cysteine Proteinases of *Schistosoma mansoni*. *Infection and Immunity* **64**, 1328-1334.

- Dalton, J.P. and Brindley, P.J. (1996) Schistosome asparaginyl endopeptidase Sm32 in hemoglobin digestion. *Parasitology Today* **12**, 125.

- Dalton, J.P. and Mulcahy, G. (2001) Parasite vaccines-a reality? *Veterinary Parasitology* **98**, 149-167.

- De Bruin, E.C., de Wolf, F.A. and Laane, N.C.M. (2000) Expression and secretion of human α1(I) procollagen fragment by *Hansenula polymorpha* as compared to *Pichia pastoris*. *Enzyme and Microbial Technology* **26**, 640-644.

- Doenhoff, M.J. (1998) A Vaccine for Schistosomiasis: Alternative Approaches. *Parasitology Today* **14** (3): 105-109.

- Dolenc, I., Turk, B., Pungercic, G., Ritonja, A. and Turk, V. (1995) Oligomeric structure and substrate induced inhibition of human cathepsin C. *The Journal of Biological Chemistry* **270**, 21626-21631.

- Dowd, A.J., Smith, A.M., Mc Gonigle, S. and Dalton, J. (1994) Purification and characterisation of a second cathepsin L proteinase secreted by the parasitic trematode *Fasciola hepatica*. *European Journal of Biochemistry* **223**, 91-98.

- Eberl, M., Langermans, J.A., Frost, P.A., Vervenne, R.A., Van Dam, G.J., Deelder, A.M., Thomas, A.W., Coulson, P.S. and Wilson, R.A. (2001) Cellular and Humoral Immune Responses and Protection against Schistosomes Induced by a Radiation-Attenuated Vaccine in Chimpanzees. *Infection and Immunity* **69** (9); 5352-5362.

- Echelard, Y. (1996) Recombinant protein production in transgenic animals. *Current Opinion in Biotechnology* **7**, 536-540.

- Esch, G.W. and Fernandez, J.C. (1993) Chapter 1: Introduction. From: A functional biology of parasitism, 1st edition, pp.1-5. Chapman and Hall, London, UK.

- Faber, K.N., Harder, K., Ab, G. and Veenhuis, M. (1995) Review: Methylotrophic Yeasts as Factories for the Production of Foreign Proteins. *Yeast* **11**, 1131-1344.

- Files, D., Ogawa, M., Scaman, C.H. and Baldwin, S.A. (2001) A *Pichia pastoris* fermentation process for producing high-levels of recombinant human cystatin-C. *Enzyme and Microbial Technology* **29**, 335-340.

- Flanigan, T.P., King, C.H., Lett, R.R., Nanaduri, J. and Mahmoud, A.A. (1989) Induction of resistance to *Schistosoma mansoni* infection in mice by purified parasite paramyosin. *Journal of Clinical Investigation* **83**, 1010-1014.

- Frye, C.C., Herschber, C.L., and Zhang, H. (2000) Bovine dipeptidylpeptidase I. U.S. Patent No: 6027911.

- Gellissen, G. and Hollenberg, C.P. (1997) Application of yeasts in gene expression studies: a comparison of *Saccharomyeces cerevisiae, Hansenula polymorpha* and *Kluyveromyces lactis*-a review. *Gene* **190**, 87-97.

- Gellissen, G. (2000) Heterologous protein production in methylotrophic yeasts. *Applied Microbiology and Biotechnology* **54**, 741-750.

- Godal, T. (1996) Chapter 1: New Dimensions for Parasitology in the 21st Century. In: Parasitology for the 21st century, Özcel, M.A. and Alkan, M.Z. Ed., pp. 1-13. CAB International, Wallingford, Oxon OX10 8DE, UK.

- Goodrick, J.C., Xu, M., Finnegan, R., Schilling, B.M., Schiavi, S., Hoppe, H. and Wan, N.C. (2001) High-Level Expression and Stabilization of Recombinant Human Chitinase Produced in a Continuous Constitutive

Pichia pastoris Expression System. *Biotechnology and Bioengineering* **74**, (6): 492-497.

- Gotz, B. and Klinkert, M.Q. (1993) Expression and partial characterisation of a cathepsin B like enzyme (Sm31) and a proposed "haemoglobinase" (Sm32) from *Schistosoma mansoni*. *Biochemical Journal* **290**, 801-806.

- Grzych, J.M., Grezel, D., Xu, C.B., Neyrinck, J.L., Capron, M., Ouma, J.H., Butterworth, A.E. and Capron, A. (1993) IgA antibodies to a protective antigen in human schistosomiasis mansoni. *The Journal of Immunology* **150**, 527-535.

- Gutman, H.R. and Frution, J.S. (1948) The proteolytic enzymes of animal tissues VIII. An intracellular enzyme related to chymotrypsin. *Journal of Biological Chemistry* **174**, 851-858.

- Hagan, P., Doenhoff, R.A., Wilson, M, Al-Sherbiny and Bergquist, R. (2000) Schistosomiasis Vaccines: A Response to a Devils' Advocate's View. *Parasitology Today* **16** (8): 322-323.

- Hart, P.S., Zhang, Y., Firatli, E., Uygur, C., Lotfazar, M., Michalec, M.D., Marks, J.J., Lu, X., Coates, B.J., Seow, W.K., Marshall, R., Williams, D., Reed, J.B., Wright, J.T. and Hart, T.C. (2000) Identification of cathepsin C mutations in ethnically diverse Papillion-Lefévre syndrome patients. *Journal of Medical Genetics* **37**, 927-932.

- Hellwig, S., Emede, F., Raven, N.P.G., Henke, M., van der Logt, P. and Fischer, R. (2001) Analysis of Single-Chain Antibody Production in *Pichia pastoris* Using On-Line Methanol Control in Fed-Batch and Mixed–Feed Fermentations. *Biotechnology and Bioengineering* **74**, (4): 344-352.

- Higgins, D.R. and Cregg, J.M. (1998) Chapter 1: Introduction to *Pichia pastoris* pp. 1-15. In: Methods in Molecular Biology, Vol. 103: *Pichia* Protocols, Higgins, D.R. and Cregg, J.M. ed. Humana Press Inc., Totowa, NJ, USA.

- Hockney, R.C. (1994) Recent developments in heterologous protein production in Escherichia coli. *Trends in Biotechnology* **12**, 456-463.

- Hola-Jamriska, L., Tort, J.F., Dalton, J.P., Day, S.R., Fan, J., Aaskov, J. and Brindley, P.J. (1998) Cathepsin C from *Schistosoma japonicum* cDNA encoding the preproenzyme and its phylogenetic relationships. European *Journal of Biochemistry* **255**, 527-534.

- Hola-Jamriska, L., Dalton, J.P., Aaskov, J. and Brindley, P.J. (1999) Dipeptidyl peptidase I and III activities of adult schistosomes. *Parasitology* **118**, 275-282.

- Hola-Jamriska, L., King, L.T., Dalton, J.P., Mann, V.H., Aaskov, J.G. and Brindley, P.J. (2000) Functional expression of dipeptidyl peptidase I (cathepsin c) of the oriental blood fluke *Schistosoma japonicum* in *Trichoplusia ni* insect cells. *Protein Expression and Purification* **19**, 384-392.

- Hollenberg, C.P. and Gellissen, G. (1997) Production of recombinant proteins by methylotrophic yeasts. *Current Opinion in Biotechnology* **8**, 554-560.

- 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 Science* **11**, 933-943.

- Howell, M.J., Board, P.G. and Boray, J.C. (1988) Glutathione-S-Transferases in *Fasciola hepatica*. *Journal of Parasitology* **74**, 715-718.

- Ishidoh, K., Muno, D., Sato, N. and Kominami, E. (1991) Molecular cloning of cDNA for rat cathepsin C. *The Journal of Biological Chemistry* **266**, 16312-16317.

- Ismail, M., Botros, S. and Metwally, A. (1999) Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *American Journal of Tropical Medicine and Hygiene* **60**, 932-935.

- Jarvis, D.L. and Finn, E.E. (1995) Biochemical analysis of the N-glycosylation pathway in baculovirus-infected lepidopteran insect cells. *Virology* **212**, 500-511.

- Jenkins, N., Parekh, R.B. and James, D.C. (1996) Getting the glycosylation right: Implications for the biotechnology industry. *Nature Biotechnology* **14**, 975-979.

- Jordan, P. (2000) From Katayama to the Dakhla Oasis: the beginning of epidemiology and control of bilharzias. *Acta Tropica* **77**, 9-40.

- Kalinna, B.H., Becker, M.M. and McManus, D.P. (1997) Engineering and expression of a full-length cDNA encoding *Schistosoma japonicum* paramyosin. Purification of the recombinant protein and its recognition by infected patient sera. *Acta Tropica* **65**, 111-115.

- Khalife, J., Cêtre, C., Pierrot, C. and Capron, M. (2000) Mechanisms of resistance to *S. mansoni* infection: the rat model. *Parasitology International* **49**, 339-345.

- Klinkert, M.Q., Felleisen, R., Link, G., Ruppel, A. and Beck, E. (1989) Primary structure of Sm31/32 diagnostic proteins of *Schistosoma mansoni* and their identification as proteases. *Molecular and Biochemical Parasitology* **33**, 113-122

- Kojima, S., Niimura, M. and Kanazawa, T. (1987) Production and properties of a mouse monoclonal IgE antibody to *Schistosoma japonicum*. *Journal of Immunology* **139**, 2044-2049.

- Kost, T.A. and Condrey, J.P. (1999) Recombinant baculovirus as expression vectors for insect and mammalian cells. *Current Opinion in Biotechnology* **10**, 428-433.

- Kummer, J.A., Kamp, A.M., Citarella, F., Horrevoets, A.J.G. and Hack, C.E. (1996) Expression of human recombinant granzyme A zymogen and its activation by the cysteine proteinase Cathepsin C. *The Journal of Biological Chemistry* **271** (16), 9281-9286.

- Lambertucci, J.R., Serufo, J.C., Gerspacher-Lara, R., Rayes, A.A., Teixeira, R., Nobre, V. and Antunes, C.M. (2000) Schistosoma mansoni: assessment of morbidity before and after control. *Acta Tropica* **77**(1): 101-9.

- Lauritzen, C., Pedersen, J., Madsen, M., Justesen, J., Martensen, P.M. and Dahl, S. (1998) Active Recombinant Rat Dipeptidyl Aminopeptidase I (cathepsin C) Produced Using the Baculovirus Expression System. *Protein Expression and Purification* **14**, 434-442.

- Lauritzen, C., Pedersen, J., Turk, V. and Turk, B. (2001) Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. *EMBO Journal*, **20**, (23): 6570-6582.

- Lee, T.K., Drohan, W.N. and Lubon, H. (1995) Proteolytic processing of human protein C in swine mammary gland. *Journal of Biochemistry* **118**, 81-87.

- Lipps, G., Fullkrug, R. and Beck,E. (1996) Purification and activation of the recombinant proenzyme secreted by *Saccharomyeces cerevisiae*. *Journal of Biological Chemistry* **271**, 1717-1725.

- Lodisch, H.F. (1988) Transport of secretory and membrane glycoproteins from the rough endoplasmic reticulum to the golgi. *The Journal of Biological Chemistry* **263**, 2107-2110.

- Lodish, H., Baltimore, D., Berk, A., Zipursky, S.L., Matsudaira, P. and Darnell, J. (1995) Chapter16: Synthesis and Sorting of Plasma Membrane, Secretory and Lysosomal Proteins. In: Molecular Cell Biology, 3rd edition, pp 699-711. Scientific American Books Inc., New York, USA.

- Maizels, R.M., Bundy, D.A., Selkirk, M.E., Smith, D.F. and Anderson, R.M. (1993) Immunological modulation and evasion by helminth parasites in human populations. *Nature* **365**, 797-805.

- McAleer, W.J., Buynak, E.B., Maigetter, R.Z., Wampler, D.E., Miller, W.J. and Hilleman, M.R. (1984) Human hepatitis B vaccine from recombinant yeast. *Nature* **307**, 178-180.

- McGuire, M.J., Lipsky, P.E. and Thiele, D.L. (1992) Purification and characterization of dipeptidyl deptidase I from human spleen. *Arch. Biochem.Biophys*. **306**, 103-110.

- McGuire, M.J., Lipsky, P.E, and Thiele, D.L. (1993) Generation of active myeloid and lymphoid granule serine proteases requires processing

by the granule thiol protease dipeptidyl peptidase I. *The Journal of Biological Chemistry* **268** (4), 2458-2467.

- McGuire, M.J., Lipsky, P.E. and Thiele, D.L. (1997) Cloning and characterization of the cDNA encoding mouse dipeptidyl peptidase 1 (Cathepsin C). *Biochimica et Biophysica Acta* **1351**, 267-273.

- McManus, D.P., Wong, J.Y.M., Zhou, J., Cai, C., Zeng, Q., Smyth, D., Li, Y., Kallina, B.H., Duke, M.J. and Yi, X. (2002) Recombinant paramyosin (rec-Sj-97) tested for immunogenicity and vaccine efficacy against *Schistosoma japonicum* in mice and water buffaloes. *Vaccine* **20**, 870-878.

- Medved, L.V., Orthner, C.L., Lubon, H., Lee, T.K., Drohan, W.N. and Ingham, K.C. (1995) Thermal stability and domain-domain interactions in natural and recombinant protein C. *Journal of Biological Chemistry* **270**, 13652-13659.

- Moir, D.T. and Dumais, D.R. (1987) Glycosylation and secretion of human alpha-1-antitrypsin by yeast. *Gene* **56**, 209-217.

- Monroy, F.P. and Dresden, M. (1996) The developmental expression of cysteine proteinases in *Schistosoma mansoni*. *International Journal for Parasitology* **26**, 109-112.

- Mulcahy, G., Joyce, P. and Dalton, J.P. (1999) Immunology of *Fasciola hepatica* infection. In: Fasciolosis Dalton, J.P. (Ed.), pp. 341-376. CAB International, Wallingford, Oxon, UK.

- Muno, D., Ishidoh, K., Ueno, T. and Kominami, E. (1993) Processing and transport of the precursor of cathepsin C during transfer to the lysosomes. *Arch. Biochem.Biophys.* **306,** 103-110.

- Nelson, G.S. (1979) Chapter 1: The Parasite and the Host. In: Parasites and Western Man, R.J. Donaldson (Ed.), pp. 1-10. MTP Press Limited, Falcon House, Lancaster, UK.

- Neurath, H. (1984) Evolution of Proteolytic Enzymes. *Science* **224**, 350-357.

- Newport, G.R. and Colley, D.G. (1993) Chapter 16: Schistosomiasis. In: Immunology and Molecular Biology of Parasitic Infections, Third Ed., Warren, K.S., (Ed.,), pp. 387-422. Blackwell Scientific Publications Inc., Cambridge MA, USA.

- Nicaud, J.M., Mackman, N. and Holland, I.B. (1986) Current status of secretion of foreign proteins by microorganisms. *Journal of Biotechnology* **3**, 255-270.

- Nuckolls, G.H. and Slavkin, H.C. (1999) Paths of Glorious Proteases. *Nature Genetics* **23**, 378-380.

- Olsen, J.G., Kadziola, A., Lauritzen, C., Pedersen, J., Larsen, S. and Dahl, S.W. (2001) Tetrameric dipeptidyl peptidase I directs substrate specificity by use of the residual pro-part domain. *FEBS Letters* **506**, 201-206.

- Paris, A., Strukelj, B., Pungercar, J., Renko, M., Dolenc, I. and Turk, V. (1995) Molecular cloning and sequence analysis of human preprocathepsin C. *FEBS Letters* **369**, 326-330.

- Pearse, E.J., James, S.L., Hieny, S., Lanar, D.E. and Sher, A. (1988) Induction of protective immunity against *Schistosoma mansoni* by vaccination with schistosome paramyosin (Sm97), anon surface parasite antigen. *Proceedings of the National Academy of Sciences USA* **85**, 5678-5682.

- Pham, C.T., Armstrong,.R.J., Zimonjic, D.B., Popescu, N.C., Payan, D.G. and Ley, T.J. (1997) Molecular cloning, chromosomal localization, and expression of murine dipeptidyl peptidase I. *The Journal of Biological Chemistry* **272** (16), 10695-10703.

- Pham, C.T.N. and Ley, T.J. (1999) Dipeptidyl peptidase I is a required for the processing and activation of granzymes A and B *in vivo*. *Proceeding of the National Academy of Science USA* **96**, 8627-8632.

- Punt, P.J., Veldhuisen, G. amd an den Hondel, C.A.M.J.J. (1994) Protein targeting and secretion in filamentous fungi. *Antonie van Leuwenhoek* **65**, 211-216.

- Punt, P.J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J. and van den Hondel, C. (2002) Filamentous fungi as cell factories for heterologous protein production. *Trends in Biotechnology* **20** (5): 200-206.

- Quackenbush, J., Liang, F., Holt, I., Pertea, G. and Upton, J. (2000) The TIGR Gene Indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Research* **28**, 141-145.

- Radzio, R. and Kück, U. (1997) Synthesis of biotechnologically relevant heterologous proteins in filamentous fungi. *Process Biochemistry* **32** (6): 529-539.

- Ramirez, B.L., Kurtis, J.D., Wiest.P.M., Arias, P., Aligui, F., Acosta, L., Peters, P. and Olds, G.R. (1996) Paramyosin: a candidate vaccine antigen against *Schistosoma japonicum*. *Parasite Immunology* **18**, 49-52.

- Rao, N.V., Rao, G.V. and Hoidal, J.R. (1997) Human dipeptidyl peptidase I. Gene characterization, localization and expression. *Journal of Biological Chemistry* **272**, 10260-10265.

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

- Redinbaugh, M.G. and Turley, R.B. (1986) Adaptation of the bicinchoninic acid protein assay for use with microtiter plates and sucrose gradient fractions. *Analytical Biochemistry* **153**, 267-271.

- Romanos, M. (1995) Advances in the use of *Pichia pastoris* for high-level gene expression. *Current Opinion in Biotechnology* **6**, 527-533.

- Romanos, M., Scorer, C., Sreekrishna, K. and Clare, J. (1998) Chapter 5: The Generation of Multicopy Recombinant Strains. In: Methods in Molecular Biology, Vol. 103: *Pichia* protocols, Higgins, D.R. and Cregg, J.M. ed. pp. 55-73. Humana Press Inc., Totowa, NJ, USA.

- Ross, A.G.P., Bartley, P.B., Sleigh, A.C., Olds, G.R., Li, Y., Williams, G.M. and McManus, D.P. (2002) Schistosomiasis. *New England Journal of Medicine* **346,** 1212-1220.

- Sajid, M. and McKerrow, J.H. (2002) Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology* **120**, 1-12.

- Schein, C.H. (1989) Production of Soluble Recombinant Proteins in Bacteria. *Biotechnology* **7**, 1141-1147.

- Scott, J.C. and McManus, D.P. (2000) Molecular cloning and enzymatic expression of the 28-kDa glutathione S-transferase of *Schistosoma japonicum*: evidence for sequence variation but lack of consistent vaccine efficacy in the murine host. *Parasitology International* **49**, 289-300.

- Seaton, D.R. (1979) Chapter 6: Cestodes and Trematodes. In: Parasites and Western Man, Donaldson, R.J. ed., pp. 114-131. MTP Press.

- Spithill, T.W., Smooker, P.M., Sexton, J.L., Bozas, E., Morrison, C.A., Creaney, J. and Parsons, J. (1999) Development of vaccines against *Fasciola hepatica*. In: Fasciolosis, Dalton, J.P. (Ed.), CAB International, pp 377-410, Wallingford, Oxon, UK.

- Smyth, J.D. (1976) Chapter 1: Parasitism, pp. 1-7. Chapter 16: Digenea: Schistosomatidae, pp. 193-205. In: Introduction to Animal Parasitology, 2nd edition. Hodder and Stoughton, London, UK

- Spier, R.E. (2000) Yeast as a Cell Factory. *Enzyme and Microbial Technology* **26**, 639.

- Sreekrishna, K., Brankamp, R.G., Kropp, K.E., Blankenship, D.T., Tsay, J. T., Smith, P.L., Wierschke, J.D., Subramaniam, A. and Birkenberger, L.A. (1997) Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* **190**, 55-62.

- Thiry, M. and Cingolani, D. (2002) Optimizing scale-up fermentaion processes. *Trends in Biotechnology* **20**, (3): 103-105.

- Toomes, C., James, J., Wood, A.J., Wu, C.L., McCormick, D., Lench, N., Hewitt, C., Moynihan, L., Roberts, E., Woods, C.G., Markham, A., Wong, M., Widmer, R., Ghaffer, K.A., Pemberton, M., Hussein, I.R., Temtamy, S.A., Davies, R., Read, A.P., Sloan, P., Dixon, M.J. and Thakker, N.S. (1999) Loss-of-function mutations in the cathepsin C gene result in periodontal disease and palmoplantar keratosis. *Nature Genetics* **23**, 421-423.

- Tort, J., Brindley, P.J., Knox, D., Wolfe, K.H. and Dalton, J.P. (1999) Proteinases and associated genes of parasitic helminths. In: *Advances in Parasitology* **43**, 162-247.

- Tschopp, J.F., Sverlow, G., Kosson, R., Craig, W. and Grinna, L. (1987) High-level secretion of glycosylated invertase in the methylotrophic yeast *Pichia pastoris*. *Biotechnology* **5**, 1305-1308.

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

- Turk, V., Turk, B. and Turk, D. (2001) Lysosomal cysteine proteases: facts and opportunities. *The EMBO Journal* **20**, (17): 4629-4633.

- Turk, V., Turk, B., Guncar, G., Turk, D. and Kos, J. (2002) Lysosomal cathepsins: structure, role in antigen processing and presentation and cancer. *Adv. Enzyme Regul.* **42**, 285-303.

- Utzinger, J., Chollet, J., Jiqing, Y., Jinyan, M., Tanner, M. and Shuhua, X. (2001) Effect of combined treatment with praziquantel and artemether on *Schistosoma japonicum* and *Schistosoma mansoni* in experimentally infected animals. *Acta Tropica* **80**, 9-18.

- Vassileva, A., Chugh, D.A., Swaminathan, S. and Khanna, N. (2001) Effect of Copy Number on the Expression Levels of Hepatitis B Surface Antigen in the Methylotrophic Yeast *Pichia pastoris*. *Protein Expression and Purification* **21**, 71-80.

- Vercruysse, J., De Bont, J. and Claerebout,E. (2001) Parasitic Vaccines: A Utopia Or Reality? *Verh K Acad Geneeskd Belg* **63** (1): 41-56.

- Verity, C.K., Mc Manus, D.P. and Brindley, P.J. (2001) Vaccine efficacy of recombinant cathepsin D aspartic protease from *Schistosoma japonicum*. *Parasite Immunology* **23** (3), 153-162.

- Walsh, G and Headon, D. (1994) Chapter 2: Protein Sources. In: Protein Biotechnology, pp. 10-38. John Wiley and Sons Inc., New York, USA.

- Wang, Y., Liang, Z.H., Zhang, Y.S., Yao, S.Y., Xu, Y.G., Tang, Y.H., Zhu, S.Q., Cui, D.F. and Feng, Y.M. (2001) Human Insulin from a Precursor Overexpressed in the Methylotrophic Yeast *Pichia pastoris* and a Simple Procedure for Purifying the Expression Product. *Biotechnology and Bioengineering* **73**, (1): 74-78.

- Weickert, M.J., Doherty, D.H., Best, E.A. and Olins, P.O. (1996) Optimization of heterologous protein production in *Escherichia coli*. *Current Opinion in Biotechnology* **7**, 494-499.

- William, S., Botros, S., Ismail, M., Farghally, A., Day, T.A. and Bennett, J.L. (2001) Praziquantel-induced tegumental damage *in vitro* is diminished in schistosomes derived from praziquantel-resistant infections. *Parasitology* **122**, 63-66.

- Wilson, R.A. and Coulson, P.S. (1998) Why don't we have a Schistosomiasis vaccine? *Parasitology Today* **14** (3), 97-98.

- Windsor, D.A. (1998) Most of the species on Earth are parasites. *International Journal for Parasitology* **28**, 1939-1941.

- Wolters, P.J., Raymond, W.W., Blount, J.L. and Caughey, G.H. (1998) Regulated expression, processing and secretion of dog mast cell dipeptidyl peptidase I. *Journal of Biological Chemistry* **273**, 15514-14520.

- Wong, J.Y.M., Harrop, S.A., Day, S.D. and Brindley, P.J. (1997) Schistosomes express two forms of cathepsin D. *Biochimica et Biophysica Acta* **1338**, 156-160.

- Wynn, T.A, and Hoffmann, K.F. (2000) Defining a Schistosomiasis vaccine-Is it really Th1 versus Th2? *Parasitology Today* **16** (11), 497-501.

- Xu, C.B., Verwaerde, C., Gras-Masse, H., Fontaine, J., Bossus, M., Trottein, F., Wolowczuk, I., Tartar, A. and Capron, A. (1993) *Schistosoma mansoni* 28-kDa glutathione S transferase and immunity against parasite fecundity and egg viability. *The Journal of Immunology* **150**, 940-949

- Yole, D.S., Reid, G.D. and Wilson, R.A. (1996) Protection against *Schistosoma mansoni* and associated immune responses induced in the vervet monkey *Cercopithecus aethiops* by the irradiated cercaria vaccine. *The American Journal of Tropical Medicine and Hygiene* **54** (3), 265-270.

- Zhang, W., Bevins, M.A., Plantz, B.A., Smith, L.A. and Meagher, M.M. (2000) Modeling *Pichia pastoris* Growth on Methanol and Optimizing the Production of a Recombinant Protein, the Heavy-Chain Fragment C of Botulinum Neurotoxin, Serotype A. *Biotechnology and Bioengineering* **70**, (1): 1-8.

- Zhang, Y., Taylor, M.G., Johansen, M.V. and Bickle, Q.D. (2002) Vaccination of mice with a cocktail DNA vaccine induces a Th 1-type immune response and partial protection against *Schistosoma japonicum* infection. *Vaccine* **20**, 724-730.

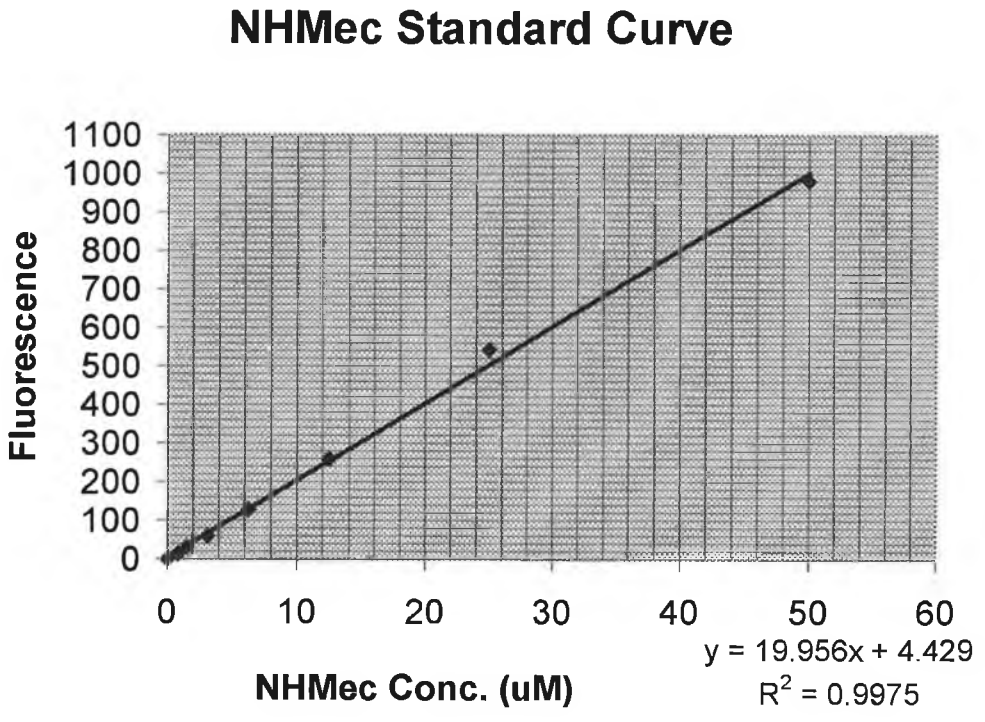
Websites used in this project

- http://www.expasy.ch (clustalw, multalign, translation tools, SOPM)
- http://www.tigr.org (TIGR gene indices)
- http://www.ncbi.nlm.gov (BLAST, Gen bank)
- http://www.npsa-pbil.ibcp.fr (Pôle Bio-Informatique Lyonnais)
- http://www.firstmarket.com (Web cutter, for restriction site analysis)

APPENDIX

An NHMec standard curve representative of those used for all enzyme assays performed in this project is shown below:

NHMec uM	Fluorescence
50	982.4
25	542.2
12.5	258.8
6.25	128.4
3.125	58.67
1.5	29.97
0.78	13.42
0	0.31



Pole Bio-Informatique Lyonnais

Network Protein Sequence @nalysis

NPS@ is the IBCP contribution to PBIL in Lyon, France

Monday, October 7th 2002 : InterProScan is available (see news)

Job **CLUSTALW** (ID: 5518) is running on NPS@ server (started on Tue Feb 11 08:54:52 CET 2003).
Results will be shown below. **Please wait and don't go back.**

n your publication cite :
IPS@: Network Protein Sequence Analysis
TIBS 2000 March Vol. 25, No 3 [291]:147-150
Combet C., Blanchet C., Geourjon C. and Deléage G.

CLUSTALW multiple alignment

Abstract Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple equence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. (1994) Nucleic Acids Research, 22, 4673-4680 [HELP]

ESPript (Toulouse, FR) | Residues repertoire

SHOW alignment with width : 60
all residues.
Conservation level of 0 % (works with "using conservation level of" option in list above).

All prediction methods and secondary consenus with percentage of secondary structure.

- DSC (King and Sternberg, 1996) (Do it alone) [HELP]
- DPM (Deleage and Roux, 1987) [HELP]
- GOR I (Garnier *et al.*,1978) Choose parameters [HELP]
- GOR III (Gibrat *et al.*,1987) [HELP]
- GOR IV (Garnier *et al.*, 1996) [HELP]
- HNN (Guermeur, 1997) [HELP]
- SIMPA96 (Levin *et al.*, 1996) [HELP]
- PHD (Rost *et al.*, 1994) (Do it alone) [HELP]
- PREDATOR (Argos *et al.*, 1996) Choose parameters [HELP]
- SOPM (Geourjon and Deleage, 1994) (Do it alone) [Choose parameters] [HELP]

A prediction method with the Do it alone sentence must be the only newly selected one between two clicks on SHOW button. The other methods can be kept if they have been computed or can be added after. All methods can be displayed at the same time in the alignement.

View CLUSTALW in: [MPSA (Mac, UNIX), About...] [AnTheProt (PC), Download...] [HELP]

```
                    10        20        30        40        50        60
                     |         |         |         |         |         |
atxxx4      DTPANCTYPDLLGTWVFQVGPRHPRSHINCSVMEPTEEKVVIHLKKLDTAYDEVGNSGYF
ousex5      DTPANCTYPDLLGTWVFQVGPRSSRSDINCSVMEATEEKVVVHLKKLDTAYDELGNSGHF
umanx3      DTPANCTYLDLLGTWVFQVGSSGSQRDVNCSVMGPQEKKVVVYLQKLDTAYDDLGNSGHF
eqxxx0      DTPANCTYEDAHGRWKFHIG----DYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHW
Bxxxx1      DTPANCTYEDAHGRWKFHIG----DYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHW
jCCxx2      DTPANCSYMDAIGHWIFHVS----RYKTKCTKQLDVSQTFSMNVQYPNIVTDSYGNMGKW
            ******:* *  * * *:.      . :*.     .:.. : :   : . *. ** * :
rim.cons.   DTPANCTY2D2LGTWVF2VGPR3SRY222CSVML22EQKVVI2L322D2AYDE2GNSGH2

                    70        80        90       100       110       120
                     |         |         |         |         |         |
atxxx4      TLIYNQGFEIVLNDYKWFAFFKYEVKGSRAISYCHETMTGWVHDVLGRNWACFVGKKMAN
ousex5      TLIYNQGFEIVLNDYKWFAFFKYEVRGHTAISYCHETMTGWVHDVLGRNWACFVGKKVES
umanx3      TIIYNQGFEIVLNDYKWFAFFKYKEEGSKVTTYCNETMTGWVHDVLGRNWACFTGKKVGT
eqxxx0      TLIYNQGFEVTINHRKWLVIFAYKSNGE---FNCHKSMPMWTHDTLIRQWKCFVAEKIGV
Bxxxx1      TLIYNQGFEVTINHRKWLVIFAYKSNGE---FNCHKSMPMWTHDTLIDSG-SVCSGKIGV
jCCxx2      TLIYNQGFEITMNHRKWLIMFAYGPNNT---YTCNKSMPMWTHDTLICQWHCFTATKVNH
            *:*******:.:*. **: :* *   ..        *:::*. *.**.*  .  ... *:
rim.cons.   TLIYNQGFEI2LN22KW2AFF2YK2NG23AI2YCH22M22W2HD2L2RNWACFVGKKVGV

                   130       140       150       160       170       180
                     |         |         |         |         |         |
atxxx4      HSEKVYVNVAHLGGLQEKYSERLYSHNHNFVKAINSVQKSWTATTYEEYEKLSIRDLIRR
ousex5      HIEKVNMNAAHLGGLQERYSERLYTHNHNFVKAINTVQKSWTATAYKEYEKMSLRDLIRR
umanx3      ASENVYVNTAHLKNSQEKYSNRLYKYDHNFVKAINAIQKSWTATTYMEYETLTLGDMIRR
eqxxx0      HDKFHINKLFGS----KSFGRTLYHINPSFVDKINAHQKSWRAEIYPELSKYTIDELRNR
Bxxxx1      HDKFHINKLFGS----KSFGRTLYHINPSFVGKINAHQKSWRGEIYPELSKYTIDELRNR
jCCxx2      FQRMIEYKSPVL----QLDGNQLYKVDTKFIKAINAKQNSWKATIYPEYSKYTIKEMRRR
              .   :       :  .. **  : .*:  **: *:** .  * * .. :: :: .*
rim.cons.   H2E2V222LAHLGGLQE2Y23RLY22NHNFVKAINA2QKSWTATIYPEY2KYTI22L2RR

                   190       200       210       220       230       240
                     |         |         |         |         |         |
atxxx4      SG-HSGRILRPKPAP----ITDEIQQQILSLPESWDWRN-VRGIN-FVSPVRNQESCGSC
ousex5      SG-HSQRIPRPKPAP----MTDEIQQQILNLPESWDWRN-VQGVN-YVSPVRNQESCGSC
umanx3      SGGHSRKIPRPKPAP----LTAEIQQKILHLPTSWDWRN-VHGIN-FVSPVRNQASCGSC
eqxxx0      AGGVKSMVTRPSVLN-RKTPSKELISLTGNLPLEFDWTSPPDGSRSPVTPIRNQGICGSC
Bxxxx1      AGGVKSMVTRPSVLN-RKTPSKELISLTGNLPLEFDWTSPPDGSRSPVTPIRNQGICGSC
jCCxx2      AGGSRSAFKRQNVQLPKKNLTSAMMLELLALPKEFDWVNRPEGLRSPVTPVRNQKTCGSC
            :*     . *  .       : :      ** .:** .   *  .   *:*:*** ****
rim.cons.   2GGHSS2I2RPK2APPRKT2T2EIQQ2ILNLP222DWRNP2DG22SPV2PVRNQ2SCGSC

                   250       260       270       280       290       300
                     |         |         |         |         |         |
atxxx4      YSFASLGMLEARIRILTNNSQTPILSPQEVVSCSPYAQGCDGGFPYLIAGKYAQDFGVVE
ousex5      YSFASMGMLEARIRILTNNSQTPILSPQEVVSCSPYAQGCDGGFPYLIAGKYAQDFGVVE
umanx3      YSFASMGMLEARIRILTNNSQTPILSPQEVVSCSQYAQGCEGGFPYLIAGKYAQDFGLVE
eqxxx0      YAFASAAALEARIRLVSNFSEQPILSPQAVVDCSPYSEGCNGGFPFLIAGKYGEDFGFVS
Bxxxx1      YASPSAAALEARIRLVSNFSEQPILSPQTVVDCSPYSEGCNGGFPFLIAGKYGEDFGLPQ
jCCxx2      YAFASTAAIEARIRLASRFRLQPILSPQDIIDCSPYSEGCDGGFPYLVAGKHGEDFGFVE
            *: .*  . :*****: :.    ****** .:.** *:**:****:*:***:.:***. .
rim.cons.   Y2FAS222LEARIR2L2N2SQ2PILSPQEVV2CSPY22GCDGGFPYLIAGKY22DFG3VE

                   310       320       330       340       350       360
                     |         |         |         |         |         |
atxxx4      ENCFPYTATDA-PCKPKENCLRYYSSEYYYVGGFYGGCNEALMKLELVKHGPMAVAFEVH
ousex5      ESCFPYTAKDS-PCKPRENCLRYYSSDYYYVGGFYGGCNEALMKLELVKHGPMAVAFEVH
umanx3      EACFPYTGTDS-PCKMKEDCFRYYSSEYHYVGGFYGGCNEALMKLELVHHGPMAVAFEVY
eqxxx0      ENCDPYTGEDTGKCTVSKNCTRYYTADYSYIGGYYGATNEKLMQLELISNGPFPVGFEVY
Bxxxx1      KIVIPYTGEDTGKCTVSKNCTRYYTTDYSYIGGYYGATNEKLMQLELISNGPFPVGFEVY
jCCxx2      EKCNPYTGVKSGTCNKLLGCTRYYTTDYHYIGGYYGATNEDLMKLELVKNGPFPVGFEVY
            :   ***. .:  *.   .* ***:::* *:**:**. ** **:***: :**:.*.***:
```

```
rim.cons.    ENCFPYTG2DSGPCK22ENCTRYY2SDY3Y2GG2YG22NEALMKLELVK2GP22V2FEVY

                    370       380       390       400       410       420
                     |         |         |         |         |         |
atxxx4       DDFLHYHSGIYHHTGLS---DPFNPFELTNHAVLLVGYGKDPVTGLDYWIVKNSWGSQWG
ousex5       DDFLHYHSGIYHHTGLS---DPFNPFELTNHAVLLVGYGRDPVTGIEYWIIKNSWGSNWG
umanx3       DDFLHYKKGIYHHTGLR---DPFNPFELTNHAVLLVGYGTDSASGMDYWIVKNSWGTGWG
eqxxx0       EDFQFYKEGIYHHTTVQNDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVKNSWGVEWG
Bxxxx1       EDFQFYKEGIYHHTTVQTDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVKNSWGVEWG
jCCxx2       GDFLQYKSGVYSHTDIINNHHPFNPFELTNHAVLLVGYGIDNSSNLPYWKIKNSWGQYWG
              **  *:.*:* ** :      ****************** *  :.  ** :*****  **
rim.cons.    DDFLHYKSGIYHHTGL2NDHDPFNPFELTNHAVLLVGYGVD22SG2PYW2VKNSWG2EWG

                    430       440
                     |         |
atxxx4       ESGYFRIRRGTDECAIESIAMAAIPIPKL
ousex5       ESGYFRIRRGTDECAIESIAVAAIPIPKL
umanx3       ENGYFRIRRGTDECAIESIAVAATPIPKL
eqxxx0       EQGYFRILRGTDECGVESLGVRFDPVL--
Bxxxx1       EQGYFRILRGTDECGVESLGVRFDPVL--
jCCxx2       EEGYFRILRGSDECGVQSIAIKFDVVL--
             *.***** **:***.::*:.:    :
rim.cons.    E2GYFRI2RGTDEC22ESIAVA2DP22KL
```

Alignment data :

Alignment length : 449
Identity (*) : 165 is 36.75 %
Strongly similar (:) : 70 is 15.59 %
Weakly similar (.) : 51 is 11.36 %
Different : 163 is 36.30 %
Sequence 0001 : Ratxxx4 (438 residues).
Sequence 0002 : Mousex5 (438 residues).
Sequence 0003 : Humanx3 (439 residues).
Sequence 0004 : Seqxxx0 (435 residues).
Sequence 0005 : GBxxxx1 (434 residues).
Sequence 0006 : SjCCxx2 (436 residues).

CLUSTALW options used :

endgaps=1
gapdist=8
gapext=0.2
gapopen=10.0
gapresidues=GPSNDQERK
ktuple=1
matrix=gonnet
maxdiv=30
outorder=aligned
pairgap=3
score=percent
topdiags=5
type=PROTEIN
window=5

Result files (text) :

CLUSTALW

Garnier parameters

Monday, October 7th 2002 : InterProScan is available (see news)

Job **SOPM** (ID: 6412) is running on **NPS@** server (started on Tue Feb 11 09:08:30 CET 2003).
Results will be shown below. **Please wait and don't go back.**

n your publication cite :
PS@: Network Protein Sequence Analysis
IBS 2000 March Vol. 25, No 3 [291]:147-150
ombet C., Blanchet C., Geourjon C. and Deléage G.

SOPM result for : SmCCxx0

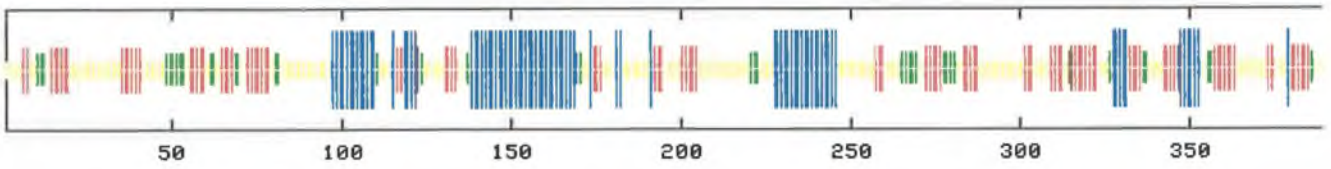
bstract Geourjon, C. & Deléage, G., SOPM: a self-optimised method for protein secondary structure prediction., Protein ngineering (1994) 7, 157-164

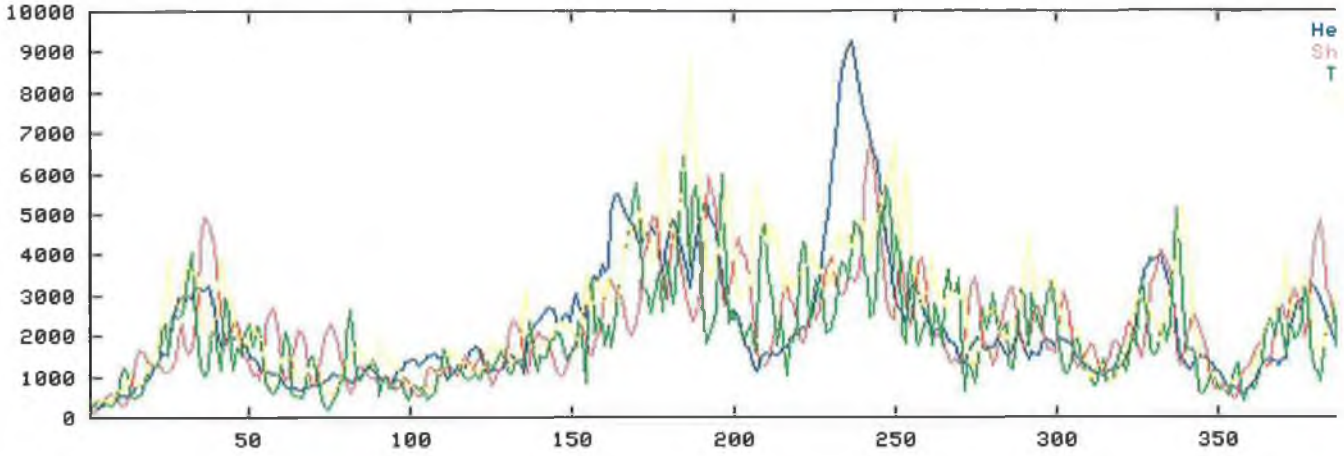
iew SOPM in: [MPSA (Mac, UNIX) , About...] [AnTheProt (PC) , Download...] [HELP]

```
        10         20         30         40         50         60         70
         |          |          |          |          |          |          |
TPANCTYEDAHGRWKFHIGDYQSKCPEKLNSKQSVVISLLYPDIAIDEFGNRGHWTLIYNQGFEVTINH
 ccccee cttteeeeee ccccccccccccccceeeeee ccccctttttt eeeee tt eeeett
KWLVIFAYKSNGEFNCHKSMPMWTHDTLIRQWKCFVAEKIGVHDKFHINKLFGSKSFGRTLYHINPSFV
 eeeeeee tt ccccccccccccccc hhhhhhhhhhhhht cccc heehhhhet ccccccc eeee ccc thh
KINAHQKSWRAEIYPELSKYTIDELRNRAGGVKSMVTRPSVLNRKTPSKELISLTGNLPLEFDWTSPPD
hhhhhhhhhhhhhhhhhhhhhhhhhhhhtt cheee cccc hh ccccccc heee ccccc eeeee ccccc
SRSPVTPIRNQGICGSCYAFASAAALEARIRLVSNFSEQPILSPQAVVDCSPYSEGCNGGFPFLIAGKY
 ccccccccttt ccc hhhhhhhhhhhhhhhhhh ccccccccccc eee ccccc ttttt cc eeeeettt
EDFGFVSENCDPYTGEDTGKCTVSKNCTRYYTADYSYIGGYYGATNEKLMQLELISNGPFPVGFEVYED
 cceeeee ccccccccccccc eee ccccc eeee c teeeeeeee ccc thhhhheeeett ccccc eeeee hhh
QFYKEGIYHHTTVQNDHYNFNPFELTNHAVLLVGYGVDKLSGEPYWKVKNSWGVEWGEQGYFRILRGTD
hh cc tteeeeeee ccccccccc ee ccc heeeeeet c ccc tt c eeee ccccccc ettttteeeee ccc
CGVESLGVRFDPVL
 tt cc eeeeec c tee
```

```
equence length :   435

OPM :
  Alpha helix      (Hh) :    84 is  19.31%
  3 10  helix      (Gg) :     0 is   0.00%
  Pi helix         (Ii) :     0 is   0.00%
  Beta bridge      (Bb) :     0 is   0.00%
  Extended strand  (Ee) :   117 is  26.90%
  Beta turn        (Tt) :    49 is  11.26%
  Bend region      (Ss) :     0 is   0.00%
  Random coil      (Cc) :   185 is  42.53%
  Ambigous states  (?)  :     0 is   0.00%
  Other states          :     0 is   0.00%
```






```
arameters :
  Window width          :  17
  Similarity threshold  :   8
  Number of states      :   4
```

rediction result file (text): [SOPM]

---

*ser : public@136.206.1.19. **Last modification time :** Tue Feb 11 09:08:35 2003. **Current time :** Tue eb 11 09:08:35 2003 This service is supported by Ministere de la recherche (ACC-SV13), CNRS IMABIO, COMI, GENOME) and Région Rhône-Alpes (Programme EMERGENCE) . Comments.*